Proceedings of Machine Learning Research vol 117:1–50, 2020 31st International Conference on Algorithmic Learning Theory

Optimal δ-Correct Best-Arm Selection for Heavy-Tailed Distributions

Shubhada Agrawal SHUBHADA.AGRAWAL@TIFR.RES.IN
Sandeep Juneja JUNEJA@TIFR.RES.IN
TIFR, Mumbai

Peter Glynn GLYNN@STANFORD.EDU
Stanford University

Editors: Aryeh Kontorovich and Gergely Neu

Abstract

Given a finite set of unknown distributions *or arms* that can be sampled, we consider the problem of identifying the one with the largest mean using a delta-correct algorithm (an adaptive, sequential algorithm that restricts the probability of error to a specified delta) that has minimum sample complexity. Lower bounds for delta-correct algorithms are well known. Delta-correct algorithms that match the lower bound asymptotically as delta reduces to zero have been previously developed when arm distributions are restricted to a single parameter exponential family. In this paper, we first observe a negative result that some restrictions are essential, as otherwise under a delta-correct algorithm, distributions with unbounded support would require an infinite number of samples in expectation. We then propose a delta-correct algorithm that matches the lower bound as delta reduces to zero under the mild restriction that a known bound on the expectation of a non-negative, continuous, increasing convex function (for example, the squared moment) of the underlying random variables, exists. We also propose batch processing and identify near optimal batch sizes to substantially speed up the proposed algorithm. The best-arm problem has many learning applications, including recommendation systems and product selection. It is also a well studied classic problem in the simulation community.

Keywords: Multi-armed bandits, best-arm identification, sequential learning, ranking and selection

1. Introduction

Given a vector of unknown arms or probability distributions that can be sampled, we consider algorithms that sequentially sample from or *pull* these arms and at termination identify the best-arm, i.e., the arm with the largest mean. The algorithms considered provide δ-correct probabilistic guarantees, that is, the probability of identifying an incorrect arm is bounded from above by a pre-specified $\delta > 0$. Further, the δ-correct algorithms that we consider aim to minimize the sample complexity, or, equivalently, the expected total number of arms pulled before they terminate. This best-arm problem is well studied in the literature (see, e.g., in learning - Garivier and Kaufmann (2016); Kaufmann et al. (2016); Russo (2016); Jamieson et al. (2014); Bubeck et al. (2011); Audibert and Bubeck (2010); Even-Dar et al. (2006); Mannor and Tsitsiklis (2004); in earlier statistics literature - Jennison et al. (1982); Bechhofer et al. (1968); Paulson et al. (1964); in simulation - Glynn and Juneja (2004); Kim and Nelson (2001); Chen et al. (2000); Dai (1996); Ho et al. (1992)).

The δ-correct guarantee imposes constraints on the expected number of times each arm must be pulled by the algorithm. These constraints are made explicit by Garivier and Kaufmann (2016) through their transportation inequality which can be used to arrive at a max-min optimization problem to develop efficient lower bounds on δ-correct algorithms. This line of work relies on *change of measure* based analysis that goes back at least to Lai and Robbins (1985). Also see Mannor and Tsitsiklis (2004) and Burnetas and Katehakis (1996). It is important to emphasize that the max-min optimization problem to develop efficient lower bound on a δ-correct algorithm, requires complete knowledge of the underlying arm distributions, and its solution is a function of these underlying distributions. The algorithms, on the other hand, acting on a given set of arms, are unaware of the underlying distributions, and, typically, adaptively learn them to decide on the sequence of arms to pull, as well as the stopping time.

Garivier and Kaufmann (2016) consider the best arm problem under the assumption that each arm distribution belongs to a single parameter exponential family (SPEF). Under this restriction, they arrive at an *asymptotically optimal* algorithm having a sample complexity matching the derived lower bound asymptotically as $\delta \to 0$. SPEF distributions include Bernoulli, Poisson and Gaussian distributions with known variance. However, in practice it is rarely the case (other than in the Bernoulli setting) that the arm distributions are from SPEF, so there is a need for a general theory as well as efficient algorithms that have wider applicability. Our paper substantially addresses this issue.

Contributions: Our first contribution is an impossibility result illustrating why some distributional restrictions on arms are necessary for δ-correct algorithms to be effective. Consider an algorithm that provides δ-correct guarantees when acting on a finite set of distributions belonging to a collection $\mathcal{U}$, where $\mathcal{U}$ comprises distributions with unbounded support that are KL *right dense* (defined in Section 2). In this set-up we show that the sample complexity of the algorithm in every instance of it acting on a finite set of distributions in $\mathcal{U}$, must be infinite. Examples of such a $\mathcal{U}$ include all light-tailed distributions with unbounded support (a distribution is said to be light tailed if its moment generating function is finite in a neighborhood of zero). Another example is a collection of unbounded distributions supported on $\Re$ that are in $\mathcal{L}^p$, for some $p \geq 1$. That is, for some $p \geq 1$, their absolute p^{th} moment is finite.

To arrive at an effective δ-correct algorithm, we restrict arm distributions to the collection

$$\mathcal{L} \triangleq \left\{ \eta \in \mathcal{P}(\Re) : \mathbb{E}_{X\sim\eta}(f(|X|)) \leq B \right\}, \tag{1}$$

where $\mathcal{P}(\Re)$ denotes the set of probability distributions with support in $\Re$, $f(\cdot)$ is a continuous, non-negative, convex function such that $f(y)/y \xrightarrow{y\to\infty} \infty$, and B is a known positive constant. For instance, we may have $f(y) = y^{1+\epsilon}$ for any $\epsilon > 0$ or $f(y) = y \log y$. In simulation models, upper bounds on moments of simulation output, as in (1), can often be found by the use of Lyapunov function based techniques (see, e.g., Glynn and Zeevi (2008)). With this mild restriction we solve the associated optimization problem to arrive at an efficient lower bound on sample complexity for δ-correct algorithms, which involves a max-min optimization problem. We also develop simple line search based procedures to solve this optimization problem. Our **main contribution** is the development of an *asymptotically optimal* δ-correct algorithm whose sample complexity matches the derived lower bound asymptotically as $\delta \to 0$.

Key to developing such a lower bound and the δ-correct algorithm is the functional $\text{KL}_{\text{inf}}(\eta, x)$ defined as follows: let $\text{KL}(\kappa_1, \kappa_2)$ denote the Kullback-Leibler divergence between probability distributions κ_1 and κ_2, and let $m(\kappa)$ denote the mean of the probability distribution κ. Then, for $\eta \in \mathcal{P}(\Re)$ and $x \in \Re$ such that $f(|x|) < B$, $\text{KL}_{\text{inf}}(\eta, x)$ is the optimal value of $\min_{\kappa\in\mathcal{L}} \text{KL}(\eta, \kappa)$, such that $m(\kappa) \geq x$, for $x \geq m(\eta)$, and $m(\kappa) \leq x$, for $x < m(\eta)$. Call this optimization problem

$\mathcal{O}_1$. Heuristically, $\mathrm{KL}_{\inf}(\eta, x)$ measures the difficulty of separating distribution η from all other distributions in $\mathcal{L}$ whose mean equals x. It equals zero when $x = m(\eta)$ and $\eta \in \mathcal{L}$.

We develop a concentration inequality for $\mathrm{KL}_{\inf}(\hat{\kappa}(n), m(\kappa))$, where for $\kappa \in \mathcal{L}$, $\hat{\kappa}(n)$ denotes the empirical distribution corresponding to n samples from κ. This plays a key role in the proof of δ-correctness of the proposed algorithm. A key step in the proof of the concentration inequality relies on arriving at a simpler dual representation of $\mathrm{KL}_{\inf}(\cdot,\cdot)$. Here, we substantially extend the representation developed by Honda and Takemura (2010) for bounded random variables to random variables belonging to $\mathcal{L}$. Honda and Takemura (2010) had focussed on the regret minimization problem for stochastic bandits (also see Burnetas and Katehakis (1996); Honda and Takemura (2011)).

To prove the δ-correctness of the proposed algorithm, we further develop a concentration inequality for $\sum_{a=1}^{K} N_a(n)\,\mathrm{KL}_{\inf}(\hat{\kappa}_a(n), m(\kappa_a))$ where $N_a(n)$ denotes the number of times arm a is pulled in a total of n arm pulls, and $\hat{\kappa}_a(n)$ denotes the empirical distribution for arm a based on $N_a(n)$ samples. While Magureanu et al. (2014) have developed these inequalities for the Bernoulli distribution, we generalize their analysis to arm distributions belonging to $\mathcal{L}$.

In bounding the sample complexity of the proposed algorithm, we exploit the continuity of $\mathrm{KL}_{\inf}(\eta, x)$ in η as well as the continuity of the solution to the max-min lower bound problem with respect to the underlying arm distributions. We achieve this by considering the Wasserstein distance in the space of probability distributions $\mathcal{L}$. The Wasserstein distance is relatively tractable to work with, and it can be seen that $\mathcal{L}$ is a compact metric space under this distance. This, in particular, allows us to use the well-known Berge's Maximum Theorem (stated in the Appendix A) to derive the requisite continuity properties.

In, e.g., Garivier and Kaufmann (2016); Kalyanakrishnan et al. (2012), the proposed algorithms solve the lower bound problem at every iteration. However, solving the corresponding max-min optimization problem can be computationally demanding particularly in the generality that we consider. We instead solve this problem in batches and arrive at near optimal batch sizes, that result in a provably substantial computational reduction.

Best arm problems arise in many settings in practice. For instance, one can view the selection of the best product version to roll out for production and sale, after a set of expensive pilot trials among many competing versions to be a best arm problem. In simulation theory, selecting the best design amongst many (based on output from a simulation model) is a classic problem, with applications to manufacturing, road and communications network design, etc. In these and many other settings, the underlying distributions can be very general and may not be modelled well by a SPEF distribution.

Roadmap: In Section 2 we review some background material and present an impossibility result illustrating the need for distributional restrictions on arms. In Section 3, an efficient lower bound for δ-correct algorithms for the best arm problem is provided when the arm distributions are restricted to $\mathcal{L}$. The algorithm that matches this lower bound asymptotically as $\delta \to 0$ is developed in Section 4. Enroute, we develop certain concentration inequalities associated with $\mathrm{KL}_{\inf}$. Discussion on optimal batch size and a numerical experiment are shown in Section 5. While key ideas in some of the proofs are outlined in the main body, proof details are all given in the Appendices.

2. Background and the impossibility result

Let $\mathcal{U}$ denote the universe of probability distributions for which we aim to devise a δ-correct algorithm. We assume that each distribution in $\mathcal{U}$ has finite mean. Let the Kullback-Leibler divergence between distributions η and κ be denoted by $\mathrm{KL}(\eta,\ \kappa) = \int \log\left(\frac{d\eta}{d\kappa}(x)\right) d\eta(x)$. For $p, q \in (0,1)$, let $\rho(p,q)$ denote the KL-divergence between Bernoulli distributions with mean p and q, respectively, that is, $\rho(p,q) = p\log\left(\frac{p}{q}\right) + (1-p)\log\left(\frac{1-p}{1-q}\right)$.

Recall that $m(\eta) = \int_{x\in\Re} x d\eta(x)$, denotes the mean of any distribution $\eta \in \mathcal{U}$. Let $\mathcal{M}_{\mathcal{U}}$ denote the collection of all $\nu = (\nu_1, \ldots, \nu_K)$ such that each $\nu_i \in \mathcal{U}$. Consider a vector of distributions $\mu = (\mu_1, \ldots, \mu_K)$ from $\mathcal{M}_{\mathcal{U}}$. Without loss of generality, henceforth we assume that the highest-mean arm in μ is arm 1, that is, $m(\mu_1) > \max_{i\neq 1} m(\mu_i)$. Let $\tilde{\mathcal{A}}$ denote the collection of all distributions $\nu = (\nu_1, \ldots, \nu_K)$ such that each $\nu_i \in \mathcal{U}$ and $m(\nu_1) \leq \max_{i\neq 1} m(\nu_i)$.

Under a δ-correct algorithm acting on μ, for $\delta \in (0,1)$, the following transportation inequality is shown by Kaufmann et al. (2016):

$$\sum_{i=1}^{K} \mathbb{E}_{\mu}(N_i(\tau))\ \mathrm{KL}(\mu_i, \nu_i) \geq \rho(\delta, 1-\delta) \geq \log\left(\frac{1}{2.4\delta}\right) \tag{2}$$

for any $\nu \in \tilde{\mathcal{A}}$, where $N_i(t)$ denotes the number of times arm i is pulled by the algorithm in t trials, and $\tau = \sum_{i=1}^{K} N_i(\tau)$ denotes the algorithm termination time. Intuitively, this specifies a lower bound on the expected number of samples that need to be generated from each arm i under μ, for an algorithm to separate it from a distribution ν belonging to the set of alternative hypotheses $\tilde{\mathcal{A}}$, with probability at least $1-\delta$.

The following lemma helps in proving our negative result in Theorem 3.

Lemma 1 *Given η with an unbounded support on the positive real line, for any finite $a > 0$ and $b > m(\eta)$, there exists a distribution κ such that*

$$\mathrm{KL}(\eta, \kappa) \leq a \qquad \textit{and} \qquad m(\kappa) \geq b. \tag{3}$$

Definition 2 *A collection of probability distributions $\mathcal{U}$ is referred to as* KL *right dense, if for every $\eta \in \mathcal{U}$, and every $a > 0$, $b > m(\eta)$, there exists a distribution $\kappa \in \mathcal{U}$ such that (3) holds.*

Observe that a necessary condition for $\mathcal{U}$ to be KL right dense is that each member does not have a real-valued upper bound on its support.

Theorem 3 *Under a δ-correct algorithm operating on* KL *right dense $\mathcal{U}$, for any $\mu \in \mathcal{M}_{\mathcal{U}}$,*

$$\mathbb{E}_k N_k(\tau) = \infty, \qquad \textit{for all}\ \ 2 \leq k \leq K. \tag{4}$$

The proof follows easily from (3) in Lemma 1, since given $\mu \in \mathcal{M}_{\mathcal{U}}$ such that arm 1 has the maximum mean, for any $k \geq 2$, one can easily find $\nu \in \tilde{\mathcal{A}}$ such that $\nu_i = \mu_i$ for $i \neq k$, $m(\nu_k) > m(\mu_1)$, and $\mathrm{KL}(\mu_k, \nu_k)$ is arbitrarily small. (4) now follows from (2).

When only information available about a distribution is that its mean exists, Bahadur and Savage (1956) prove a related impossibility result that there does not exist an effective test of hypothesis for testing whether the mean of the distribution is zero (also see Lehmann and Romano (2006)). However, to the best of our knowledge, Theorem 3 is the first impossibility result in the best arm setting (or, equivalently, the ranking and selection setting; see Glynn and Juneja (2018) for further discussion in the best arm settings).

3. Lower bound for a δ-correct algorithm

Theorem 3 suggests that some restrictions are needed on $\mathcal{U}$ for δ-correct algorithms to provide reasonable performance guarantees. To this end, we limit our analysis to δ-correct algorithms acting on the class $\mathcal{L}$ defined in (1) earlier. Let $\mathcal{M}_{\mathcal{L}}$ denote the collection of vectors $\nu = (\nu_1, \ldots, \nu_K)$, such that each $\nu_i \in \mathcal{L}$. Let $\mu \in \mathcal{M}_{\mathcal{L}}$. Recall that without loss of generality, $m(\mu_1) > \max_{j\geq 2} m(\mu_j)$. Let $\mathcal{A}$ denote the collection of all distributions $\nu = (\nu_1, \ldots, \nu_K)$ such that each $\nu_i \in \mathcal{L}$ and $m(\nu_1) \leq \max_{i\neq 1} m(\nu_i)$. From (2) it follows that for any δ-correct algorithm acting on μ:

$$\mathbb{E}_\mu(\tau) \inf_{\nu\in\mathcal{A}} \sum_{i=1}^{K} \frac{\mathbb{E}_\mu(N_i(\tau))}{\mathbb{E}_\mu(\tau)} \, \mathrm{KL}(\mu_i, \nu_i) \geq \log\left(\frac{1}{2.4\delta}\right).$$

Let Σ^K denote probability simplex in $\Re^K$. It follows that $\mathbb{E}_\mu(\tau)$ is bounded from below by $\log\left(\frac{1}{2.4\delta}\right)$ times the inverse of

$$\sup_{t\in\Sigma^K} \inf_{\nu\in\mathcal{A}} \sum_{i=1}^{K} t_i \, \mathrm{KL}(\mu_i, \nu_i), \tag{5}$$

and hence the problem of computing the lower bound on $\mathbb{E}_\mu(\tau)$ reduces to solving the above max-min problem. To characterize the solution to (5), we need some definitions.

Recall that for $\eta \in \mathcal{P}(\Re)$ and $x \in \Re$ such that $f(|x|) < B$, $\mathrm{KL}_{\inf}(\eta, x)$ is defined as the value of $\mathcal{O}_1$. As mentioned in the introduction, we study the continuity of $\mathrm{KL}_{\inf}(\eta, x)$ as a function of η in the Wasserstein metric.

Wasserstein metric: (see, e.g., Villani (2003)). Recall that the Wasserstein metric $d_W(\cdot, \cdot)$ between probability distributions κ and η on $\Re$ is given by:

$$d_W(\kappa, \eta) = \inf_{\gamma\in\Gamma(\kappa,\eta)} \int_{\Re} \int_{\Re} d(x, y) d\gamma(x, y), \tag{6}$$

where $\Gamma(\kappa, \eta)$ denotes the collection of measures on $\Re^2$ with marginals κ, η on the first and second coordinates, respectively, and $d(\cdot, \cdot)$ is any metric on $\Re$. For simplicity, we consider $d(x, y) = |x - y|$. We endow $\mathcal{P}(\Re)$ with the corresponding Wasserstein metric, d_W. Then, $(\mathcal{P}(\Re), d_W)$ is a metric space (see Section 7.1 Villani (2003)). $\mathcal{L}$ as a subset of $\mathcal{P}(\Re)$, is also a metric space with d_W being the metric. Hence, we can define continuity of functions from $(\mathcal{L}, d_W)$ to $(\Re, d)$. Furthermore, for x and y in $\Re^K$, $d_K(x, y) = \sum_{i=1}^{K} d(x_i, y_i)$ is a metric on $\Re^K$. Thus, $\mathcal{M}_{\mathcal{L}}$ endowed with d_W defined with $d = d_K$ in (6), is a metric space and we can define continuous functions from $(\mathcal{M}_{\mathcal{L}}, d_W)$ to $(\Re^k, d_K)$. Lemma 4 lists some properties of the set $\mathcal{L}$ and $\mathrm{KL}_{\inf}$ that give insights into geometrical structure of $\mathrm{KL}_{\inf}$ and are useful for our analysis. These are proved in Appendix B.1.

Lemma 4 *The set $\mathcal{L}$ is uniformly-integrable and compact in the Wasserstein metric. Moreover, for $x \in \Re$ such that $f(|x|) < B$ and $\eta \in \mathcal{P}(\Re)$, $\mathrm{KL}_{\inf}(\eta, x)$ is increasing for $x > m(\eta)$, and decreasing for $x < m(\eta)$. It is continuous and convex in η, and convex and twice differentiable in x. Furthermore, for $\eta \in \mathcal{L}$, it satisfies $\mathrm{KL}_{\inf}(\eta, x) \leq \frac{f^{-1}(B)}{f^{-1}(B)-x}$, $\mathrm{KL}_{\inf}(\eta, m(\eta)) = 0$, and $\frac{\partial \mathrm{KL}_{\inf}(\eta, m(\eta))}{\partial x} = 0$.*

Let $\mathcal{B} = \left\{\boldsymbol{x} \in \Re^K : \text{ for } i \leq K,\ x_i = m(\nu_i) \text{ for some } \nu_i \in \mathcal{L}, \text{ and } x_1 \leq \max_{i\neq 1} x_i\right\}$. Observe that the max-min problem (5) may be re-expressed as

$$\sup_{t\in\Sigma^K} \inf_{\boldsymbol{x}\in\mathcal{B}} \sum_{i=1}^{K} t_i \, \mathrm{KL}_{\inf}(\mu_i, x_i). \tag{7}$$

The inner optimization problem in (7) can be further simplified. Given $\mu \in \mathcal{M}_{\mathcal{L}}$, let $V(\mu)$ denote the optimal value of the expression in (7) (or, equivalently, (5)) and let $T(\mu)$ denote the set of $t \in \Sigma_K$ that maximize (7). Furthermore, recall that $m(\mu_1) > \max_{j \neq 1} m(\mu_j)$. For $j \in \{2, \ldots, K\}$, and μ fixed, consider functions $g_j : \Re^3 \to \Re$ and $G_j : \Re^2 \to \Re$ given by

$$g_j(y, z, x) \triangleq y \, \mathrm{KL}_{\mathrm{inf}}(\mu_1, x) + z \, \mathrm{KL}_{\mathrm{inf}}(\mu_j, x), \quad \text{and} \quad G_j(y, z) \triangleq \inf_{x \in [m(\mu_j), m(\mu_1)]} g_j(y, z, x). \tag{8}$$

The following theorem characterizes the solution to $V(\mu)$, given that μ is known.

Theorem 5 *The set $T(\mu)$ is a singleton. Moreover, the optimal value of the max-min problem (7),*

$$V(\mu) = \max_{t \in \Sigma_K} \min_{j > 1} G_j(t_1, t_j). \tag{9}$$

Further, the max-min problem (7) is solved by $t^* \in \Sigma_K$ that uniquely satisfies

1. $\forall i, \; t_i^* > 0$ *and* $\sum_i t_i^* = 1$,
2. $\sum_{j=2}^{K} \frac{\mathrm{KL}_{\mathrm{inf}}(\mu_1, x_j(t_1^*, t_j^*))}{\mathrm{KL}_{\mathrm{inf}}(\mu_j, x_j(t_1^*, t_j^*))} = 1$, *where $x_j(y, z)$ denotes the unique* $\arg\inf_{x \in [m(\mu_j), m(\mu_1)]} g_j(y, z, x)$.
3. $G_2(t_1^*, t_2^*) = G_j(t_1^*, t_j^*)$, *for $j \in \{3, \ldots, K\}$.*

Moreover, the optimal value, $V(\mu)$ equals $G_2(t_1^*, t_2^*)$, and the optimal proportion vector, $t^* : \mathcal{M}_{\mathcal{L}} \to \Sigma_K$, is a continuous function (in the Wasserstein metric) of μ.

The above characterization is analogous to that in Theorem 1, Glynn and Juneja (2004) where they considered the fixed budget best-arm problem. It is easily seen that the fixed budget setting also lends itself to solving a max-min problem analogous to (5), where the arguments μ_i and ν_i in each KL term are switched. The above characterization also generalizes that in Garivier and Kaufmann (2016) for SPEF of distributions. Observe that when η belongs to SPEF and $\mathcal{L}$ is restricted to the same SPEF, $\mathrm{KL}_{\mathrm{inf}}(\eta, x)$ corresponds to $\mathrm{KL}(\eta, \kappa)$ where κ denotes the corresponding SPEF distribution with mean x.

Remark 6 The proposed algorithm discussed in Section 4, relies on solving the max-min problem (7) repeatedly with μ replaced by its empirical estimator. Since this estimator may not necessarily belong to $\mathcal{M}_{\mathcal{L}}$, it is important to note that the lower bound in (5) and the results in Theorem 5 hold for every $\mu \in (\mathcal{P}(\Re))^K$. Furthermore, since the max-min problem (7) needs to be solved multiple times in our proposed algorithm, efficiently solving it for the optimal proportions $t^*(\nu)$ for any $\nu \in (\mathcal{P}(\Re))^K$ is crucial to it.

Remark 7 (Numerically solving the max-min problem:) Let $c^* = V(\mu)$ denote the common value of $G_j(t_1^*, t_j^*)$ for $j \in \{2, \ldots, K\}$. We develop an algorithm that relies on repeated single dimensional line-searches to solve for t^* and c^*. Appendix B.3 contains the details of the algorithm and proofs of its convergence to the correct value. To get an idea of the computational effort needed in solving (7), let τ_0 denote the average time taken to compute $\mathrm{KL}_{\mathrm{inf}}$ using efficient solvers (Theorem 12 below shows that $\mathrm{KL}_{\mathrm{inf}}$ has a dual representation, where it is a solution to a two variable concave program). Let ϵ_L denote the tolerance for each line search. Then, numerically solving for $t^*(\mu)$ and $V(\mu)$ takes time $\tau_0(K-1)\times O\left(\log^3 \frac{1}{\epsilon_L}\right)$. To decrease this computation time, we pre-compute values of $\mathrm{KL}_{\inf}(\mu_i, y)$ for each y along a grid, for each μ_i. For y not in this set, we linearly interpolate from the computed values. This substantially reduces the computation time of the algorithm to $(K-1)\times O\left(\log^3 \frac{1}{\epsilon_L}\right)+\tilde{\tau}_0$, where $\tilde{\tau}_0$ is time for the pre-processing step.

4. The δ-correct algorithm

We now propose a δ-correct algorithm and show that its sample complexity matches the lower bound up to the first order as $\delta \to 0$. Recall that a δ-correct algorithm has a sampling rule that at any stage, based on the information available, decides which arm to sample next. Further, it has a stopping rule, and at the stopping time it announces the arm with the largest mean while ensuring that the probability of incorrect assessment is at most a pre-specified $\delta \in (0,1)$.

It can be shown that if the distribution of the K arms, μ, is not known, but there exists an oracle that informs us the optimal $t^*(\mu)$ that solves (7), then sampling arms to closely match the proportions in $t^*(\mu)$ leads to an asymptotically optimal algorithm (this can be seen, for instance, by using the stopping rule that is analogous to ours, and essentially repeating the arguments in our proof where approximations to $t^*(\mu)$ are used). This suggests that the fraction of times a good algorithm pulls an arm j should converge to $t^*_j(\mu)$. We propose a sampling rule to ensure this. Our stopping rule (discussed above (10)) relies on a generalized likelihood ratio statistic taking sufficiently large value.

Sampling rule: Our sampling algorithm relies on solving the max-min lower bound problem with the vector of empirical distributions used as a proxy for the unknown true distribution μ. The computed optimal proportions then guide the sampling strategy. Garivier and Kaufmann (2016) and Juneja and Krishnasamy (2019) follow a similar plug-in strategy for SPEF distributions, where empirically observed means are used as a proxy for true means. The proposed algorithm conducts some exploration to ensure that no arm is starved with insufficient samples. Because solving the max-min lower bound problem can be computationally demanding, we solve it periodically after fixed, well chosen $m > 1$ samples (which is allowed to be a function of δ), where m may be optimised to minimize the overall computation effort.

The specific algorithm, $\mathbf{AL_1}$, is as follows:

1. Initialize by allocating m samples in round-robin way to generate at least $\lfloor \frac{m}{K} \rfloor$ samples from each arm. Set $l = 1$ and let lm denote the total number of samples generated.
2. Compute optimal proportions $t^*(\hat{\mu}(lm))$. Check if the stopping criteria (shown above (10)) is met. If not,
3. Compute starvation s_a for each arm as $s_a := (((l+1)m)^{1/2} - N_a(lm))^+$.
4. If $m \geq \sum_a s_a$, generate s_a samples from each arm a. Specifically, first generate s_1 samples from arm 1, then s_2 samples from arm 2 and so on. In addition, generate $\max\{m - \sum_a s_a, 0\}$ independent samples from probability distribution $t^*(\hat{\mu}(lm)) \in \Sigma_K$. For each arm i, count the number of occurrences of i in the generated samples and sample arm i that many times.
5. Else, if $\sum_a s_a > m$, generate $\hat{s}_a$ samples from each arm a, where $\{\hat{s}_a\}_{a=1}^K$ are a solution to the load balancing problem: $\min\left(\max_a \{s_a - \hat{s}_a\}\right)$ s.t. $s_a \geq \hat{s}_a \geq 0\ \forall a \in \{1,\ldots,K\}$, and $\sum_a \hat{s}_a = m$. Again, first generate $\hat{s}_1$ samples from arm 1, then $\hat{s}_2$ samples from arm 2 and so on.

6. Increment l by 1 and return to step 2.

In step 4 above we randomly generate samples from $t^*(\hat{\mu}(lm)) \in \Sigma_K$. Since $t^*(\hat{\mu}(lm)) \in \Sigma_K$ can be seen to be close to $t^*(\mu) \in \Sigma_K$ with high probability, this ensures that under our algorithm, the proportion of samples allocated to each arm are close to $t^*(\mu)$, with small amount of noise due to randomization. Our algorithm differs from Garivier and Kaufmann (2016) where they use a deterministic *C and* D tracking rules. Numerically we observe that the proposed randomized strategy performs similarly to the deterministic C and D tracking rules proposed earlier.

Lemma 8 *Set $m \geq (K+1)^2$. Algorithm $\mathbf{AL_1}$ ensures that $N_a(lm) \geq (lm)^{1/2} - 1$ for all $l \geq 1$.*

When to stop: At any step of the algorithm, the generated data suggests a unique arm, say j, with the largest mean (arbitrarily breaking ties, if any). Call this (μ_j has maximum mean) the null hypothesis, and its complement (arm j does not have maximum mean) the alternate hypothesis. For a stopping rule we consider the generalized likelihood ratio test (see Chernoff (1959)). The numerator in this ratio has value of the likelihood under most likely K-vector of distributions with arm j having the maximum mean, that explains the observed data. The denominator equals the value of likelihood of observed data under most likely distribution of arms under the alternative hypothesis.

In this spirit, at time n, since among all K-vectors of distributions in $(\mathcal{P}(\Re))^K$ with distribution j having maximum mean, $\hat{\mu}(n) = \{\hat{\mu}_1(n), \ldots, \hat{\mu}_K(n)\}$ maximizes the likelihood of the observed data, we take numerator to be the likelihood under $\hat{\mu}(n)$ and the denominator to be that under $\nu \in \mathcal{M}_\mathcal{L}$ that maximizes the likelihood of given data under alternative hypothesis. Our stopping rule corresponds to the logarithm of this 'generalized likelihood ratio' becoming sufficiently large.

Specifically, let $\mathcal{A}_j = \left\{\nu \in \mathcal{M}_\mathcal{L} : m(\nu_j) > \max_{i \neq j} m(\nu_i)\right\}$ denote the set of arms with arm j having the largest mean. Denote $\mathcal{A}_j^c$ to be the set $\left\{\nu \in \mathcal{M}_\mathcal{L} : m(\nu_j) \leq \max_{i \neq j} m(\nu_i)\right\}$. If at stage n, $m(\hat{\mu}_j(n)) > \max_{i \neq j} m(\hat{\mu}_i(n))$, the log (generalized likelihood ratio), $Z_j(n)$, can be seen to equal

$$\inf_{\mu' \in \mathcal{A}_j^c} \sum_{a=1}^{K} N_a(n) \, \mathrm{KL}(\hat{\mu}_a(n), \mu'_a)$$

(see Appendix C.2 for the proof).

Stopping rule: If at stage n, $m(\hat{\mu}_j(n)) > \max_{i \neq j} m(\hat{\mu}_i(n))$, check if $Z_j(n)$ exceeds the threshold function

$$\beta(n, \delta) \triangleq \log\left(\frac{Cn^\alpha}{\delta} (\log n)^K \left(\log \frac{1}{\delta}\right)^{2K+1}\right), \tag{10}$$

where $C > 0$ is specified later in (16), and $\alpha \geq 2K+2$. The algorithm stops if $Z_j(n) \geq \beta(n, \delta)$, announcing arm j as the one having the largest mean. If the threshold function is not exceeded, the algorithm continues.

We prove in Theorem 10 that $\beta(n, \delta)$ given by (10) ensures δ-correctness of $\mathbf{AL_1}$, as well as that the sample complexity matches lower bound asymptotically as $\delta \to 0$ when the batch size $m = o(\log(1/\delta))$. Using arguments as in proof of Theorem 5, it can be shown that if $m(\hat{\mu}_j(n)) > \max_{i \neq j} m(\hat{\mu}_i(n))$, then:

$$Z_j(n) = \min_{b \neq j} \inf_{x \leq y} N_j(n) \, \mathrm{KL_{inf}}(\hat{\mu}_j(n), x) + N_b(n) \, \mathrm{KL_{inf}}(\hat{\mu}_b(n), y), \tag{11}$$

and thus our stopping rule corresponds to evaluating if (11) exceeds $\beta(n, \delta)$.

Remark 9 As mentioned earlier in Remark 6, a mild nuance in our analysis is that while computing $Z_j(n)$, the empirical distribution need not lie in $\mathcal{M}_{\mathcal{L}}$. Also, recall that the stopping condition is checked only after intervals of m, i.e., every time after m samples are generated.

Let τ_δ denote stopping time for the algorithm for a given δ. The algorithm makes an error if at time τ_δ, $m(\hat{\mu}_j(\tau_\delta)) > \max_{i\neq j} m(\hat{\mu}_i(\tau_\delta))$, for some $j \neq 1$. Let $\mathcal{E}$ denote the error event.

Theorem 10 *The algorithm $\mathbf{AL_1}$, with $\beta(n,\delta)$ as in (10), and $m = o(\log(1/\delta))$, is δ-correct, i.e.,*

$$\mathbb{P}(\mathcal{E}) \leq \delta. \tag{12}$$

Further,

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_\mu(\tau_\delta)}{\log(1/\delta)} \leq \frac{1}{V(\mu)}. \tag{13}$$

We first analyse δ-correctness of algorithm $\mathbf{AL_1}$ below. Analysis for the sample complexity is presented towards the end of this section.

The proof of δ-correctness relies on the concentration inequality for $\mathrm{KL_{inf}}(\hat{\kappa}(n), m(\kappa))$, where for $\kappa \in \mathcal{L}$, $\hat{\kappa}(n)$ denotes the empirical distribution corresponding to n samples from κ (Theorem 11). Proof of Theorem 11 in turn relies on the dual representation of $\mathrm{KL_{inf}}(\cdot,\cdot)$ (Theorem 12). These results are proved in Appendix C.4 and C.3 respectively and may also be of independent interest.

Set $c_1 = \mathbb{E}_\kappa(|X - m(\kappa)|)$, $c_2 = \mathbb{E}_\kappa\left(|B - f(|X|)|\right)$, $d_1 = \frac{c_1}{2(f^{-1}(B) - m(\kappa))}$, $d_2 = \frac{c_2}{2(B - f(|m(\kappa)|))}$ and $\tilde{B}_1 \triangleq d_1 + d_2$.

Theorem 11 *For $\kappa \in \mathcal{L}$, and $u \geq 0$,*

$$\mathbb{P}\left(\mathrm{KL_{inf}}(\hat{\kappa}(n), m(\kappa)) \geq u\right) \leq (n+1)^2 e^{\tilde{B}_1} e^{-nu}.$$

Let $\eta \in \mathcal{P}(\Re)$, and $\mathrm{Supp}(\eta)$ denote the support of the measure η. Let

$$\mathcal{R}_2 = \left\{(\lambda_1,\lambda_2) : \lambda_1 \geq 0, \lambda_2 \geq 0, \text{ and } \forall y \in \Re, 1 - (y-x)\lambda_1 - (B - f(|y|))\lambda_2 \geq 0\right\}.$$

Note that for $(\lambda_1,\lambda_2) \in \mathcal{R}_2$, due to strict convexity of f, there can be at most one y_0 such that $1 - (y_0 - x)\lambda_1 - (B - f(|y_0|))\lambda_2 = 0$.

Theorem 12 *For $\eta \in \mathcal{P}(\Re)$ and x such that $x \geq m(\eta)$ and $f(|x|) < B$,*

$$\mathrm{KL_{inf}}(\eta, x) = \max_{(\lambda_1,\lambda_2)\in\mathcal{R}_2} \mathbb{E}_\eta\left(\log\left(1 - (X-x)\lambda_1 - (B - f(|X|))\lambda_2\right)\right). \tag{14}$$

Maximum in RHS above is attained at a unique $(\lambda_1^*, \lambda_2^*)$ in $\mathcal{R}_2$. For $x > m(\eta)$, or $x = m(\eta)$ and $\eta \notin \mathcal{L}$ any probability measure $\kappa^* \in \mathcal{L}$ achieving infimum in the primal problem $\mathcal{O}_1$, satisfies $m(\kappa^*) = x$, $\mathbb{E}_{\kappa^*}(f(|X|)) = B$, and

$$\frac{d\kappa^*}{d\eta}(y) = \frac{1}{1 - (y-x)\lambda_1^* - (B - f(|y|))\lambda_2^*}, \quad \text{for } y \in \mathrm{Supp}(\eta).$$

Furthermore, if $\int_{y\in\mathrm{Supp}(\eta)} d\kappa^*(y) < 1$, then $\mathrm{Supp}(\kappa^*) \setminus \mathrm{Supp}(\eta) = \{y_0\}$, where

$$1 - (y_0 - x)\lambda_1^* - (B - f(|y_0|))\lambda_2^* = 0.$$

In the Appendix C.5.2 we briefly discuss how the algorithm and the analysis simplify if it is known apriori that the underlying distributions have a known bounded support.

Proof of (12) in Theorem 10: Recall that the algorithm makes an error if at the stopping time τ_δ, $m(\hat{\mu}_j(\tau_\delta)) > \max_{i\neq j} m(\hat{\mu}_i(\tau_\delta))$ for some $j \neq 1$. Let the event $\left\{\hat{\mu}(lm) \in \mathcal{A}_j\right\}$ be denoted by $\tilde{\mathcal{E}}_l(j)$. Then the error event $\mathcal{E}$ is contained in the event

$$\left\{\exists l \bigcup_{j\neq 1}\left\{\bigcap_{b\neq j}\inf_{x\leq y} N_j(lm)\,\mathrm{KL}_{\inf}(\hat{\mu}_j(lm),x) + N_b(lm)\,\mathrm{KL}_{\inf}(\hat{\mu}_b(lm),y) \geq \beta(lm,\delta), \tilde{\mathcal{E}}_l(j)\right\}\right\},$$

which is a subset of

$$\left\{\exists l \bigcup_{j\neq 1}\left\{N_j(lm)\,\mathrm{KL}_{\inf}(\hat{\mu}_j(lm),m(\mu_j)) + N_1(lm)\,\mathrm{KL}_{\inf}(\hat{\mu}_1(lm),m(\mu_1)) \geq \beta(lm,\delta), \tilde{\mathcal{E}}_l(j)\right\}\right\}.$$

The above event, and hence the event $\mathcal{E}$, is further contained in

$$\left\{\exists l \sum_{a=1}^{K} N_a(lm)\,\mathrm{KL}_{\inf}(\hat{\mu}_a(lm),m(\mu_a)) \geq \beta(lm,\delta)\right\}. \tag{15}$$

Let $\mathcal{E}_l = \left\{\sum_{a=1}^{K} N_a(lm)\,\mathrm{KL}_{\inf}(\hat{\mu}_a(lm),m(\mu_a)) \geq \beta(lm,\delta)\right\}$. To prove that the probability of making an error is bounded by δ, it is sufficient to prove that $\sum_{l=1}^{\infty}\mathbb{P}\left(\mathcal{E}_l\right)$ is less than δ. For this, we upper bound $\mathbb{P}(\mathcal{E}_l)$ using Theorem 13 below.

Let $c_1^a = \mathbb{E}_{\mu_a}(|X-x|)$, $c_2^a = \mathbb{E}_{\mu_a}\left(|B - f(|X|)|\right)$, $d_1^a = \frac{c_1^a}{2(f^{-1}(B)-m(\mu_a))}$, and $d_2^a = \frac{c_2^a}{2(B-f(|\mu_a|))}$ be non-negative constants and let $\tilde{B}_a \triangleq c_1^a + c_2^a$.

Theorem 13 *For $\mu \in \mathcal{M}_{\mathcal{L}}$, $n \in \mathbb{N}$, and $\Gamma > K+1$,*

$$\mathbb{P}\left(\sum_{a=1}^{K} N_a(n)\,\mathrm{KL}_{\inf}(\hat{\mu}_a(n),m(\mu_a)) \geq \Gamma\right) \leq e^{K+1}\left(\frac{4n^2\Gamma^2\log(n)}{K}\right)^K e^{-\Gamma}\prod_{a=1}^{K} e^{\tilde{B}_a}.$$

Using Theorem 13 with $n = lm$ and $\Gamma = \beta(lm,\delta)$,

$$\mathbb{P}\left(\mathcal{E}\right) \leq \sum_{l=1}^{\infty}\mathbb{P}(\mathcal{E}_l) \leq \sum_{l=1}^{\infty}\left(\frac{4l^2m^2e\log(lm)\beta^2(lm,\delta)}{K}\right)^K e^{-\beta(lm,\delta)+1}\prod_{a=1}^{K} e^{\tilde{B}_a}.$$

Choosing $\alpha \geq 2K+2$ in expression (10) for $\beta(n,\delta)$, ensures that the summation in the above expression is finite. Further, choosing the constant C as:

$$\sum_{l=1}^{\infty}\frac{\left(4e\beta^2(lm,\delta)\right)^K e\prod_{a=1}^{K} e^{\tilde{B}_a}}{K^K (lm)^{\alpha-2K}\left(\log\frac{1}{\delta}\right)^{2K+1}} \leq C, \tag{16}$$

proves that $\sum_{l=1}^{\infty}\mathbb{P}\left(\mathcal{E}_l\right)$, and hence $\mathbb{P}\left(\mathcal{E}\right)$ is bounded from above by δ. ■

We refer the reader to Appendix C.5 for a proof of Theorem 13 and related results.

Proof for (13) in Theorem 10: To see that the sample complexity of $\mathbf{AL_1}$ matches lower bound as $\delta \to 0$, i.e., (13) holds, first observe that our sampling algorithm ensures that fraction of times each arm is pulled is close to its optimal proportion $t_a^*(\mu)$. In particular,

Lemma 14 *For all* $a \in \{1, \ldots, K\}$, $\frac{N_a(lm)}{lm} \xrightarrow{a.s.} t^*_a(\mu)$, *as* $l \to \infty$.

The proof of Lemma 14 is given in Appendix C.6. It uses the fact that eventually all the samples are allocated according to optimal proportions computed for the empirical distribution vector, $\hat{\mu}$, which in turn converges to μ. We first heuristically argue that (13) holds. Let $[K]$ denote the set $\{1, \ldots, K\}$. Recall that for $\beta(lm, \delta)$ defined in (10), the stopping time, τ_δ equals

$$\inf\left\{ lm, \max_{a \in [K]} \min_{b \neq a} \inf_x \left\{ \frac{N_a(lm)}{lm} \mathrm{KL}_{\inf}(\hat{\mu}_a(lm), x) + \frac{N_b(lm)}{lm} \mathrm{KL}_{\inf}(\hat{\mu}_b(lm), x) \right\} \geq \frac{\beta(lm, \delta)}{lm} \right\},$$

and satisfies

$$\max_{a \in [K]} \min_{b \neq a} \inf_x \left\{ \frac{N_a(\tau_\delta)}{\tau_\delta} \mathrm{KL}_{\inf}(\hat{\mu}_a(\tau_\delta), x) + \frac{N_b(\tau_\delta)}{\tau_\delta} \mathrm{KL}_{\inf}(\hat{\mu}_b(\tau_\delta), x) \right\} \approx \frac{\beta(\tau_\delta, \delta)}{\tau_\delta}. \tag{17}$$

Furthermore, for sufficiently large l, with high probability, $\forall a, \;\; \hat{\mu}_a(lm) \approx \mu_a$, and from Lemma 14, $N_a(lm)/lm \approx t^*_a(\mu)$. When this is true, arm 1 is the best arm, and τ_δ satisfies

$$\tau_\delta \approx \beta(\tau_\delta, \delta) \left(\min_{b \neq 1} \inf_x t^*_a(\mu) \, \mathrm{KL}_{\inf}(\mu_a, x) + t^*_b(\mu) \, \mathrm{KL}_{\inf}(\mu_b, x) \right)^{-1} = \frac{\beta(\tau_\delta, \delta)}{V(\mu)}. \tag{18}$$

With constants C and α as in (10), τ_δ that satisfies (18) is given by

$$\tau_\delta = \log\left(C\delta^{-1} \left(\log C\delta^{-1} \right)^\alpha \right) V(\mu)^{-1} + o\left(\log \delta^{-1} \right). \tag{19}$$

Dividing both sides of (19) by $\log(1/\delta)$, we get $\frac{\tau_\delta}{\log(1/\delta)} \approx \frac{1}{V(\mu)}$, for sufficiently small δ.

Complement of this high-probability event contributes only lower order terms (with respect to $\log(1/\delta)$) to $\mathbb{E}_\mu(\tau_\delta)$. Combining these, we get an upper bound on $\mathbb{E}_\mu(\tau_\delta)$ that asymptotically (as $\delta \to 0$) matches the lower bound in (5).

Rigorous proof of the sample complexity result in Theorem 10, i.e., proof for (13), is given in the Appendix C.6. Our proof builds upon that in Garivier and Kaufmann (2016) where the authors consider a restricted SPEF, while we allow arm distributions to belong to a more general class $\mathcal{L}$. Our proof differs in that we work in space of probability measures instead of Euclidian space. This leads to additional nuances. To work in the space of probability measures, we use Wasserstein metric to define continuity of functions and convergence of sequences in this space of probability measures. Furthermore, we check for stopping condition only once in m samples, instead of doing so in every sample, and construct the proof that allows for this flexibility.

5. Optimizing batch sizes and numerical results

We now discuss batch size selection in $\mathbf{AL_1}$ to minimize the overall experiment cost. Suppose that the average cost of generating a sample is given by c_1. This may be large when generating a sample is costly, for instance, if that corresponds to an output of a massive simulation model, or a result of a clinical trial. It may be small, e.g., in an online recommendation system. The cost of solving the max-min problem (7) may be measured by the computational effort involved. The total experiment cost of $\mathbf{AL_1}$ is the sum of the total cost of sampling ($c_1 \times$ number of samples generated) and the total computational effort involved in solving the max-min optimization problems till termination.

Recall that in order to efficiently solve the max-min problem iteratively (see Section 3), at each stage when this evaluation is done, letting $\hat{\mu}_i$ denote the empirical distribution of arm i, we pre-compute values of $\mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_i, y)$ for each y along a grid and linearly interpolate for values of y not in the grid, for each arm i. Empirically we see that the the cost of computing $\mathrm{KL}_{\mathrm{inf}}$ increases linearly with the number of samples of the corresponding empirical distribution (also see Cappé et al. (2013) for similar observations). This suggests that the computational cost of $\mathbf{AL_1}$ increases linearly in the total number of samples generated. To this end, we observe that the overall computational cost of solving the max-min problem (7) is modelled well as $c_{21} + c_{22}n$, where n denotes the total number of samples generated by $\mathbf{AL_1}$ till that stage, and c_{21} and c_{22} are fitted to the data. In our numerical experiment below, this cost is approximately (in computational time) $1854 + 0.6n$ seconds. Since, n runs in many thousands in a typical setting, the linear term cannot be ignored.

To approximate the optimal batch size, we need to approximate the sample complexity. To this end, let $\tilde{\beta}(\delta) \triangleq \log\left(C/\delta\left(\log(C/\delta)\right)^{\alpha}\right)$, where recall that C and α were defined in the stopping rule for $\mathbf{AL_1}$. For small values of δ, the sample complexity of $\mathbf{AL_1}$ (see (78) in Appendix C.6),

$$\mathbb{E}_{\mu}\left(\tau_{\delta}\right) \leq \tilde{\beta}(\delta) V(\mu)^{-1} + m + \text{ lower order terms}, \tag{20}$$

where m denotes the batch size. Equation (20), remains valid if we use $\log(1/\delta)$ in place of $\tilde{\beta}(\delta)$. However, for reasonable values of δ, the two may differ significantly, and empirically we find that $\log(1/\delta)V(\mu)^{-1} + m$ substantially underestimates $\mathbb{E}_{\mu}\left(\tau_{\delta}\right)$, while $\tilde{\beta}(\delta)V(\mu)^{-1} + m$ is much closer. Using $\tilde{\beta}(\delta)V(\mu)^{-1} + m$ as a proxy for $\mathbb{E}_{\mu}\left(\tau_{\delta}\right)$, and assuming that the total number of batches till the stopping time is approximated by $(\tilde{\beta}(\delta)V(\mu)^{-1} + m)m^{-1}$, the total cost $\mathcal{C}$ of $\mathbf{AL_1}$ approximately equals

$$\left(c_1 + c_{21}m^{-1}\right)\left(\tilde{\beta}(\delta)V(\mu)^{-1} + m\right) + 0.5\, c_{22}m^{-1}\left(\tilde{\beta}(\delta)V^{-1}(\mu) + m\right)^2.$$

Observe that for m constant, independent of δ, $\mathcal{C}$ is $\Theta\left(\log^2(1/\delta)\right)$ since $\tilde{\beta}(\delta)$ is $\Theta(\log(1/\delta))$. For $m = \Theta(\log(1/\delta))$, it is $\Theta(\log(1/\delta))$.

Optimizing over m to minimize $\mathcal{C}$, we get

$$m^* = \left(c_{21}\tilde{\beta}(\delta)V(\mu)^{-1} + 0.5\, c_{22}\tilde{\beta}(\delta)^2 V(\mu)^{-2}\right)^{0.5}\left(c_1 + 0.5\, c_{22}\right)^{-0.5}, \tag{21}$$

i.e., $m^* = \Theta\left(\log\left(1/\delta\right)\right)$. Notice that even though $m = m^*$ minimizes $\mathcal{C}$, (20) suggests that with this choice of m, the ratio of expected number of samples until termination for $\mathbf{AL_1}$ to the corresponding max-min lower bound no longer converges to 1 as $\delta \to 0$, that is, (13) no longer holds. It can however be seen that the δ-correct property still holds for $\mathbf{AL_1}$ even for $m = \Theta(\log(1/\delta))$. If, however, $\mathrm{KL}_{\mathrm{inf}}$ could be estimated using computational effort that is independent of the size of the empirical distribution, that is, if $c_{22} = 0$, then $m^* = \Theta\left((\log(1/\delta))^{0.5}\right)$, and $\mathbf{AL_1}$ is asymptotically optimal, so that (13) holds. One way to achieve this may be to approximate the empirical distribution by a fixed size distribution (eg., by bucketing the generated samples into finitely many bins). This may substantially reduce the computation time. The overall issue of developing efficient implementations for the best arm problem for general distributions is an interesting area for further research.

Numerical experiment: We consider a 4-arm bandit seeting. Each arm has a Pareto distribution with pdf $f_{\alpha,\beta}(x) = \frac{\alpha\beta^{\alpha}}{x^{\alpha+1}}$, supported on $[\beta, \infty)$. The parameters (α, β) of these arms are set to

$(4, 1.875)$, $(4, 1.5)$, $(4, 1.25)$, and $(4, 0.75)$. The resulting arm-means are $(2.5, 2.0, 1.67, 1.0)$. $\mathcal{L}$ corresponds to $f(y) = y^2$ and $B = 9$. In the Appendix C.7 we show that the average number of samples needed by $\mathbf{AL_1}$, slowly approaches the lower bound as δ approaches zero (their ratio equals 28 for $\delta = 0.001$ and 16 for $\delta = 10^{-8}$). We also compute the average cost (averaged over 20 independent experiments) of $\mathbf{AL_1}$ measured as c_1 times average sample complexity plus observed average computational effort, as a function of the batch size and c_1, for $\delta = 0.01$. We observe that for $c_1 = 0.0001$ optimal batch size is approximately $30,000$. For $c_1 = 3$ the optimal batch size is close to $4,000$.

Acknowledgments

We acknowledge the support of Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.01-0500. We also thank ICTS, TIFR to allow the authors to collaborate on this work during the Applied Probability Program at ICTS. The first author thanks Pierre Menard for some useful comments on an earlier version of this work.

References

Jean-Yves Audibert and Sébastien Bubeck. Best arm identification in multi-armed bandits. In *COLT-23th Conference on Learning Theory-2010*, pages 13–p, 2010.

Raghu R Bahadur and Leonard J Savage. The nonexistence of certain statistical procedures in nonparametric problems. *The Annals of Mathematical Statistics*, 27(4):1115–1122, 1956.

Robert Eric Bechhofer, Jack Kiefer, and Milton Sobel. *Sequential identification and ranking procedures: with special reference to Koopman-Darmois populations*, volume 3. University of Chicago Press, 1968.

Claude Berge. *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*. Courier Corporation, 1997.

P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley, 2013. ISBN 9781118625965. URL https://books.google.co.in/books?id=6ItqtwaWZZQC.

Patrick Billingsley. *Weak convergence of measures: Applications in probability*, volume 5. Siam, 1971.

Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theor. Comput. Sci.*, 412(19):1832–1852, 2011.

A.N. Burnetas and M.N. Katehakis. Optimal adaptive policies for sequential allocation problems. *Advances in Applied Mathematics*, 17(2):122–142, 1996.

Olivier Cappé, Aurélien Garivier, Odalric-Ambrym Maillard, Rémi Munos, Gilles Stoltz, et al. Kullback–leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, 41(3):1516–1541, 2013.

D Aliprantis Charalambos and Kim C Border. *Infinite dimensional analysis: a hitchhiker's guide*. Springer, 2006.

C.H. Chen, J. Lin, E. Yucesan, and S.E. Chick. Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Discrete Event Dynamic Systems*, 10(3):251–270, 2000.

Herman Chernoff. Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3): 755–770, 1959.

L Dai. Convergence properties of ordinal comparison in the simulation of discrete event dynamic systems. *Journal of Optimization Theory and Applications*, 91(2):363–388, 1996.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(Jun):1079–1105, 2006.

Anthony V Fiacco and Yo Ishizuka. Sensitivity and stability analysis for nonlinear programming. *Annals of Operations Research*, 27(1):215–235, 1990.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027, 2016.

P. W. Glynn and A. Zeevi. Bounding stationary expectations of markov processes. *In Markov processes and related topics: a Festschrift for Thomas G. Kurtz*, pages 195 – 214, 2008.

Peter Glynn and Sandeep Juneja. A large deviations perspective on ordinal optimization. In *Proceedings of the 36th conference on Winter simulation*, pages 577–585. Winter Simulation Conference, 2004.

Peter Glynn and Sandeep Juneja. Selecting the best system, large deviations, and multi-armed bandits. *arXiv preprint arXiv:1507.04564v2*, 2018.

Yu-Chi Ho, R_S Sreenivas, and P Vakili. Ordinal optimization of deds. *Discrete event dynamic systems*, 2(1):61–88, 1992.

Junya Honda and Akimichi Takemura. An asymptotically optimal bandit algorithm for bounded support models. In *In Proceedings of the Twenty-third Conference on Learning Theory (COLT 2010*, pages 67–79. Omnipress, 2010.

Junya Honda and Akimichi Takemura. An asymptotically optimal policy for finite support models in the multiarmed bandit problem. *Machine Learning*, 85(3):361–391, Dec 2011. ISSN 1573-0565. doi: 10.1007/s10994-011-5257-4. URL https://doi.org/10.1007/s10994-011-5257-4.

Junya Honda and Akimichi Takemura. Non-asymptotic analysis of a new bandit algorithm for semi-bounded rewards. *The Journal of Machine Learning Research*, 16(1):3721–3756, 2015.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439, 2014.

Christopher Jennison, Iain M Johnstone, and Bruce W Turnbull. Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. In *Statistical decision theory and related topics III*, pages 55–86. Elsevier, 1982.

Sandeep Juneja and Subhashini Krishnasamy. Sample complexity of partition identification using multi-armed bandits. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1824–1852, Phoenix, USA, 25–28 Jun 2019. PMLR. URL http://proceedings.mlr.press/v99/juneja19a.html.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *ICML*, volume 12, pages 655–662, 2012.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

Seong-Hee Kim and Barry L Nelson. A fully sequential procedure for indifference-zone selection in simulation. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 11(3): 251–273, 2001.

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

Erich Lehmann and Joseph Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.

D.G. Luenberger. *Optimization by Vector Space Methods*. Series in Decision and Control. Wiley, 1969. ISBN 9780471553595. URL https://books.google.co.in/books?id=NAKuQgAACAAJ.

S Magureanu, R Combes, and A Proutiere. Lipschitz bandits: Regret lower bounds and optimal algorithms. In *27th Conference on Learning Theory*, 2014.

Shie Mannor and John N Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5(Jun):623–648, 2004.

Alfred Müller and Dietrich Stoyan. *Comparison methods for stochastic models and risks*, volume 389. Wiley New York, 2002.

Edward Paulson et al. A sequential procedure for selecting the population with the largest mean from k normal populations. *The Annals of Mathematical Statistics*, 35(1):174–180, 1964.

E Posner. Random coding strategies for minimum entropy. *IEEE Transactions on Information Theory*, 21(4):388–391, 1975.

Daniel Russo. Simple bayesian algorithms for best arm identification. In *Conference on Learning Theory*, pages 1417–1418, 2016.

Rangarajan K. Sundaram. *A First Course in Optimization Theory*. Cambridge University Press, 1996. URL https://EconPapers.repec.org/RePEc:cup:cbooks:9780521497701.

C. Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003. ISBN 9780821833124. URL https://books.google.co.in/books?id=GqRXYFxe0l0C.

David Williams. *Probability with martingales*. Cambridge university press, 1991.

Appendix A. Background and proofs related to the impossibility result

We first recall the Berge's Maximum Theorem, which we use in our proofs that follow.

A.1. Berge's Maximum Theorem

Berge's maximum theorem (Sundaram (1996, Chapter 9), Berge (1997, Chapter 6)) provides conditions for (upper and lower) continuity of the optimal value and set of optimizers with respect to the underlying parameters in a constrained optimization problem. We first give some definitions before stating the theorem.

Definition 15 *A correspondence Γ from X to Y (denoted as $\Gamma : X \rightarrow\rightarrow Y$) is a map from X to the set of all non-empty subsets of Y. Γ is a compact (closed) valued correspondence if $\Gamma(x)$ is a compact (closed) subset of Y for each x.*

Definition 16 *For $S \subset X$, define $\Gamma(S) := \bigcup_{x\in S}\Gamma(x)$. The correspondence $\Gamma : X \rightarrow\rightarrow Y$ is upper-hemicontinuous at $x \in X$ if for every open subset $V \subset Y$ such that $\Gamma(x) \subset V$, $\exists\delta > 0$ such that $\Gamma(B_\delta(x)) \subset V$, where $B_\delta(x)$ is a δ neighbourhood of x in X.*

Definition 17 *Correspondence $\Gamma : X \rightarrow\rightarrow Y$ is lower-hemicontinuous at $x \in X$ if for every open subset $V \subset Y$ such that $\Gamma(x) \cap V \neq \emptyset$, $\exists\delta > 0$ such that $\forall z \in B_\delta(x)$, $\Gamma(z) \cap V \neq \emptyset$.*

Definition 18 *Correspondence $\Gamma : X \rightarrow\rightarrow Y$ is continuous at $x \in X$ if it is both upper and lower hemicontinuous at x.*

Theorem 19 (Berge's Maximum Theorem) *Let Θ and X be two metric spaces. Let $\Gamma : \Theta \rightarrow\rightarrow X$ be a compact valued correspondence, $f : X \times \Theta \longrightarrow \Re$ be a real-valued function and*

$$f^*(\theta) = \max_{x\in\Gamma(\theta)} f(x,\theta) \text{ and } \Gamma^*(\theta) = \operatorname*{argmax}_{x\in\Gamma(\theta)} f(x,\theta).$$

- *If f is lower semicontinuous on $X \times \Theta$ and Γ is upper hemicontinuous, then f^* is lower semicontinuous at θ.*
- *If f is a continuous function and Γ is continuous at $\theta \in \Theta$, then f^* is continuous at θ and Γ^* is compact valued and upper-hemicontinuous at θ.*

A.2. Proof of Lemma 1

For notational ease, let probability measure η also denote the associated distribution function. Consider a large y whose value will be fixed later. Furthermore, take $\gamma \in (0, 1)$. Construct another distribution function κ as follows: Set

$$\kappa(x) = (1-\gamma)\eta(x)$$

for all $x \leq y$, and,

$$\bar{\kappa}(x) = \beta\bar{\eta}(x)$$

for $x > y$, where, $\bar{\kappa}(x) = 1 - \kappa(x)$ and $\bar{\eta}(x) = 1 - \eta(x)$.

Since $\kappa(x)$ integrates to 1, β satisfies

$$\beta = 1 + \gamma\frac{\eta(y)}{\bar{\eta}(y)}.$$

Since $\beta > 1$,

$$0 \leq \int_{x\in\Re} \log\left(\frac{d\eta}{d\kappa}(x)\right) d\eta(x) \leq -\eta(y)\log(1-\gamma). \tag{22}$$

By selecting $\gamma = 1 - \exp(-a)$, we get

$$-\eta(y)\log(1-\gamma) \leq a.$$

Also, for y such that $\eta(y^+) = \eta(y^-)$,

$$\begin{aligned} m(\kappa) &= (1-\gamma)\int_{-\infty}^{y} x d\eta(x) + \left(1 + \gamma\frac{\eta(y)}{\bar{\eta}(y)}\right)\int_{y}^{\infty} x d\eta(x) \\ &\geq \exp(-a)m(\eta) + (1-\exp(-a))y. \end{aligned}$$

Since, RHS increases to infinity as $y \to \infty$, one can select y sufficiently large so that $m(\kappa) \geq b$. $\square$

Appendix B. Proofs related to lower bound

B.1. Proof of Lemma 4

Clearly, $\mathcal{L}$ is a uniformly integrable family of measures (see, e.g., Williams (1991)).

Recall that $\mathcal{L}$ is a subset of $\mathcal{P}(\Re)$ and the topology on $\mathcal{L}$ is the subset topology corresponding to that induced by Wasserstein metric on $\mathcal{P}(\Re)$. Furthermore, the topology generated by the Wasserstein metric on $\mathcal{P}(\Re)$ is equivalent to that of the weak convergence of probability measures (see, e.g., Theorem 7.12 in Villani (2003) for the equivalence). We first show that $\mathcal{L}$ is a closed and relatively-compact and thus, compact (in the topology discussed) set of probability measures.

Consider a sequence of measures $\eta_n \in \mathcal{L}$ such that η_n converge weakly to some $\bar{\eta}$ (denoted as $\eta_n \overset{\mathcal{D}}{\Rightarrow} \bar{\eta}$). To show that $\mathcal{L}$ is **closed**, it is sufficient to show that $\bar{\eta} \in \mathcal{L}$, i.e., $\mathbb{E}_{\bar{\eta}}(f(|X|)) \leq B$. Since η_n and $\bar{\eta}$ are measures on $\Re$ and $\eta_n \overset{\mathcal{D}}{\Rightarrow} \bar{\eta}$, there exist random variables X_n and X on $\Re$ such that X_n is distributed according to η_n and X according to $\bar{\eta}$ and $X_n \xrightarrow{a.s.} X$ (by Skorohod's Representation Theorem. See, e.g., Theorem 6.7 Billingsley (2013)). Since $f(|\cdot|)$ is a continuous function, $X_n \xrightarrow{a.s.}$

X implies that $f(|X_n|) \xrightarrow{a.s.} f(|X|)$. Applying Fatou's Lemma (see, e.g., Williams (1991)) to the sequence of non-negative random variables $f(|X_n|)$,

$$\mathbb{E}\left(\liminf_{n\to\infty} f\left(|X_n|\right)\right) \leq \liminf_{n\to\infty} \mathbb{E}\left(f\left(|X_n|\right)\right). \tag{23}$$

Since $\eta_n \in \mathcal{L}$, r.h.s. of (23) is upper bounded by B. This gives $\mathbb{E}_{\bar{\eta}}(f(|X|)) \leq B$, as desired.

Next, to show that $\mathcal{L}$ is a **compact** set under the topology generated by Wasserstein metric, let

$$K_\epsilon = \left[-f^{-1}\left(\frac{B}{\epsilon}\right), f^{-1}\left(\frac{B}{\epsilon}\right)\right],$$

where

$$f^{-1}(x) = \sup\left\{y : f(y) \leq x\right\}.$$

For $\bar{\eta} \in \mathcal{L}$ and $\epsilon > 0$, by Chebyshev's inequality, $1 - \bar{\eta}(K_\epsilon) \leq \epsilon$, and thus, $\mathcal{L}$ is a tight subset of $\mathcal{P}(\Re)$. Furthermore, since $\mathcal{L}$ is closed, by Prokhorov's theorem, $\mathcal{L}$ is a compact set under the topology generated by the Wasserstein metric (see page 25,Villani (2003)).

Thus, $\mathcal{L}$ is a uniformly integrable, and compact collection of probability measures. Now, for $\eta \in \mathcal{P}(\Re)$ and x such that $f(|x|) < B$, we prove the properties of $\text{KL}_{\text{inf}}(\eta, x)$ as a function of η and x. We prove these only for the case when $x \geq m(\eta)$. Exactly the same proofs hold for the case when $x < m(\eta)$.

It is clear from the definition that $\text{KL}_{\text{inf}}(\eta, x)$ is non-decreasing for $x > m(\eta)$ and non-increasing for $x < m(\eta)$. Next, recall that $\text{KL}(\eta, \eta) = 0$ for all $\eta \in \mathcal{P}(\Re)$. Furthermore, $\text{KL}_{\text{inf}}(\eta, x)$ is non-negative for all feasible x and η. In particular, for $\eta \in \mathcal{L}$, η is a feasible solution and $\text{KL}_{\text{inf}}(\eta, m(\eta)) = 0$.

To see the **strict convexity** of $\text{KL}_{\text{inf}}(\eta, x)$ in x, let x_1 and x_2 be such that $f(|x_1|) < B$ and $f(|x_2|) < B$. Let $\eta \in \mathcal{P}(\Re)$. Let κ_1^* and κ_2^* the denote optimal solutions for $\text{KL}_{\text{inf}}(\eta, x_1)$ and $\text{KL}_{\text{inf}}(\eta, x_2)$, respectively (we show their existence in Section C.3). For $\lambda \in [0, 1]$, let

$$\kappa_{12} = \lambda\kappa_1^* + (1-\lambda)\kappa_2^*, \quad \text{and} \quad x_{12} = \lambda x_1 + (1-\lambda)x_2.$$

Then, $m(\kappa_{12}) \geq x_{12}$ and $\mathbb{E}_{\kappa_{12}}(f(|X|)) \leq B$. Since $\kappa_{12} \in \mathcal{L}$ and $\text{KL}(\cdot,\cdot)$ is strictly-convex in the second argument,

$$\text{KL}_{\text{inf}}(\eta, x_{12}) \leq \text{KL}(\eta, \kappa_{12}) < \lambda\,\text{KL}(\eta, \kappa_1^*) + (1-\lambda)\,\text{KL}(\eta, \kappa_2^*). \tag{24}$$

Now, by optimality of κ_1^* and κ_2^* it follows that the r.h.s. of (24) equals $\lambda\,\text{KL}_{\text{inf}}(\eta, x_1) + (1-\lambda)\,\text{KL}_{\text{inf}}(\eta, x_2)$ thus proving strict convexity of $\text{KL}_{\text{inf}}(\eta, x)$ in x. Similarly, by using joint convexity of KL in both the arguments, one can show **convexity** of $\text{KL}_{\text{inf}}(\eta, x)$ in η.

Next, we show that for $\eta \in \mathcal{L}$, $\text{KL}_{\text{inf}}(\eta, x)$ is **bounded** by $\frac{f^{-1}(B)}{f^{-1}(B)-x}$. Consider

$$\kappa := \frac{f^{-1}(B) - x}{f^{-1}(B) - m(\eta)}\eta + \frac{x - m(\eta)}{f^{-1}(B) - m(\eta)}\delta_{f^{-1}(B)},$$

where δ_y denotes a point mass at y. Clearly, $\kappa \in \mathcal{L}$ and $m(\kappa) = x$. Thus,

$$\text{KL}_{\text{inf}}(\eta, x) \leq \text{KL}(\eta, \kappa) \leq \int_{\text{Supp}(\eta)} \ln\left(\frac{f^{-1}(B) - m(\eta)}{f^{-1}(B) - x}\right) d(\eta(y)) \leq \frac{f^{-1}(B)}{f^{-1}(B) - x}.$$

To prove **continuity** of $\text{KL}_{\text{inf}}(\eta, x)$ in η for a fixed x, we show that it is both upper and lower semi continuous function. Let

$$R \triangleq \{\kappa \in \mathcal{L} : \; m(\kappa) \geq x\}.$$

Recall that

$$\text{KL}_{\text{inf}}(\eta, x) = \inf_{\kappa \in R} \text{KL}(\eta, \kappa).$$

Clearly, $\text{KL}(\cdot, \cdot)$ is a lower semicontinuous function from $\mathcal{P}(\Re) \times \mathcal{P}(\Re)$ to $\Re$ (Posner, 1975, Theorem 1) in topology of weak convergence on the domain. Define the correspondence Γ from $\mathcal{P}(\Re)$ to $\mathcal{L}$ as $\Gamma(\eta) = R$ for each η in $\mathcal{P}(\Re)$. Then, Γ is a compact valued correspondence. To see this, consider a sequence of measures $\{\kappa_n\} \in R$ weakly converging to $\kappa \in \mathcal{L}$. Since $\mathcal{L}$ is a uniformly integrable family, $m(\kappa_n) \longrightarrow m(\kappa)$, (Billingsley (1971, Corollary 5, page 9)). From the definition of set R,

$$m(\kappa_n) \geq x \implies m(\kappa) \geq x.$$

Thus, $\kappa \in R$ and hence, R is a closed subset of $\mathcal{L}$. Since $\mathcal{L}$ is a compact set, R is compact. Furthermore, since Γ is a constant correspondence, it can easily be verified to be continuous. Now, from Berge's Theorem it follows that KL_{inf} is lower semicontinuous in η.

Next, we show that $\text{KL}_{\text{inf}}(\eta, x)$ is upper semicontinuous in η. Key idea in proving upper semicontinuity of KL_{inf} in η is showing that it is continuous in the interior of $\mathcal{P}(\Re)$ (see, Charalambos and Border (2006, Theorem 5.43)). Upper semicontinuity then follows since it is a convex function over $\mathcal{P}(\Re)$.

To see continuity in interior of the domain, let $C := \mathcal{P}(\Re)$ and consider $\eta_0 \in \mathcal{L}^o$, where $\mathcal{L}^o$ denotes relative interior of $\mathcal{L}$ in $\mathcal{P}(\Re)$. Furthermore, $\mathcal{L}^o \subset C^o$. Let

$$\mathcal{B}^o(\eta_0, y) = \{\eta \in \mathcal{P}(\Re) : d_W(\eta_0, \eta) < y\}$$

denote an open ball in $\mathcal{P}(\Re)$, which is of radius y and is centered at η_0. Since $\eta_0 \in \mathcal{L}^o$, there exists $\mathcal{B}^o(\eta_0, r) \subset \mathcal{L}^o \subset C^o$. Next, consider $\kappa_1 \in C^o$. Since C^o is convex, there exists $\kappa_2 \in C^o$ such that

$$\kappa_1 = (1 - \lambda)\eta_0 + \lambda \kappa_2, \text{ for some } \lambda \in [0, 1].$$

It can be shown that for all κ in $\mathcal{B}^o(\kappa_1, (1 - \lambda)r)$ the function KL_{inf} is bounded from above by $\max\{M, \text{KL}_{\text{inf}}(\kappa_2, x)\}$, where $M = \frac{f^{-1}(B)}{f^{-1}(B) - x}$. This then gives that KL_{inf} is locally Lipschitz on C^o and hence, continuous. Since KL_{inf} is convex on C and continuous on C^o, it follows that it is upper semicontinuous on C.

From Theorem 2.1 Fiacco and Ishizuka (1990),

$$\frac{\partial \, \text{KL}_{\text{inf}}(\eta, x)}{\partial x} = \lambda_1^*(x), \tag{25}$$

where $\lambda_1^*(x)$ denotes the optimal dual parameter corresponding to the first-moment constraint in the definition of KL_{inf} (see Section C.3 for the dual representation of KL_{inf}).

To prove **double differentiability** of $\text{KL}_{\text{inf}}(\eta, x)$ in x, it is sufficient to prove that $\lambda_1^*(x)$ is differentiable function of x. For $x \neq m(\eta)$, or $x = m(\eta)$ and $\eta \notin \mathcal{L}$, Theorem 12 (discussion and proof in Section C.3) gives that the constraints are tight, i.e., if κ^* denotes the optimal distribution achieving the infimum in $\text{KL}_{\text{inf}}(\eta, x)$, then $m(\kappa^*) = x$ and $\mathbb{E}_{\kappa^*}(f(|X|)) = B$. Furthermore, let

$$\int_{\text{Supp}(\eta)} d\kappa^*(y) = p_0.$$

As shown in the Theorem 12, if $p_0 < 1$ then there is a unique point, say y_0, such that $\text{Supp}(\kappa^*) = \text{Supp}(\eta) \cup \{y_0\}$. Then, using the form of κ^* from Theorem 12 we get that y_0, λ_1^* and λ_2^* solve

$$1 - (y_0 - x)\lambda_1^* - (B - f(|y_0|))\lambda_2^* = 0,$$

$$\int_{\Re} \frac{y d\eta(y)}{1 - (y-x)\lambda_1^* - (B - f(|y|))\lambda_2^*} + y_0(1 - p_0) = x,$$

and

$$\int_{\Re} \frac{f(|y|)\, d\eta(y)}{1 - (y-x)\lambda_1^* - (B - f(|y|))\lambda_2^*} + f(|y_0|)(1 - p_0) = B.$$

Using Implicit function theorem it can be shown that the dual variable λ_1^* is differentiable with respect to the parameter x.

To see that partial derivative of $\text{KL}_{\text{inf}}(\eta, x)$ with respect to x, when $\eta \in \mathcal{L}$ and derivative is evaluated at $m(\eta)$ is 0, from (25), it is sufficient to show that $\lambda_1^*(m(\eta)) = 0$. However, for $\eta \in \mathcal{L}$, $\text{KL}_{\text{inf}}(\eta, m(\eta)) = 0$, with η being the minimizer. Using the form of the minimizer distribution from Section C.3, it follows that $\lambda_1^*(m(\eta)) = 0$. □

B.2. Proof of Theorem 5

To prove Theorem 5, we need a few other results, which we prove first. Let

$$\mathcal{A}_j = \left\{ \nu \in \mathcal{M}_{\mathcal{L}} : m(\nu_j) \geq \max_{i \neq j} m(\nu_i) \right\}.$$

Also, recall that for $\mu \in (\mathcal{P}(\Re))^K$, $g_j(t_1, t_j, x) = t_1 \text{KL}_{\text{inf}}(\mu_1, x) + t_j \text{KL}_{\text{inf}}(\mu_j, x)$ and $G_j(t_1, t_j) = \inf\limits_{x \in [m(\mu_j), m(\mu_1)]} g_j(t_1, t_j, x)$.

Lemma 20 *For t_1, t_2 such that $\max\{t_1, t_2\} > 0$, infimum in expression for $G_j(t_1, t_j)$ is achieved at unique point, $x_j(t_1, t_j)$, and satisfies:*

$$\frac{\partial G_j(t_1, t_j)}{\partial t_1} = \text{KL}_{\text{inf}}(\mu_1, x_j(t_1, t_j)), \quad \frac{\partial G_j(t_1, t_j)}{\partial t_j} = \text{KL}_{\text{inf}}(\mu_j, x_j(t_1, t_j)). \tag{26}$$

Furthermore, for $t \in \Sigma_K$, $\min_j G_j(t_1, t_j)$ is a strictly concave function of t.

Proof We first prove that the infimum over x of $g_j(t_1, t_j, x)$ is attained at a unique x, which satisfies (26). To this end, we first argue that $g_j(t_1, t_j, x)$ is a strictly convex function of x, that has its unique global minima in the region $[m(\mu_j), m(\mu_1)]$. Furthermore, the global minima of a continuous and strictly convex function should satisfy the first order conditions for the global optimality. Using these first order conditions, we arrive at (26).

From Lemma 4, $\text{KL}_{\text{inf}}(\mu_j, x)$ is continuous and strictly convex in x and hence, $g_j(t_1, t_j, x)$ is continuous and strictly convex in x. Furthermore, for $x > m(\mu_1)$ (or for $x < m(\mu_j)$), both $\text{KL}_{\text{inf}}(\mu_1, x)$ and $\text{KL}_{\text{inf}}(\mu_j, x)$ are increasing (or decreasing) functions of x. Thus, $g_j(t_1, t_j, x)$ is minimized for some x in $[m(\mu_j), m(\mu_1)]$, i.e.,

$$\inf_{x \in \Re} g_j(t_1, t_j, x) = \inf_{x \in [m(\mu_j), m(\mu_1)]} g_j(t_1, t_j, x). \tag{27}$$

Since the above expression computes the infimum of a non-negative, continuous, and strictly-convex function over a compact set, there exists a unique $x_j \in [m(\mu_j), m(\mu_1)]$ that minimizes $g_j(t_1, t_j, x)$, giving

$$G_j(t_1, t_j) = t_1 \, \mathrm{KL_{inf}}(\mu_1, x_j(t_1, t_j)) + t_j \, \mathrm{KL_{inf}}(\mu_j, x_j(t_1, t_j)).$$

Furthermore, x_j being point of global minima of g_j, satisfies the following first order condition for optimality:

$$t_1 \frac{\partial \, \mathrm{KL_{inf}}(\mu_1, x_j)}{\partial x} + t_j \frac{\partial \, \mathrm{KL_{inf}}(\mu_j, x_j)}{\partial x} = 0. \tag{28}$$

For $(t_1, t_j) > 0$, by Implicit Function Theorem, x_j is a differentiable function of (t_1, t_j), denoted by $x_j(t_1, t_j)$. Differentiating G_j with respect to t_j,

$$\frac{\partial G_j(t_1, t_j)}{\partial t_j}$$
$$= \mathrm{KL_{inf}}(\mu_j, x_j(t_1, t_j)) + \left(t_j \frac{\partial \, \mathrm{KL_{inf}}(\mu_j, x_j(t_1, t_j))}{\partial x} + t_1 \frac{\partial \, \mathrm{KL_{inf}}(\mu_1, x_j(t_1, t_j))}{\partial x} \right) \frac{\partial x_j(t_1, t_j)}{\partial t_j}.$$

Using (28) in the above expression for the derivative, we get one term in (26). Similarly, differentiating with respect to t_1, one can get the other term in (26).

We now prove strict concavity in t of the function $\min_j G_j(t_1, t_j)$. Consider $t, \tilde{t} \in \Sigma_K$ such that $t \neq \tilde{t}$. For $\lambda \in [0, 1]$ and $\forall j \in \{2, \ldots, K\}$, by linearity of $g_j(t_1, t_j, x)$ in the vector t,

$$G_j(\lambda t_1 + (1-\lambda)\tilde{t_1}, \lambda t_j + (1-\lambda)\tilde{t_j}) = \inf_{x \in [m(\mu_j), m(\mu_1)]} \left\{ \lambda g_j(t_1, t_j, x) + (1-\lambda) g_j(\tilde{t_1}, \tilde{t_j}, x) \right\}, \tag{29}$$

which in turn gives,

$$G_j(\lambda t_1 + (1-\lambda)\tilde{t_1}, \lambda t_j + (1-\lambda)\tilde{t_j}) \geq \lambda G_j(t_1, t_j) + (1-\lambda) G_j(\tilde{t_1}, \tilde{t_j}). \tag{30}$$

To prove strict concavity, it is sufficient to prove that the inequality above is strict for at least one j. Let the unique point of infimum in (29) be x^* and those for $G_j(t_1, t_j)$ and $G_j(\tilde{t}_1, \tilde{t}_j)$ in r.h.s. of (30) be denoted by $x_j(t_1, t_j)$ and $x_j(\tilde{t}_1, \tilde{t}_j)$ respectively.

Suppose $x^* \neq x_j(t_1, t_j)$ or $x^* \neq x_j(\tilde{t}_1, \tilde{t}_j)$, then

$$\begin{aligned} G_j(\lambda t_1 + (1-\lambda)\tilde{t_1}, \lambda t_j + (1-\lambda)\tilde{t_j}) &= \lambda g_j(t_1, t_j, x^*) + (1-\lambda) g_j(\tilde{t}_1, \tilde{t}_j, x^*) \\ &> \lambda G_j(t_1, t_j) + (1-\lambda) G_j(\tilde{t}_1, \tilde{t}_j). \end{aligned}$$

Hence, if (30) holds as an equality for some j, then $x_j(t_1, t_j) = x_j(\tilde{t}_1, \tilde{t}_j) = x^*$ and these must satisfy,

$$\begin{aligned} t_1 \frac{\partial \, \mathrm{KL_{inf}}(\mu_1, x)}{\partial x}\bigg|_{x*} + t_j \frac{\mathrm{KL_{inf}}(\mu_j, x)}{\partial x}\bigg|_{x*} &= \tilde{t_1} \frac{\partial \, \mathrm{KL_{inf}}(\mu_1, x)}{\partial x}\bigg|_{x*} + \tilde{t_j} \frac{\mathrm{KL_{inf}}(\mu_j, x)}{\partial x}\bigg|_{x*} \\ &= 0. \end{aligned}$$

This implies $t_1/t_j = \tilde{t_1}/\tilde{t_j}$. But t and $\tilde{t}$ are both distinct elements of Σ_K. Hence $\exists k \in \{2, \ldots K\}$ such that $t_1/t_k \neq \tilde{t_1}/\tilde{t_k}$ and hence, the corresponding $G_k(t_1, t_k)$ is strictly concave, proving the lemma. ■

B.2.1. PROOF OF (9) IN THEOREM 5

We first prove that,

$$\inf_{\nu \in \mathcal{A}} \left(\sum_{a=1}^{K} t_a \, \mathrm{KL}(\mu_a, \nu_a) \right) = \min_{j \neq 1} \inf_{\nu \in \mathcal{A}_j} \left(\sum_{a=1}^{K} t_a \, \mathrm{KL}(\mu_a, \nu_a) \right) = \min_{j \neq 1} G_j(t_1, t_j). \tag{31}$$

The first equality above is trivial. Notice that infimum of the summation in the expression above in the middle is achieved by some vector $\nu \in \mathcal{A}_j$ such that $\nu_i = \mu_i$, for all $i \notin \{1, j\}$, since for any other $\nu' \in \mathcal{A}_j$ not satisfying this, the value of this expression can be minimized by replacing ν'_i by μ_i for all $i \notin \{1, j\}$. This gives,

$$\begin{aligned} \inf_{\nu \in \mathcal{A}_j} \sum_{a=1}^{K} t_a \, \mathrm{KL}(\mu_a, \nu_a) &= \inf_{\nu \in \mathcal{A}_j : \nu_i = \mu_i, \, i \notin \{1,j\}} t_1 \, \mathrm{KL}(\mu_1, \nu_1) + t_j \, \mathrm{KL}(\mu_j, \nu_j) \\ &= \inf_{x \leq y} \left(t_1 \, \mathrm{KL}_{\mathrm{inf}}(\mu_1, x) + t_j \, \mathrm{KL}_{\mathrm{inf}}(\mu_j, y) \right). \end{aligned}$$

Also, notice that the infimum in the r.h.s. of above equation is attained at a common point ($x = y$). Suppose not, i.e., suppose that the infimum is achieved at $x^* < y^*$. Then, increasing x^* to x' (or decreasing y^* to y', depending on whether $x^* < m(\mu_1)$ or $y^* > m(\mu_j)$) such that $x^* < x' < y^*$ (or $x^* < y' < y^*$) reduces $\mathrm{KL}_{\mathrm{inf}}(\mu_1, x')$ (or $\mathrm{KL}_{\mathrm{inf}}(\mu'_j, y')$) while keeping the other term unchanged (see Lemma 4 for properties of $\mathrm{KL}_{\mathrm{inf}}$), thus reducing the overall value of the function. Thus,

$$\inf_{\nu \in \mathcal{A}_j} \sum_{a=1}^{K} t_a \, \mathrm{KL}(\mu_a, \nu_a) = \inf_{x} \left(t_1 \, \mathrm{KL}_{\mathrm{inf}}(\mu_1, x) + t_j \, \mathrm{KL}_{\mathrm{inf}}(\mu_j, x) \right).$$

Substituting this into (31), we have the following equalities:

$$\inf_{\nu \in \mathcal{A}} \left(\sum_{a=1}^{K} t_a \, \mathrm{KL}(\mu_a, \nu_a) \right) = \min_{j \neq 1} \inf_{x} \left(t_1 \, \mathrm{KL}_{\mathrm{inf}}(\mu_1, x) + t_j \, \mathrm{KL}_{\mathrm{inf}}(\mu_j, x) \right) = \min_{j \neq 1} G_j(t_1, t_j).$$

The last equality above follows from (27). This gives $V(\mu) = \sup_{t \in \Sigma_K} \min_j G_j(t_1, t_j)$.

Furthermore, to prove that the set of maximizers in the expression above is a singleton, notice that $\forall j, G_j(t_1, t_j) \leq \max \left\{ \mathrm{KL}_{\mathrm{inf}}(\mu_1, m(\mu_j)), \mathrm{KL}_{\mathrm{inf}}(\mu_j, m(\mu_1)) \right\}$. Finiteness of $\mathrm{KL}_{\mathrm{inf}}(\mu_1, m(\mu_j))$ follows from the definition of $\mathrm{KL}_{\mathrm{inf}}$, and by considering measure $\mu' = p\mu_1 + (1-p)\delta_{m(\mu_j) - \epsilon}$, for some $\epsilon > 0$ and $p \in (0, 1)$ such that $\mu' \in \mathcal{L}$ and $m(\mu') \leq m(\mu_j)$. Similarly, one can argue the finiteness of $\mathrm{KL}_{\mathrm{inf}}(\mu_j, m(\mu_1))$. Hence, $\min_j G_j(t_1, t_j) < \infty$.

From Lemma 20, for $t \in \Sigma_K$, $\min_j G_j(t_1, t_j)$ is strictly concave function of t. Since Σ_K is a compact set, $\min_j G_j(t_1, t_j)$ attains the maximum at a **unique point** $t^* \in \Sigma_K$. $\square$

B.2.2. PROOF OF CHARACTERIZATION OF $t^*(\mu)$ IN THEOREM 5

For $\mu \in \mathcal{M}_{\mathcal{L}}$ and $t \in \Sigma_K$,

$$h(\mu, t) = \min_{i \in \{2, \cdots, K\}} G_i(t_1, t_i).$$

Let t^* be the unique optimizer in $T(\mu)$. For $i \neq 1$,

$$t_i^* = 0 \implies G_i(t_1^*, t_i^*) = 0, \text{ and } h(\mu, t^*) = 0.$$

Similarly, $h(\mu, t^*) = 0$ if $t_1^* = 0$. But if $t_i = 1/K$ for all i, $h(\mu, t) > 0$ contradicting the optimality of t^*. Hence $t_i^* > 0$ for all i.

Since $\forall i \; t_i^* > 0$, $\partial G_i(t_1, t_i)/\partial t_i$ and $\partial G_i(t_1, t_i)/\partial t_1$ are as given in (26). Furthermore, by Lemma 20, $h(\mu, t)$ is a strictly concave function of t. Hence, first order conditions are necessary and sufficient to find its optimal solution. Re-writing $\max_{t \in \Sigma_K} h(\mu, t)$ as the following optimization problem:

$$\max \; z \quad \text{s.t. } G_j(t_1, t_j) \geq z, \; \forall j; \; \sum_{i=1}^{K} t_i = 1; \;\; t_j \geq 0, \; \forall j. \tag{32}$$

From the first order conditions for (32), it follows that there exist $(\lambda_j : j = 2, \ldots, K)$ and γ satisfying:

$$\sum_{j=2}^{K} \lambda_j = 1, \qquad \lambda_i \left. \frac{\partial G_i(t_1, t_i)}{\partial t_i} \right|_{t=t^*} = \gamma, i = 2, \ldots, K, \tag{33}$$

$$\sum_{i=2}^{K} \lambda_i \left. \frac{\partial G_i(t_1, t_i)}{\partial t_1} \right|_{t=t^*} = \gamma, \qquad \lambda_i (G_i(t_1^*, t_i^*) - z) = 0, \; i = 2, \ldots, K. \tag{34}$$

Equation (33) implies that $\lambda_i > 0$ for some i. Also, since $\partial G_i(t_1^*, t_i^*)/\partial t_i > 0$ for all i, it follows that $\gamma > 0$ and hence each $\lambda_i > 0$.

Part 2 of the Theorem 5 follows from (33) and (34). Part 3 follows from (34).

To show continuity of optimal proportions $t^*(\mu)$ in μ, we use Berge's Theorem (reproduced in Appendix A.1) for the problem in (32), treating μ as a parameter.

Since $\mathrm{KL}_{\inf}(\eta, x)$ is continuous in η (Lemma 4), $g_j(t_1, t_j, x)$ is jointly continuous function of (μ, t). $G_j(t_1, t_j)$ being infimum over a compact set of continuous functions, is jointly continuous in μ and t.

Let the correspondence $\Gamma(\mu) = \Sigma_K$ for all μ. Clearly, Γ is a compact-valued correspondence. Furthermore, since it is independent of μ, it can be easily verified to be both upper and lower hemicontinuous, and hence is continuous in the parameter μ.

Berge's Theorem then gives that the set $T(\mu)$ of optimal solutions to (32) is upper hemicontinuous. However, $T(\mu) = \{t^*(\mu)\}$ being singleton (Theorem 5) and upper hemicontinuous, we conclude that $t^*(\mu)$ is continuous in μ (a correspondence Γ such that $\Gamma(x) = \{\gamma(x)\}$ for some function γ, is upper-hemicontinuous iff $\Gamma(x)$ is lower-hemicontinuous and iff $\gamma(x)$ is continuous).

B.3. Algorithm for solving the max-min lower bound

Recall that solving for the max-min lower bound efficiently is crucial to the performance of the proposed δ-correct algorithm. Theorem 5 characterizes the solution to the max-min lower bound problem, given that μ is known. As discussed earlier in Remark 6, we allow μ to lie not just in $\mathcal{M}_{\mathcal{L}}$, but in $(\mathcal{P}(\Re))^K$.

In this section, we formally describe the algorithm for computing t^* and $V(\mu)$ and prove the monotonicity properties of the relevant equations in the characterization, that are used in the algorithm. These are shown in Lemmas 21, 22, 23, and 24. Algorithm is presented in Section B.3.1.

For fixed $\mu \in (\mathcal{P}(\Re))^K$ such that $m(\mu_1) > \max_{j\geq 2} m(\mu_j)$, define functions $\mathcal{I}_j : \Re \times \Re^+ \to \Re$ and $\tilde{\mathcal{I}}_j : \Re^+ \to \Re^+$ for $j \in \{2, \dots, K\}$ as

$$\mathcal{I}_j(x,y) = \mathrm{KL}_{\mathrm{inf}}(\mu_1, x) + y\,\mathrm{KL}_{\mathrm{inf}}(\mu_j, x) \quad \text{ and } \quad \tilde{\mathcal{I}}_j(y) = \inf_{x\in\Re} \mathcal{I}_j(x,y).$$

As in Lemma 20, with change of variables (set $y_j = \frac{t_j}{t_1}$) and minor modifications, it can be seen that there is a unique x, denoted by $x_j(y)$, that attains the infimum in $\tilde{\mathcal{I}}_j(y)$. Furthermore, $x_j(y)$ belongs to the interval $[m(\mu_j), m(\mu_1)]$ and satisfies

$$\left.\frac{\partial\, \mathrm{KL}_{\mathrm{inf}}(\mu_1, x)}{\partial x}\right|_{x_j(y)} + y \left.\frac{\partial\, \mathrm{KL}_{\mathrm{inf}}(\mu_j, x)}{\partial x}\right|_{x_j(y)} = 0. \tag{35}$$

Below, we prove some monotonicity results, relevant for the algorithm.

Lemma 21 *$\tilde{\mathcal{I}}_j(y)$ is a monotonically strictly increasing function of y.*

Proof Let $x_j(y)$ denote the unique x attaining infimum in $\tilde{\mathcal{I}}_j(y)$. Then,

$$\tilde{\mathcal{I}}_j(y) = \mathrm{KL}_{\mathrm{inf}}(\mu_1, x_j(y)) + y\,\mathrm{KL}_{\mathrm{inf}}(\mu_j, x_j(y)).$$

Differentiating with respect to y and using (35),

$$\begin{aligned}\frac{\partial \tilde{\mathcal{I}}_j(y)}{\partial y} &= \frac{\partial x_j(y)}{\partial y}\left(\left.\frac{\partial\, \mathrm{KL}_{\mathrm{inf}}(\mu_1, x)}{\partial x}\right|_{x_j(y)} + y \left.\frac{\partial\, \mathrm{KL}_{\mathrm{inf}}(\mu_j, x)}{\partial x}\right|_{x_j(y)} \right) + \mathrm{KL}_{\mathrm{inf}}(\mu_j, x_j(y)) \\ &= \mathrm{KL}_{\mathrm{inf}}(\mu_j, x_j(y)) > 0.\end{aligned}$$

■

Clearly, $\tilde{\mathcal{I}}_j(0) = \mathrm{KL}_{\mathrm{inf}}(\mu_1, m(\mu_1))$. Note that this may be non-zero if $\mu_1 \notin \mathcal{L}$. Furthermore, $\tilde{\mathcal{I}}_j(y) \xrightarrow{y\to\infty} d_j$, where

$$d_j = \begin{cases} \mathrm{KL}_{\mathrm{inf}}(\mu_1, m(\mu_j)), & \text{if } \mu_j \in \mathcal{L} \\ \infty, & \text{otherwise.} \end{cases}$$

From Lemma 21, the function $\tilde{\mathcal{I}}_j$ is invertible in $[\mathrm{KL}_{\mathrm{inf}}(\mu_1, m(\mu_1)), d_j)$.

For $c \in [\mathrm{KL}_{\mathrm{inf}}(\mu_1, m(\mu_1)), d_j)$, let $y_j(c)$ denote $\tilde{\mathcal{I}}_j^{-1}(c)$, and let $x_j(c)$ denote the unique x attaining infimum in $\tilde{\mathcal{I}}_j(y_j(c))$.

Lemma 22 *$y_j(c)$ is a monotonically strictly increasing function of c in $[\mathrm{KL}_{\mathrm{inf}}(\mu_1, m(\mu_1)), d_j)$ for $j \in \{2, \dots, K\}$.*

Proof Recall that $x_j(c)$ and $y_j(c)$ satisfy

$$\mathrm{KL}_{\mathrm{inf}}(\mu_1, x_j(c)) + y_j(c)\,\mathrm{KL}_{\mathrm{inf}}(\mu_j, x_j(c)) = c.$$

Differentiating the above equation with respect to c,

$$\left(\frac{\partial\, \mathrm{KL}_{\mathrm{inf}}(\mu_1, x_j(c))}{\partial x} + y_j(c) \frac{\partial\, \mathrm{KL}_{\mathrm{inf}}(\mu_j, x_j(c))}{\partial x} \right) \frac{\partial x_j(c)}{\partial c} + \mathrm{KL}_{\mathrm{inf}}(\mu_j, x_j(c)) \frac{\partial y_j(c)}{\partial c} = 1.$$

Since, $x_j(c)$ attains infimum in $\tilde{\mathcal{I}}_j(y_j(c))$, it satisfies the first order conditions for optimality. Thus,

$$\frac{\partial y_j(c)}{\partial c} = \frac{1}{\mathrm{KL_{inf}}(\mu_j, x_j(c))} > 0.$$

■

Recall that for $x \geq m(\eta)$, $\mathrm{KL_{inf}}(\eta, x)$ is the optimal value of the following constrained optimization problem,

$$\min_{\kappa \in \mathcal{P}(\Re)} \mathrm{KL}(\eta, \kappa) \;\text{ s.t. }\; m(\kappa) \geq x, \quad \mathbb{E}_\kappa(f(|X|)) \leq B.$$

Let $\lambda^*_{1,j}(x)$ denote the optimal dual parameter corresponding to the first-moment constraint in $\mathrm{KL_{inf}}(\mu_j, x)$ (dual formulation of $\mathrm{KL_{inf}}$ and existence of optimal primal and dual variables is argued in Section C.3). Then,

$$\frac{\partial \, \mathrm{KL_{inf}}(\mu_j, x)}{\partial x} = \lambda^*_{1,j}(x), \tag{36}$$

(see, e.g., Fiacco and Ishizuka (1990, Theorem 2.1)). Furthermore, as argued above, $x_j(c)$, lies in the interval $[m(\mu_j), m(\mu_1)]$. Thus, from dual formulation of $\mathrm{KL_{inf}}(\mu_j, x_j(c))$ and $\mathrm{KL_{inf}}(\mu_1, x_j(c))$ (Section C.3),

$$\lambda^*_{1,j}(x_j(c)) \geq 0, \text{ and } \lambda^*_{1,1}(x_j(c)) \leq 0. \tag{37}$$

Lemma 23 *$x_j(c)$ is a monotonically-decreasing function of c in $[\mathrm{KL_{inf}}(\mu_1, m(\mu_1)), d_j)$, for all $j \in \{2, \ldots, K\}$.*

Proof Recall from (35) and (36) that $x_j(c)$ and $y_j(c)$ satisfy,

$$\lambda^*_{1,1}(x_j(c)) + \lambda^*_{1,j}(x_j(c)) y_j(c) = 0.$$

Differentiating with respect to x and re-arranging terms,

$$\frac{\partial x_j(c)}{\partial c} = \frac{-\lambda^*_{1,j}(x_j(c)) \partial y_j(c)/\partial c}{\partial^2 \, \mathrm{KL_{inf}}(\mu_1, x)/\partial x^2 + y_j(c) \partial^2 \, \mathrm{KL_{inf}}(\mu_j, x)/\partial x^2}.$$

Since $\mathrm{KL_{inf}}(\cdot, x)$ is a strictly convex function of x, denominator is clearly positive. Non-positivity of numerator follows from Lemma 22, and (37), thus giving:

$$\frac{\partial x_j(c)}{\partial t_1} \leq 0.$$

■

Lemma 24 $\sum\limits_{j=2}^{K} \frac{\mathrm{KL_{inf}}(\mu_1, x_j(c))}{\mathrm{KL_{inf}}(\mu_j, x_j(c))}$ *is a monotonically-increasing function of c in $[\mathrm{KL_{inf}}(\mu_1, m(\mu_1)), d)$.*

Proof To show the required result, it is sufficient to show that each term in the summation is monotonically-increasing in c. To this end, consider

$$S_j = \frac{\mathrm{KL_{inf}}(\mu_1, x_j(c))}{\mathrm{KL_{inf}}(\mu_j, x_j(c))}.$$

The sign (denoted by Sgn) of derivative of S_j with respect to c equals product of

$$\mathrm{Sgn}\left(\mathrm{KL_{inf}}(\mu_j, x_j(c))\frac{\partial\,\mathrm{KL_{inf}}(\mu_1, x_j(c))}{\partial x} - \mathrm{KL_{inf}}(\mu_1, x_j(c))\frac{\partial\,\mathrm{KL_{inf}}(\mu_j, x_j(c))}{\partial x}\right)$$

and

$$\mathrm{Sgn}\left(\frac{\partial x_j(c)}{\partial c}\right).$$

Recall that $x_j(c)$ lies in the interval $[m(\mu_j), m(\mu_1)]$ and from (36) and (37),

$$\frac{\partial\,\mathrm{KL_{inf}}(\mu_j, x_j(c))}{\partial x} \geq 0 \quad \text{and} \quad \frac{\partial\,\mathrm{KL_{inf}}(\mu_1, x_j(c))}{\partial x} \leq 0.$$

Furthermore, using Lemma 23,

$$\mathrm{Sgn}\left(\frac{\partial S_j}{\partial t_1}\right) = -\,\mathrm{Sgn}\left(\frac{\partial x_j(c)}{\partial c}\right) \geq 0.$$

■

B.3.1. ALGORITHM FOR SOLVING LOWER BOUND OPTIMIZATION PROBLEM

Now we formally describe the algorithm. Let $t^* \in \Sigma_K$ denote the optimal weights vector and let c^* denote the common value of $\tilde{\mathcal{I}}_j(t^*_j/t^*_1)$ for $j \in \{2, \ldots, K\}$. Note that $c^* \in [\mathrm{KL_{inf}}(\mu_1, m(\mu_1)), d_j)$ for all $j \in \{2, \ldots, K\}$. Let $d = \min_j d_j$. Then, $c^* \in [\mathrm{KL_{inf}}(\mu_1, m(\mu_1)), d)$.

1. Fix $c = \mathrm{KL_{inf}}(\mu_1, m(\mu_1))$.

2. For a fixed $c \in [0, d)$, and for each $j \in \{2, \ldots, K\}$, solve the following for $y_j = y_j(c)$ (set $y_1(c) = 1$) and let $x_j(c)$ for each $j \geq 2$ denote the corresponding minimizer:

$$\inf_{x \in [m(\mu_j), m(\mu_1)]} \mathrm{KL_{inf}}(\mu_1, x) + y_j\,\mathrm{KL_{inf}}(\mu_j, x) = c.$$

3. Line search for c^* in the interval $[\mathrm{KL_{inf}}(\mu_1, m(\mu_1)), d)$, so that with $y_j(c^*)$ and the corresponding $x_j(c^*)$ computed using Step 2, the following holds:

$$\sum_{j=2}^{K} \frac{\mathrm{KL_{inf}}(\mu_1, x_j(c^*))}{\mathrm{KL_{inf}}(\mu_j, x_j(c^*))} = 1.$$

4. Output $t^*_j = y_j(c^*)/\sum_{i=1}^{K} y_i(c^*)$ for all $y \in \{1, \ldots, K\}$ and $V(\mu) = c^* t^*_1$.

Appendix C. Proofs related to the sampling algorithm

We first prove that at the end of each interval of length m, say l, the sampling algorithm ensures a minimum ($\sqrt{lm}-1$) number of samples to each arm, i.e., $N_a(lm) \geq \sqrt{lm}-1$.

C.1. Proof of Lemma 8

The given statement is true for $l=1$ as $N_a(m) \geq \frac{m}{K}-1 \geq m^{1/2}-1$. Now suppose that at step lm each arm has at least $(lm)^{1/2}-1$ samples, i.e., $N_a(lm) \geq (lm)^{1/2}-1$ for each arm a. Then, it needs at most $((l+1)m)^{1/2}-(lm)^{1/2}$ samples to ensure that $N_a((l+1)m) \geq ((l+1)m)^{1/2}-1$.

Since $lm \geq K\left((lm)^{1/2}-1\right)$, and $m \geq (K+1)^2$, $m^{1/2}((l+1)^{1/2}-l^{1/2}) < m^{1/2}/l^{1/2} < m/K$, where the first inequality is trivially true. Now, since the maximum number of samples required is an integer, each arm requires at most $\left\lfloor \frac{m}{K} \right\rfloor$ samples and the algorithm has sufficient samples to distribute. This guarantees that all arms reach the minimum threshold.

C.2. Simplification of stopping rule

Recall that $\mathcal{A}_j = \left\{\nu \in \mathcal{M}_{\mathcal{L}} : m(\nu_j) > \max_{i\neq j} m(\nu_i)\right\}$. Let $\boldsymbol{Y^a} := \left(Y^a_i : 1 \leq i \leq N_a(n)\right)$ denote the $N_a(n)$ samples from arm a. For $\nu \in (\mathcal{P}(\Re))^K$, let $L_\nu(\boldsymbol{Y^1}, \dots, \boldsymbol{Y^K})$ denote the likelihood of observing the given samples under ν. Furthermore, if at stage n, $m(\hat{\mu}_j(n)) > \max_{i\neq j} m(\hat{\mu}_i(n))$, then the log of generalized likelihood ratio is given by

$$Z_j(n) = \log\left(\frac{L_{\hat{\mu}(n)}\left(\boldsymbol{Y^1}, \dots, \boldsymbol{Y^K}\right)}{\max\limits_{\mu' \in \mathcal{A}_j^c} L_{\mu'}\left(\boldsymbol{Y^1}, \dots, \boldsymbol{Y^K}\right)}\right).$$

Since each sample is independent of all the other samples,

$$\begin{aligned} Z_j(n) = \log\left(\frac{\prod\limits_{a=1}^{K}\prod\limits_{i=1}^{N_a(n)} \hat{\mu}_a(n)\left(Y^a_i\right)}{\sup\limits_{\mu'\in\mathcal{A}_j^c}\prod\limits_{a=1}^{K}\prod\limits_{i=1}^{N_a(n)} \mu'_a\left(Y^a_i\right)}\right) &= \inf_{\mu'\in\mathcal{A}_j^c} \sum_{a=1}^{K}\sum_{i=1}^{N_a(n)} \left(\log\left(\hat{\mu}_a(n)\left(Y^a_i\right)\right) - \log\left(\mu'_a(Y^a_i)\right)\right) \\ &= \inf_{\mu'\in\mathcal{A}_j^c} \sum_{a=1}^{K} N_a(n)\,\mathrm{KL}(\hat{\mu}_a(n), \mu'_a). \end{aligned}$$

C.3. Dual representation of $\mathrm{KL_{inf}}$

Let $M^+(\Re)$ denote the collection of all non-negative measures on $\Re$. Extend the Kullback-Leibler Divergence to a function on $M^+(\Re) \times M^+(\Re)$, that is, $\mathrm{KL} : M^+(\Re) \times M^+(\Re) \to \Re$ defined as:

$$\mathrm{KL}(\kappa_1, \kappa_2) \triangleq \int_{y\in\Re} \log\left(\frac{d\kappa_1}{d\kappa_2}(y)\right) d\kappa_1(y).$$

Note that for $\kappa_1 \in \mathcal{P}(\Re)$ and $\kappa_2 \in \mathcal{P}(\Re)$, $\mathrm{KL}(\kappa_1, \kappa_2)$ is the usual Kullback-Leibler Divergence between the probability measures.

Recall that for $\eta \in \mathcal{P}(\Re)$, $x \geq m(\eta)$, and $B > f(|x|)$, $\mathrm{KL_{inf}}(\eta, x)$ is defined as the solution to the following optimization problem, $(\mathcal{O}_1)$:

$$\inf_{\kappa\in M^+(\Re)} \mathrm{KL}(\eta,\kappa); \ \text{ s.t. } \int_{y\in\Re} y d\kappa(y) \geq x, \ \int_{y\in\Re} f(|y|)d\kappa(y) \leq B, \ \int_{y\in\Re} d\kappa(y) = 1.$$

We point out that the results that we present below are for $f^{-1}(B) > x \geq m(\eta)$. Symmetric results hold when $-f^{-1}(B) < x \leq m(\eta)$, with $\mathrm{KL_{inf}}$ defined with the corresponding constraints.

Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$. For $\kappa \in M^+(\Re)$, the Lagrangian, denoted by $L(\kappa, \boldsymbol{\lambda})$, for the Problem $(\mathcal{O}_1)$ is given by,

$$\mathrm{KL}(\eta,\kappa) + \lambda_3 \left(1 - \int\limits_{y\in\Re} d\kappa(y)\right) + \lambda_1 \left(x - \int\limits_{y\in\Re} y d\kappa(y)\right) + \lambda_2 \left(\int\limits_{y\in\Re} f(|y|)d\kappa(y) - B\right). \tag{38}$$

Define

$$L(\boldsymbol{\lambda}) \triangleq \inf_{\kappa\in M^+(\Re)} L(\kappa,\boldsymbol{\lambda}). \tag{39}$$

The Lagrangian dual problem corresponding to the Problem $(\mathcal{O}_1)$ is given by the following problem:

$$\max_{\lambda_3\in\Re,\lambda_1\geq 0,\lambda_2\geq 0} \left(\inf_{\kappa\in M^+(\Re)} L(\kappa,\boldsymbol{\lambda})\right). \tag{40}$$

Let $\mathrm{Supp}(\kappa)$ denote the support of measure κ,

$$h(y,\boldsymbol{\lambda}) \triangleq -\lambda_3 - y\lambda_1 + f(|y|)\lambda_2, \qquad \mathcal{Z}(\boldsymbol{\lambda}) = \{y \in \Re : h(y,\boldsymbol{\lambda}) = 0\},$$

and

$$\mathcal{R}_3 = \left\{\boldsymbol{\lambda} \in \Re^3 : \lambda_1 \geq 0, \ \lambda_2 \geq 0, \lambda_3 \in \Re, \inf_{y\in\Re} h(y,\boldsymbol{\lambda}) \geq 0\right\}.$$

Observe that for $\boldsymbol{\lambda} \in \mathcal{R}_3$ $\mathcal{Z}(\boldsymbol{\lambda})$ is either a singleton or an empty set. This is easy to see since $f(\cdot)$ is strictly convex and continuous function. In particular, if $\mathcal{Z}(\boldsymbol{\lambda})$ is non-empty, y_0 that minimizes $h(y,\boldsymbol{\lambda})$ is the unique element in $\mathcal{Z}(\boldsymbol{\lambda})$.

Lemma 25 *The Lagrangian dual problem (40) is simplified as below.*

$$\max_{\lambda_3\in\Re,\lambda_1\geq 0,\lambda_2\geq 0} \left(\inf_{\kappa\in M^+(\Re)} L(\kappa,\boldsymbol{\lambda})\right) = \max_{\boldsymbol{\lambda}\in\mathcal{R}_3} \left(\inf_{\kappa\in M^+(\Re)} L(\kappa,\boldsymbol{\lambda})\right).$$

We call the problem on right as $\boldsymbol{\mathcal{O}_2}$.

Proof Let $\boldsymbol{\lambda} \in \Re^3 \setminus \mathcal{R}_3$. Then, there exists $y_0 \in \Re$ such that $h(y_0,\boldsymbol{\lambda}) < 0$. We show below that for such a $\boldsymbol{\lambda}$, $L(\boldsymbol{\lambda}) = -\infty$, where $L(\boldsymbol{\lambda})$ is defined in (39). Thus, to maximize $L(\boldsymbol{\lambda})$, it is sufficient to consider $\boldsymbol{\lambda} \in \mathcal{R}_3$.

For every $M > 0$, there exists a measure $\kappa_M \in M^+(\Re)$ satisfying $\kappa_M(y_0) = M$ and for $y \in \mathrm{Supp}(\eta) \setminus \{y_0\}$,

$$\frac{d\eta}{d\kappa_M}(y) = 1.$$

Then, (38) can be re-written as:

$$L(\kappa_M, \boldsymbol{\lambda}) = \underbrace{\int\limits_{y\in\Re} \log\left(\frac{d\eta}{d\kappa_M}(y)\right) d\eta(y)}_{\triangleq A} + \underbrace{\int\limits_{y\in\Re} h(y,\boldsymbol{\lambda}) d\kappa_M(y)}_{\triangleq B} + \lambda_3 + \lambda_1 x - \lambda_2 B.$$

From above, it can be easily seen that $L(\kappa_M, \boldsymbol{\lambda}) \xrightarrow{M\to\infty} -\infty$, since $A + B \to -\infty$. Thus, for $\boldsymbol{\lambda} \in \Re^3 \setminus \mathcal{R}_3$, $L(\boldsymbol{\lambda}) = -\infty$ and we get the desired result. ■

Lemma 26 *For $\boldsymbol{\lambda} \in \mathcal{R}_3$, $\kappa^* \in M^+(\Re)$ that minimizes $L(\kappa, \boldsymbol{\lambda})$, satisfies*

$$\text{Supp}(\kappa^*) \subset \{\text{Supp}(\eta) \cup \mathcal{Z}(\boldsymbol{\lambda})\}. \tag{41}$$

Furthermore, for $y \in \text{Supp}(\eta)$, $h(y, \boldsymbol{\lambda}) > 0$, and

$$\frac{d\kappa^*}{d\eta}(y) = \frac{1}{-\lambda_3 - \lambda_1 y + \lambda_2 f(|y|)}. \tag{42}$$

Proof First observe that for $\boldsymbol{\lambda} \in \mathcal{R}_3$, $L(\kappa, \boldsymbol{\lambda})$ is a strictly convex function of κ and that $M^+(\Re)$ is a convex set. Hence, if the minimizer of $L(\kappa, \boldsymbol{\lambda})$ exists, it is unique. Next, we show that any measure, say κ^*, satisfying (41) and (42) minimizes $L(\kappa, \boldsymbol{\lambda})$. This combined with uniqueness of minimizer in $M^+(\Re)$ ensures that any measure satisfying (41) and (42) minimizes $L(\kappa, \boldsymbol{\lambda})$.

Let κ_1 be any measure in $M^+(\Re)$ that is different from κ^*. Since $M^+(\Re)$ is a convex set, for $t \in [0, 1]$, $\kappa_{2,t} \triangleq (1-t)\kappa^* + t\kappa_1$ belongs to $M^+(\Re)$. Since $L(\kappa, \boldsymbol{\lambda})$ is convex in κ, to show that κ^* minimizes $L(\kappa, \boldsymbol{\lambda})$, it suffices to show

$$\left.\frac{\partial L(\kappa_{2,t}, \boldsymbol{\lambda})}{\partial t}\right|_{t=0} \geq 0.$$

Substituting for $\kappa_{2,t}$ in (38),

$$L(\kappa_{2,t}, \boldsymbol{\lambda}) = \int\limits_{y\in\text{Supp}(\eta)} \log\left(\frac{d\eta}{d\kappa_{2,t}}(y)\right) d\eta(y) + (\lambda_3 + \lambda_1 x - \lambda_2 B) + \int\limits_{\Re} h(y,\boldsymbol{\lambda}) d\kappa_{2,t}(y).$$

Evaluating the derivative with respect to t at $t = 0$,

$$\left.\frac{\partial L(\kappa_{2,t}, \boldsymbol{\lambda})}{\partial t}\right|_{t=0} = \int\limits_{y\in\text{Supp}(\eta)} \frac{d\eta}{d\kappa^*}(y)(d\kappa^* - d\kappa_1)(y) + \int\limits_{\Re} (\lambda_3 + \lambda_1 y - \lambda_2 f(|y|))(d\kappa^* - d\kappa_1)(y).$$

For $y \in \text{Supp}(\eta)$, $\partial\eta/\partial\kappa^* = h(y)$. Substituting this in the above expression, we get:

$$\left.\frac{\partial L\left(\kappa_{2,t}, \boldsymbol{\lambda}\right)}{\partial t}\right|_{t=0} = \int_{y\in\text{Supp}(\eta)} h(y)(d\kappa^* - d\kappa_1)(y) - \int_{\Re} h(y)(d\kappa^* - d\kappa_1)(y)$$
$$= -\int_{y\in\{\Re\setminus\text{Supp}(\eta)\}} h(y)d\kappa^*(y) + \int_{y\in\{\Re\setminus\text{Supp}(\eta)\}} h(y)d\kappa_1(y)$$
$$\geq 0.$$

where, for the last inequality, we have used the fact that for $y \in \{\text{Supp}(\kappa^*) \setminus \text{Supp}(\eta)\}$, $h(y) = 0$ and $h(y) \geq 0$, otherwise. ■

Now, we are ready to prove the dual representation of KL_{inf} given in Theorem 12.

C.3.1. PROOF OF THEOREM 12

Let $\mathcal{S}$ denote the rectangle,

$$\mathcal{S} = \left[0, \frac{1}{f^{-1}(B) - x}\right] \times \left[0, \frac{1}{B - f(|x|)}\right],$$

and $\mathcal{R}_2 \subset \mathcal{S}$ denote the region:

$$\mathcal{R}_2 = \left\{(\lambda_1, \lambda_2) : \lambda_1 \geq 0,\ \lambda_2 \geq 0, \inf_{y\in\Re}\left\{1 - (y-x)\lambda_1 - (B - f(|y|))\lambda_2\right\} \geq 0\right\}. \tag{43}$$

Furthermore, define

$$\tilde{h}(y, (\lambda_1, \lambda_2)) \triangleq 1 - (y-x)\lambda_1 - (B - f(|y|))\lambda_2. \tag{44}$$

To prove the alternative expression for KL_{inf} given by this theorem, we first show that both the primal and dual problems ($\mathcal{O}_1$ and $\mathcal{O}_2$, respectively) are feasible. Further, we argue that strong duality holds for the Problem $\mathcal{O}_1$ and show that the expression on the right in (14) is the corresponding optimal Lagrangian dual.

Let δ_y denote a unit mass at point y. Since $f(|x|) < B$, there exists a positive ϵ such that $f(|x+\epsilon|) < B$. Consider $\kappa_0 = \delta_{x+\epsilon}$. Consider distribution κ' which is a convex combination of η and κ_0, given by: $\kappa' = p\kappa_0 + (1-p)\eta$, for $p \in [0,1]$ chosen to satisfy the following two conditions.

$$p(x+\epsilon) + (1-p)m(\eta) \geq x \text{ and } pf(|x+\epsilon|) + (1-p)\mathbb{E}_\eta(f(|X|)) \leq B.$$

It is easy to check that such a p always exists. κ' thus obtained satisfies the constraints of $\mathcal{O}_1$ and $\text{KL}(\eta, \kappa') < \infty$, since $\text{Supp}(\eta) \subset \text{Supp}(\kappa')$. Hence, primal problem $\mathcal{O}_1$ is feasible.

Next, we claim that $\boldsymbol{\lambda^1} = (0, 0, -1)$ is a dual feasible solution. To this end, it is sufficient to show that $\min_{\kappa\in M^+(\Re)} L(\kappa, (0,0,-1)) > -\infty$. Observe that for $\kappa \in \mathcal{M}^+(\Re)$, $\text{KL}(\eta, \kappa)$ defined to extend the usual definition of Kullback-Leibler Divergence to include all measures in $M^+(\Re)$, can be negative with arbitrarily large magnitude. From (38),

$$L(\kappa, \boldsymbol{\lambda^1}) = \text{KL}(\eta, \kappa) - 1 + \int_{y\in\Re} d\kappa(y).$$

Let $\tilde{\kappa}$ denote the minimizer of $L(\kappa, \boldsymbol{\lambda}^1)$. First, observe that $\text{Supp}(\tilde{\kappa}) = \text{Supp}(\eta)$. If there is a y in $\text{Supp}(\eta)$ but outside $\text{Supp}(\tilde{\kappa})$, then $L(\tilde{\kappa}, \boldsymbol{\lambda}^*)$ is ∞. On the other hand, if there exists y in $\{\text{Supp}(\tilde{\kappa}) \setminus \text{Supp}(\eta)\}$, it only contributes to increase the integral in the above expression and thus increases $L(\tilde{\kappa}, \boldsymbol{\lambda}^1)$. Thus, $\text{Supp}(\tilde{\kappa}) = \text{Supp}(\eta)$. Furthermore, from Lemma 26, for y in $\text{Supp}(\eta)$, the optimal measure $\tilde{\kappa}$ must satisfy

$$\frac{d\tilde{\kappa}}{d\eta}(y) = 1.$$

Thus, $\tilde{\kappa} = \eta$ and $\min\limits_{\kappa \in M^+(\Re)} L(\kappa, \boldsymbol{\lambda}^1) = 0$. This proves the feasibility of the dual problem $\mathcal{O}_2$.

Since both primal and dual problems are feasible, both have optimal solutions. Furthermore, $\kappa_0 = \delta_{x+\epsilon}$ defined earlier, satisfies all the inequality constraints of $(\mathcal{O}_1)$ strictly, hence lies in the interior of the feasible region (Slater's conditions are satisfied). Thus strong duality holds for the problem $(\mathcal{O}_1)$ and there exists optimal dual variable $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \lambda_3^*)$ that attains maximum in the problem $\mathcal{O}_2$ (See Theorem 1, Page 224, Luenberger (1969)). Also, since the primal problem is minimization of a strictly-convex function (which is non-negative on the feasible set) with an optimal solution over a closed and convex set (see Lemma 4 for properties if the feasible regioin, $\mathcal{L}$), it attains its infimum within the set.

Strong duality implies

$$\text{KL}_{\text{inf}}(\eta, x) = \max_{\boldsymbol{\lambda} \in \mathcal{R}_3} \inf_{\kappa \in M^+(\Re)} L(\kappa, \boldsymbol{\lambda}).$$

Let κ^* and $\boldsymbol{\lambda}^*$ denote the optimal primal and dual variables. Since strong duality holds, and the problem $(\mathcal{O}_1)$ is a convex optimization problem, KKT conditions are necessary and sufficient for κ^* and $\boldsymbol{\lambda}^*$ to be optimal variables (See page 224, Boyd and Vandenberghe (2004)). Hence $\kappa^*, \lambda_3^* \in \Re, \lambda_1^* \geq 0$, and $\lambda_2^* \geq 0$ must satisfy the following conditions (KKT):

$$\kappa^* \in M^+(\Re), \int\limits_{\Re} y d\kappa^*(y) \geq x, \int\limits_{\Re} f(|y|) d\kappa^*(y) \leq B, \int\limits_{y \in \Re} d\kappa^*(y) = 1, \tag{45}$$

$$\lambda_3^* \left(1 - \int\limits_{\Re} d\kappa^*(y)\right) = 0,\ \lambda_1^* \left(x - \int\limits_{\Re} y d\kappa^*(y)\right) = 0,\ \lambda_2^* \left(\int\limits_{\Re} f(|y|) d\kappa^*(y) - B\right) = 0, \tag{46}$$

and

$$(\lambda_1^*, \lambda_2^*, \lambda_3^*) \in \mathcal{R}_3. \tag{47}$$

Furthermore, κ^* minimizes $L(\kappa, \boldsymbol{\lambda}^*)$. From conditions (46), and Lemma 26,

$$L(\kappa^*, \boldsymbol{\lambda}^*) = \mathbb{E}_\eta \left(\log\left(-\lambda_3^* - \lambda_1^* X + \lambda_2^* f(|X|)\right)\right),$$

where X is the random variable distributed as η.

Adding the equations in (46), and using the form of κ^* from Lemma 26, we get

$$\lambda_3^* = -1 - \lambda_1^* x + \lambda_2^* B.$$

With this condition on λ_3^*, the region $\mathcal{R}_3$ reduces to the region $\mathcal{R}_2$ defined earlier in (43).

Since we know that the optimal $\boldsymbol{\lambda}^*$ in $\mathcal{R}_3$ with the corresponding minimizer, κ^*, satisfies the conditions in (46) and that λ_3^* has the specific form given above, the dual optimal value remains unaffected by adding these conditions as constraints in the dual optimization problem. With these conditions, the dual reduces to

$$\max_{(\lambda_1,\lambda_2)\in\mathcal{R}_2} \mathbb{E}_\eta \left(\log\left(1-(X-x)\lambda_1-(B-f(|X|)\lambda_2)\right)\right),$$

and by strong duality, this is also the value of $\mathrm{KL}_{\inf}(\eta, x)$.

Tightness of the constraint, $m(\kappa^*) = x$: Notice that if η does not have full support, κ^* may have support outside $\mathrm{Supp}(\eta)$. For some $c \geq 0$ and $x > m(\eta)$,

$$\begin{aligned} 1-c &= \mathbb{E}_\eta\left(\left(1-(X-x)\lambda_1^*-(B-f(|X|))\lambda_2^*\right)^{-1}\right) \\ &\geq \left(1-(m(\eta)-x)\lambda_1^*-(B-\mathbb{E}_\eta(f(|X|)))\lambda_2^*\right)^{-1} \\ &\geq \left(1-(m(\eta)-x)\lambda_1^*\right)^{-1}. \end{aligned} \tag{48}$$

In (48) we use Jensen's inequality. Furthermore, if η is a degenerate distribution, then $c > 0$, otherwise (48) is strict inequality as $1/y$ is a strictly convex function of y. Thus,

$$\frac{1}{1-(m(\eta)-x)\lambda_1^*} < 1,$$

and hence, $\lambda_1^* > 0$. Condition (46) then implies $m(\kappa^*) = x$.

Tightness of the constraint, $\mathbb{E}_{\kappa^*}(f(|X|)) = B$: Recall that for $x > m(\eta)$ and $\eta \in \mathcal{P}(\Re)$, $\lambda_1^* > 0$. Also, since $(\lambda_1^*, \lambda_2^*) \in \mathcal{R}_2$, for all y in $\Re$, $\tilde{h}(y,(\lambda_1,\lambda_2)) \geq 0$, where $\tilde{h}(y,(\lambda_1,\lambda_2))$ is defined in (44). However, for $y \to \infty$, $\tilde{h}(y,(\lambda_1,\lambda_2)) < 0$, iff $\lambda_2 = 0$. Thus, $\lambda_2 > 0$ and hence by (46), $\mathbb{E}_{\kappa^*}(f(|X|)) = B$.

C.4. Proofs of concentration result for $\mathrm{KL}_{\inf}$

Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, X_i denote the i^{th} sample from the distribution κ, and recall from (44) that

$$\tilde{h}(X,\boldsymbol{\lambda}) = 1-(X-m(\kappa))\lambda_1-(B-f(|X|))\lambda_2,$$

and define

$$L(\boldsymbol{\lambda}, m(\kappa), \hat{\kappa}(n)) \triangleq \frac{1}{n}\sum_{i=1}^{n}\log\left(\tilde{h}(X_i,\boldsymbol{\lambda})\right).$$

Since $\tilde{h}(y,\boldsymbol{\lambda})$ is a linear function of $\boldsymbol{\lambda}$, and log is a non-decreasing, concave function, $L(\boldsymbol{\lambda}, x, \hat{\kappa}(n))$ is a concave function of $\boldsymbol{\lambda}$ (see, e.g., Page 84, Boyd and Vandenberghe (2004)).

Recall that region $\mathcal{R}_2 \subset \Re^2$ is given by

$$\mathcal{R}_2 = \left\{(\lambda_1,\lambda_2) : \lambda_1 \geq 0,\ \lambda_2 \geq 0,\ \inf_{y\in\Re}\{1-(y-x)\lambda_1-(B-f(|y|))\lambda_2\} \geq 0\right\}, \tag{49}$$

and $\mathcal{R}_2 \subset \mathcal{S}$, where

$$\mathcal{S} = \left[0, \frac{1}{f^{-1}(B)-m(\kappa)}\right] \times \left[0, \frac{1}{B-f(|m(\kappa)|)}\right].$$

Proof of Theorem 11 uses Lemma 27, which we state below.

Lemma 27 *For $\kappa \in \mathcal{L}, \lambda_1 > 0, \lambda_2 > 0, u \geq 0$,*

$$\mathbb{P}\left(L(\boldsymbol{\lambda}, m(\kappa), \hat{\kappa}(n)) \geq u\right) \leq e^{-nu}.$$

Proof Let $X \sim X_i$, for some i. Observe that $\log\left(\tilde{h}(X_i, \boldsymbol{\lambda})\right)$ are i.i.d. For $\gamma \geq 0$, exponentiating and using Markov's inequality,

$$\mathbb{P}\left(L(\boldsymbol{\lambda}, m(\kappa), \hat{\kappa}(n)) \geq u\right) \leq \prod_{i=1}^{n} \mathbb{E}\left(e^{\gamma \log\left(\tilde{h}(X_i, \boldsymbol{\lambda})\right)}\right) e^{-\gamma n u} = \left(\mathbb{E}\left(h(X, \boldsymbol{\lambda})\right)^{\gamma} e^{-\gamma u}\right)^{n}.$$

In particular, the above inequality holds for $\gamma = 1$. Furthermore, by Jensen's inequality,

$$\mathbb{E}(\tilde{h}(X, \boldsymbol{\lambda})) \leq 1.$$

Thus, we have the desired inequality. ■

C.4.1. PROOF OF THEOREM 11

From Theorem 12, if $m(\hat{\kappa}(n)) \leq m(\kappa)$,

$$\mathrm{KL}_{\mathrm{inf}}(\hat{\kappa}(n), m(\kappa)) = \max_{\boldsymbol{\lambda} \in \mathcal{R}_2} L(\boldsymbol{\lambda}, x, \hat{\kappa}(n)),$$

where $\mathcal{R}_2$ is given in (49). In the other case, a symmetric dual representation for $\mathrm{KL}_{\mathrm{inf}}$ holds. We work with only one of these. To get a bound on the probability of maximum of $L(\boldsymbol{\lambda}, m(\kappa), \hat{\kappa}(n))$ over the region $\mathcal{R}_2$, taking values away from 0, we divide the rectangular region, $\mathcal{S}$, into a grid of small rectangles and bound this probability within each rectangle in the grid that intersects with $\mathcal{R}_2$.

To this end, we first describe the grid of $\mathcal{S}$ that we consider. Let $\delta_1 > 0$ and $\delta_2 > 0$ be constants denoting the side lengths of the rectangles in each direction. We will choose their values later. Let

$$M_{\delta_1} = \left\lfloor \frac{1}{\delta_1(f^{-1}(B) - m(\kappa))} \right\rfloor \text{ and } M_{\delta_2} = \left\lfloor \frac{1}{\delta_2(B - f(|m(\kappa)|))} \right\rfloor.$$

Let the rectangle points along λ_1 axis be indexed by l_1 and those along λ_2 axis by l_2, such that $l_1 \in \{0, \ldots, M_{\delta_1}\}$ and $l_2 \in \{0, \ldots, M_{\delta_2}\}$. For $l_1 \in \{0, \ldots, M_{\delta_1}\}$ and $l_2 \in \{0, \ldots, M_{\delta_2}\}$, $\lambda_{1,l_1} \triangleq l_1\delta_1$ and $\lambda_{2,l_2} \triangleq l_2\delta_2$, and

$$\lambda_{1, M_{\delta_1}+1} \triangleq \frac{1}{f^{-1}(B) - m(\kappa)} \quad \text{and} \quad \lambda_{2, M_{\delta_2}+1} \triangleq \frac{1}{B - f(|m(\kappa)|)}.$$

Denote by G_{l_1,l_2} the rectangle $[\lambda_{1,l_1}, \lambda_{1,l_1+1}] \times [\lambda_{2,l_2}, \lambda_{2,l_2+1}]$. Then, using union bound,

$$\begin{aligned}\mathbb{P}\left(\max_{\boldsymbol{\lambda} \in \mathcal{R}_2} L(\boldsymbol{\lambda}, m(\kappa), \hat{\kappa}(n)) \geq u\right) &= \mathbb{P}\left(\bigcup_{l_2=0}^{M_{\delta_2}} \bigcup_{l_1=0}^{M_{\delta_1}} \left\{\max_{\boldsymbol{\lambda} \in G_{l_1,l_2} \cap \mathcal{R}_2} L(\boldsymbol{\lambda}, m(\kappa), \hat{\kappa}(n)) \geq u\right\}\right) \\ &\leq \sum_{l_1=0}^{M_{\delta_1}} \sum_{l_2=0}^{M_{\delta_2}} \mathbb{P}\left\{\max_{\boldsymbol{\lambda} \in G_{l_1,l_2} \cap \mathcal{R}_2} L(\boldsymbol{\lambda}, m(\kappa), \hat{\kappa}(n)) \geq u\right\}. \end{aligned} \quad (50)$$

Let us now focus on the summand of the above expression. Note that if the rectangle G_{l_1,l_2} does not intersect with the region $\mathcal{R}_2$, then its contribution to the summation in the r.h.s. above is 0. Thus, we only consider the rectangles that have a non-trivial intersection with $\mathcal{R}_2$. Using Markov's Inequality,

$$\mathbb{P}\left\{\max_{\boldsymbol{\lambda}\in G_{l_1,l_2}\cap\mathcal{R}_2} L(\boldsymbol{\lambda}, m(\kappa), \hat{\kappa}(n)) \geq u\right\}$$
$$\leq \mathbb{E}_\kappa\left(\exp\left\{\max_{\boldsymbol{\lambda}\in G_{l_1,l_2}} \sum_{i=1}^n \log\left(1-(X_i-m(\kappa))\lambda_1-(B-f(|X_i|))\lambda_2\right)\right\}\right)e^{-nu}. \tag{51}$$

Observe that for G_{l_1,l_2},

$$\max_{\boldsymbol{\lambda}\in G_{l_1,l_2}} \log\left(1-(X_i-m(\kappa))\lambda_1-(B-f(|X|))\lambda_2\right)$$
$$\leq \log\left(1-(X_i-m(\kappa))\lambda_{1,l_1}-(B-f(|X|))\lambda_{2,l_2}+|X_i-m(\kappa)|\,\delta_1+|B-f(|X_i|)|\,\delta_2\right) \tag{52}$$

Using this in (51), $\mathbb{P}\left\{\max_{\boldsymbol{\lambda}\in G_{l_1,l_2}\cap\mathcal{R}_2} L(\boldsymbol{\lambda}, m(\kappa), \hat{\kappa}(n)) \geq u\right\}$ can be bounded from above by e^{-nu} times

$$\mathbb{E}_\kappa\left(\prod_{i=1}^n\left(1-(X_i-m(\kappa))\lambda_{1,l_1}-(B-f(|X_i|))\lambda_{2,l_2}+|X_i-m(\kappa)|\,\delta_1+|B-f(|X_i|)|\,\delta_2\right)\right).$$

The expectation above satisfies the following:

$$\mathbb{E}_\kappa\left(\prod_{i=1}^n\left(1-(X_i-m(\kappa))\lambda_{1,l_1}-(B-f(|X_i|))\lambda_{2,l_2}+|X_i-m(\kappa)|\,\delta_1+|B-f(|X_i|)|\,\delta_2\right)\right)$$
$$=\prod_{i=1}^n\mathbb{E}_\kappa\left(1-(X_i-m(\kappa))\lambda_{1,l_1}-(B-f(|X_i|))\lambda_{2,l_2}+|X_i-m(\kappa)|\,\delta_1+|B-f(|X_i|)|\,\delta_2\right)$$
$$\leq\left(1+d_1\delta_1+d_2\delta_2\right)^n,$$

where $d_1=\mathbb{E}_\kappa(|X-m(\kappa)|)$ and $d_2=\mathbb{E}_\kappa(|B-f(|X|)|)$.

Set $\delta_1=\frac{1}{n(f^{-1}(B)-m(\kappa))}$ and $\delta_2=\frac{1}{n(B-f(|m(\kappa)|))}$. Let $c_1=\frac{d_1}{(f^{-1}(B)-m(\kappa))}$ and $c_2=\frac{d_2}{(B-f(|m(\kappa)|))}$ and $c_1+c_2=\tilde{B}_1$. Then,

$$\mathbb{P}\left\{\max_{\boldsymbol{\lambda}\in G_{l_1,l_2}\cap\mathcal{R}_2} L(\boldsymbol{\lambda}, m(\kappa), \hat{\kappa}(n)) \geq u\right\}\leq\left(1+\frac{c_1}{n}+\frac{c_2}{n}\right)^n e^{-nu}\leq e^{\tilde{B}_1}e^{-nu}.$$

Furthermore, from the choice of δ_1 and δ_2, $M_{\delta_1}\leq n+1,$ and $M_{\delta_2}\leq n+1$. Substituting these and the above inequality back into (50) gives:

$$\mathbb{P}\left(\max_{\boldsymbol{\lambda}\in\mathcal{R}_2} L(\boldsymbol{\lambda}, x, \hat{\kappa}(n))\geq u\right)\leq(n+1)^2 e^{\tilde{B}_1}e^{-nu}.$$

C.5. Proofs related to δ-correctness

Recall that to prove δ-correctness of the proposed algorithm, it is sufficient to prove Theorem 13, i.e., for $\mu \in \mathcal{M}_{\mathcal{L}}$, $n \in \mathbb{N}$, and $\Gamma > K+1$,

$$\mathbb{P}\left(\sum_{a=1}^{K} N_a(n)\,\mathrm{KL}_{\inf}(\hat{\mu}_a(n), m(\mu_a)) \geq \Gamma\right) \leq e^{K+1}\left(\frac{4n^2\Gamma^2\log(n)}{K}\right)^{K} e^{-\Gamma}\prod_{a=1}^{K} e^{\tilde{B}_a},$$

where $N_a(n)$ is the number of times arm a has been pulled in n trials, $\hat{\mu}_a(n)$ is the empirical distribution corresponding to $N_a(n)$ samples from arm a (distribution μ_a), $m(\mu_a)$ denotes the mean of the distribution μ_a, and $\tilde{B}_a$ is a constant corresponding to arm a.

Some notation is needed to this end. For $\tilde{e} > 0$, let $D = \lceil \log(n)/\log(1+\tilde{e})\rceil$ and $\mathcal{D} = \left\{d \in \mathbb{N}^K \ s.t., \forall a \ \ 1 \leq d_a \leq D\right\}$. For $d \in \mathcal{D}$, let

$$\mathcal{C}^d = \bigcap_{a=1}^{K}\left\{(1+\tilde{e})^{d_a-1} \leq N_a(n) \leq (1+\tilde{e})^{d_a}\right\} = \bigcap_{a=1}^{K}\mathcal{C}^d_a,$$

where $\mathcal{C}^d_a = \left\{(1+\tilde{e})^{d_a-1} \leq N_a(n) \leq (1+\tilde{e})^{d_a}\right\}$. Let $\underline{t}_a = (1+\tilde{e})^{d_a-1}$ and $\bar{t}_a = (1+\tilde{e})^{d_a}$. Furthermore, for each arm a, let $\mathcal{S}_a$ denote the following rectangle:

$$\mathcal{S}_a = \left[0, \frac{1}{f^{-1}(B) - m(\mu_a)}\right] \times \left[0, \frac{1}{B - f(|m(\mu_a)|)}\right],$$

and let

$$\mathcal{R}^a_2 = \left\{(\lambda_1,\lambda_2) : \lambda_1 \geq 0,\ \lambda_2 \geq 0, \inf_{y\in\Re}\{1-(y-m(\mu_a))\lambda_1 - (B - f(|y|))\lambda_2\} \geq 0\right\}. \quad (53)$$

On set $\mathcal{C}^d_a$, let there be a grid of the rectangular region $\mathcal{S}_a$, for each a, similar to that in the proof of Theorem 11, with

$$\delta^a_1 = \frac{1}{\underline{t}_a\left(f^{-1}(B) - m(\mu_a)\right)} \qquad \text{and} \qquad \delta^a_2 = \frac{1}{\underline{t}_a\left(B - f(|m(\mu_a)|)\right)}$$

being the side lengths of each rectangle in the grid. Let a typical rectangle that intersects with the region $\mathcal{R}^a_2$, defined in (53), be denoted by G_a. Recall from Theorem 12 that to bound the probability of empirical $\mathrm{KL}_{\inf}$ taking large values, it is sufficient to consider only such rectangles since, the optimal dual parameters, $(\lambda^{a*}_1, \lambda^{a*}_2)$ lie in $\mathcal{R}^a_2$. Henceforth, in our discussion, we consider only such rectangles.

Lemma 28 *For any $u_a \in \Re$, non-negative constants $\tilde{B}_a$ and rectangle G_a,*

$$\mathbb{P}\left(\bigcap_{a=1}^{K}\left\{\max_{\boldsymbol{\lambda_a}\in G_a} L(\boldsymbol{\lambda_a}, m(\mu_a), \hat{\mu}_a(n)) \geq u_a, \mathcal{C}^d_a\right\}\right) \leq \prod_{a=1}^{K} e^{\tilde{B}_a} e^{-\underline{t}_a u_a}.$$

Proof Recall that G_a is a rectangle that intersects the region $\mathcal{R}^a_2$. Let $\boldsymbol{\lambda_{a0}} = (\lambda^a_{10}, \lambda^a_{20})$ denote one of the corner points of the rectangle G_a such that $\lambda^a_{10} > 0$, $\lambda^a_{20} > 0$. Let

$$\Lambda_a(\theta, \boldsymbol{\lambda_{a0}}) = \log \mathbb{E}_{\mu_a}\left(e^{\left\{\theta\log\left(1-(X-m(\mu_a))\lambda^a_{10} - (B-f(|X_i|))\lambda^a_{20} + |X-m(\mu_a)|\delta^a_1 + |B-f(|X|)|\delta^a_2\right)\right\}}\right),$$

and

$$\theta_a = \underset{\theta \geq 0}{\operatorname{argmax}} \left\{ \theta u_a - \Lambda_a(\theta, \boldsymbol{\lambda_{a0}}) \right\}.$$

Clearly, $\theta_a \geq 0$ and $\{\theta_a u_a - \Lambda_a(\theta_a, \boldsymbol{\lambda_{a0}})\} \geq 0$. Let $S^a(n, \boldsymbol{\lambda_{a0}})$ denote the sum below:

$$\sum_{i=1}^{N_a(n)} \log\left(1 - (X_i - m(\mu_a))\lambda_{10}^a - (B - f(|X_i|))\lambda_{20}^a + |X - x|\, \delta_1^a + |B - f(|X|)|\, \delta_2^a\right).$$

Observe from (52) that $\max\limits_{\boldsymbol{\lambda_a} \in G_a} N_a(n) L(\boldsymbol{\lambda_a}, m(\mu_a), \hat{\mu}_a) \leq S^a(n, \boldsymbol{\lambda_{a0}})$. Thus,

$$\mathbb{P}\left(\bigcap_{a=1}^{K} \left\{ \max_{\boldsymbol{\lambda_a} \in G_a} L(\boldsymbol{\lambda_a}, m(\mu_a), \hat{\mu}_a(n)) \geq u_a,\ \mathcal{C}_a^d \right\} \right)$$
$$\leq \mathbb{P}\left(\bigcap_{a=1}^{K} \left\{ S^a(n, \boldsymbol{\lambda_{a0}}) \geq N_a(n) u_a,\ \mathcal{C}_a^d \right\} \right) \leq \mathbb{P}\left(\mathbb{1}_{\mathcal{C}^d} e^{\left\{ \sum\limits_{a=1}^{K} \theta_a S_a(n, \boldsymbol{\lambda_{a0}}) \right\}} \geq e^{\left\{ \sum\limits_{a=1}^{K} \theta_a N_a(n) u_a \right\}} \right).$$

Multiplying by $\exp\{-N_a(n)\Lambda_a(\theta_a, \boldsymbol{\lambda_{a0}})\}$ on both sides of the inequality in the above expression, the probability of intersection can be upper bounded by:

$$\mathbb{P}\left(\mathbb{1}_{\mathcal{C}^d} e^{\left\{ \sum\limits_{a=1}^{K} \theta_a S_a(n, \boldsymbol{\lambda_{a0}}) - N_a(n)\Lambda_a(\theta_a, \boldsymbol{\lambda_{a0}}) \right\}} \geq e^{\left\{ \sum\limits_{a=1}^{K} N_a(n)(\theta_a u_a - \Lambda_a(\theta_a, \boldsymbol{\lambda_{a0}})) \right\}} \right). \tag{54}$$

Since $\{\theta_a u_a - \Lambda_a(\theta_a, \boldsymbol{\lambda_{a0}})\}$ is non-negative (by choice of θ_a), on set $\mathcal{C}_a^d$,

$$N_a(n)\left\{\theta_a u_a - \Lambda_a(\theta_a, \boldsymbol{\lambda_{a0}})\right\} \geq \underline{t_a}\left\{\theta_a u_a - \Lambda_a(\theta_a, \boldsymbol{\lambda_{a0}})\right\}.$$

Furthermore, let

$$G(n) = \exp\left\{ \sum_{a=1}^{K} \theta_a S_a(n, \boldsymbol{\lambda_{a0}}) - N_a(n)\Lambda_a(\theta_a, \boldsymbol{\lambda_{a0}}) \right\}.$$

Using these substitutions with Markov's inequality in (54),

$$\mathbb{P}\left(\bigcap_{a=1}^{K} \left\{ \max_{\boldsymbol{\lambda_a} \in G_a} L(\boldsymbol{\lambda_a}, m(\mu_a), \hat{\mu}_a(n)) \geq u_a,\ \mathcal{C}_a^d \right\} \right) \leq \mathbb{E}\left(\mathbb{1}_{\mathcal{C}^d} G(n)\right) e^{\left\{ -\sum\limits_{a=1}^{K} \underline{t_a}(\theta_a u_a - \Lambda_a(\theta_a, \boldsymbol{\lambda_{a0}})) \right\}}.$$

Since $G(n)$ is a mean-1 martingale, $\mathbb{E}(\mathbb{1}_{\mathcal{C}^d} G_n) \leq 1$. Using definition of θ_a along with this, we get the following:

$$\mathbb{P}\left(\bigcap_{a=1}^{K} \left\{ \max_{\boldsymbol{\lambda_a} \in G_a} L(\boldsymbol{\lambda_a}, m(\mu_a), \hat{\mu}_a(n)) \geq u_a,\ \mathcal{C}_a^d \right\} \right) \leq \prod_{a=1}^{K} e^{\left\{ -\underline{t_a} \sup\limits_{\theta \geq 0} (\theta u_a - \Lambda_a(\theta, \boldsymbol{\lambda_{a0}})) \right\}}. \tag{55}$$

Notice that $\Lambda_a(1, \boldsymbol{\lambda_{a0}})$ equals

$$\log\left(\mathbb{E}_{\mu_a}\left(1 - (X - m(\mu_a))\,\lambda_{10}^a - (B - f(|X|))\lambda_{20}^a + |X - m(\mu_a)|\, \delta_1^a + |B - f(|X|)|\, \delta_2^a\right),\right.$$

which is upper bounded by

$$\log\left(1+\frac{c_1^a}{\underline{t}_a}+\frac{c_2^a}{\underline{t}_a}\right),$$

where $c_1^a = \underline{t}_a\delta_1^a \mathbb{E}_{\mu_a}\left(|X-m(\mu_a)|\right)$ and $c_2^a = \underline{t}_a\delta_2^a \mathbb{E}_{\mu_a}\left(|B-f(|X|)|\right)$. Let $\tilde{B}_a = c_1^a + c_2^a$. Substituting this back into (55) and choosing $\theta = 1$ for all a, we get the following desired upper bound:

$$\mathbb{P}\left(\bigcap_{a=1}^{K}\left\{\max_{\boldsymbol{\lambda_1}\in G_a} L(\boldsymbol{\lambda_a}, m(\mu_a), \hat{\mu}_a(n)) \geq u_a,\ \mathcal{C}_a^d\right\}\right) \leq \prod_{a=1}^{K} e^{\tilde{B}_a} e^{-\underline{t_a} u_a}. \tag{56}$$

■

Using the above result, we prove following inequality, which will assist in the proof of Theorem 13.

Lemma 29 *Let $\tilde{e} > 0$. For $\zeta_a \geq 0$,*

$$\mathbb{P}\left(\bigcap_{a=1}^{K}\left\{N_a(n)\,\mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \zeta_a, \mathcal{C}_a^d\right\}\right) \leq \prod_{a=1}^{K}\left\{(n+1)^2 e^{\tilde{B}_a} e^{-\zeta_a/(1+\tilde{e})}\right\}.$$

Proof On set $\mathcal{C}_a^d$,

$$\mathbb{P}\left(\bigcap_{a=1}^{K}\left\{N_a(n)\,\mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \zeta_a, \mathcal{C}_a^d\right\}\right)$$
$$\leq \mathbb{P}\left(\bigcap_{a=1}^{K}\left\{\mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \frac{\zeta_a}{\bar{t}_a}, \mathcal{C}_a^d\right\}\right).$$

Recall that corresponding to $\mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a))$,

$$\mathcal{S}_a = \left[0, \frac{1}{f^{-1}(B)-m(\mu_a)}\right] \times \left[0, \frac{1}{B-f(|m(\mu_a)|)}\right],$$

and

$$\mathcal{R}_2^a = \left\{(\lambda_1, \lambda_2) : \lambda_1 \geq 0,\ \lambda_2 \geq 0, \inf_{y\in\Re}\left\{1-(y-m(\mu_a))\lambda_1 - (B-f(|y|))\lambda_2\right\} \geq 0\right\}.$$

Let $G_{l_1^a, l_2^a}$ denote a rectangle in $\mathcal{S}_a$ that also intersects with the region $\mathcal{R}_2$, and is given by

$$G_{l_1^a, l_2^a} = \left[\lambda_{1,l_1^a}, \lambda_{1,l_1^a+1}\right] \times \left[\lambda_{2,l_2^a}, \lambda_{2,l_2^a+1}\right],$$

as in Theorem 11. Then,

$$\mathbb{P}\left(\bigcap_{a=1}^{K}\left\{\mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \frac{\zeta_a}{\bar{t}_a}, \mathcal{C}_a^d\right\}\right)$$
$$= \mathbb{P}\left(\bigcap_{a=1}^{K}\left\{\bigcup_{l_1^a}\bigcup_{l_2^a}\max_{\substack{\boldsymbol{\lambda_a}\in\\ G_{l_1^a, l_2^a}}} L(\boldsymbol{\lambda_a}, m(\mu_a), \hat{\mu}_a(n)) \geq \frac{\zeta_a}{\bar{t}_a}, \mathcal{C}_a^d\right\}\right).$$

Using union bound,

$$\mathbb{P}\left(\bigcap_{a=1}^{K}\left\{\mathrm{KL_{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \frac{\zeta_a}{\bar{t}_a}, \mathcal{C}_a^d\right\}\right)$$
$$\leq \sum_{l_1^1}\sum_{l_2^1}\cdots\sum_{l_1^K}\sum_{l_2^K}\mathbb{P}\left(\bigcap_{a=1}^{K}\left\{\max_{\boldsymbol{\lambda_a}\in G_{l_1^a,l_2^a}} L(\boldsymbol{\lambda_a}, m(\mu_a), \hat{\mu}_a(n)) \geq \underbrace{\frac{\zeta_a}{\bar{t}_a}}_{:=u_a},\ \mathcal{C}_a^d\right\}\right).$$

Recall that $\underline{t}_a = (1+\tilde{e})^{d_a-1}$ and $\bar{t}_a = (1+\tilde{e})^{d_a}$. Using Lemma 28, we upper bound the summand in the above inequality, to get

$$\mathbb{P}\left(\bigcap_{a=1}^{K}\left\{\mathrm{KL_{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \frac{\zeta_a}{\bar{t}_a}, \mathcal{C}_a^d\right\}\right) \leq \sum_{l_1^1}\sum_{l_2^1}\cdots\sum_{l_1^K}\sum_{l_2^K}\left(\prod_{a=1}^{K} e^{\tilde{B}_a} e^{-\underline{t}_a u_a}\right). \tag{57}$$

Furthermore, choosing $\delta_1^a = \frac{1}{\underline{t}_a(f^{-1}(B)-x)}$ and $\delta_2^a = \frac{1}{\underline{t}_a(B-f(|x|))}$ gives $\underline{t}_a u_a = \frac{\zeta_a}{1+\tilde{e}}$. Substituting in inequality (57), we get the following desired inequality:

$$\mathbb{P}\left(\bigcap_{a=1}^{K}\left\{N_a(n)\,\mathrm{KL_{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \zeta_a, \mathcal{C}_a^d\right\}\right) \leq \prod_{a=1}^{K}\left\{(n+1)^2 e^{\tilde{B}_a} e^{-\frac{\zeta_a}{1+\tilde{e}}}\right\}.$$

■

Lemma 30

$$\mathbb{P}_\mu\left(\mathbb{1}_{\mathcal{C}^d}\sum_{a=1}^{K} N_a(n)\,\mathrm{KL_{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \Gamma\right) \leq e^{-\frac{\Gamma}{(1+\tilde{e})}}\left(\frac{e\Gamma}{K(1+\tilde{e})}\right)^K \prod_{a=1}^{K}\left\{(n+1)^2 e^{\tilde{B}_a}\right\}.$$

Proof Proof of the Lemma depends on the concentration result of Theorem 11 and a careful use of martingales. Let $Y = \{Y_1, Y_2, \ldots, Y_K\}$, where Y_a are i.i.d. distributed as $\mathrm{Exp}(1/\,(1+\tilde{e}))$. Then, for non-negative x_a,

$$\mathbb{P}\left(\bigcap_{a=1}^{K}\{Y_a \geq x_a\}\right) = \prod_{a=1}^{K} e^{-\frac{x_a}{1+\tilde{e}}}.$$

Let $\boldsymbol{X} = \{X_1, X_2, \ldots, X_K\}$, where $X_a = \mathbb{1}_{\mathcal{C}_a^d} N_a(n)\,\mathrm{KL_{inf}}(\hat{\mu}_a(n), m(\mu_a))$. Then, from Lemma 29, we have:

$$\mathbb{P}_\mu\left(\bigcap_{a=1}^{K}\{X_a \geq x_a\}\right) \leq \mathbb{P}\left(\bigcap_{a=1}^{K}\{Y_a \geq x_a\}\right)\prod_{a=1}^{K}\left\{(n+1)^2 e^{\tilde{B}_a}\right\}.$$

Let

$$R_1(n) := \prod_{a=1}^{n}\left\{(n+1)^2 e^{\tilde{B}_a}\right\}.$$

Define $Z = \{Z_1, Z_2, \ldots, Z_K\}$, where each Z_i is a non-negative random variable, and for $A \subset (\mathbb{R}^+)^K$ and $\mathbf{0} = (0, 0, \ldots, 0)$, Z has a distribution given by:

$$\mathbb{P}(Z \in A) = \frac{1}{R_1(n)}\mathbb{P}(X \in A) + \left(1 - \frac{1}{R_1(n)}\right) \mathbb{1}\left\{\mathbf{0} \in A\right\}. \tag{58}$$

Clearly, for all $x \in \mathbb{R}^K$,

$$\mathbb{P}\left(\bigcap_{a=1}^{K} \{Z_a \geq x_a\}\right) \leq \mathbb{P}\left(\bigcap_{a=1}^{K} \{Y_a \geq x_a\}\right).$$

Hence, (see Theorem 3.3.16, Müller and Stoyan (2002)) for all collections of non negative increasing function $f_1, f_2, \ldots, f_K$ we have $\mathbb{E}\left(\prod_{a=1}^{K} f_a(Z_a)\right) \leq \mathrm{E}\left(\prod_{a=1}^{K} f_a(Y_a)\right)$. Consider $f_a(x) = e^{\theta x}$. Clearly, $f_a(x)$ is non-negative and increasing in x. Using this, (58),

$$\frac{\mathbb{E}\left(\prod_{a=1}^{K} e^{\theta X_a}\right)}{R_1(n)} + \left(1 - \frac{1}{R_1(n)}\right) = \mathbb{E}\left(\prod_{a=1}^{K} e^{\theta Z_a}\right) \leq \mathbb{E}\left(\prod_{a=1}^{K} e^{\theta Y_a}\right),$$

which implies,

$$\mathbb{E}\left(\prod_{a=1}^{K} e^{\theta X_a}\right) \leq R_1(n)\mathbb{E}\left(\prod_{a=1}^{K} e^{\theta Y_a}\right) = R_1(n)\prod_{a=1}^{K}\mathbb{E}\left(e^{\theta Y_a}\right). \tag{59}$$

Let $b = 1/(1+\tilde{e})$. For $0 \leq \theta \leq b$, $\forall a \in [K]$, observe that $\mathbb{E}\left(e^{\theta Y_a}\right) = \frac{b}{b-\theta}$. Then, exponentiating and using Markov's Inequality, below:

$$\mathbb{P}\left(\mathbb{1}_{\mathcal{C}^d}\sum_{a=1}^{K} N_a(n)\,\mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \Gamma\right) \leq \mathbb{E}\left(\prod_{a=1}^{K} e^{\theta X_a}\right)\exp\left\{-\theta\Gamma\right\}.$$

Using (59), we further upper bound the above quantity as:

$$\begin{aligned}\mathbb{P}\left(\mathbb{1}_{\mathcal{C}^d}\sum_{a=1}^{K} N_a(n)\,\mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \Gamma\right) &\leq R_1(n)\mathbb{E}\left(\prod_{a=1}^{K} e^{\theta Y_a}\right)\exp\left\{-\theta\Gamma\right\}\\ &= \exp\left\{-\theta\Gamma\right\}R_1(n)\left(\frac{b}{b-\theta}\right)^K.\end{aligned}$$

Letting $\theta = b - K/\Gamma$, substituting for b and $R_1(n)$ in the above expression, we get:

$$\begin{aligned}\mathbb{P}\left(\mathbb{1}_{\mathcal{C}^d}\sum_{a=1}^{K} N_a(n)\,\mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \Gamma\right) &\leq \exp\left\{-b\Gamma\right\}R_1(n)\left(\frac{e\Gamma b}{K}\right)^K\\ &= e^{-\frac{\Gamma}{(1+\tilde{e})}}\left(\frac{e\Gamma}{K(1+\tilde{e})}\right)^K\prod_{a=1}^{K}\left\{(n+1)^2 e^{\tilde{B}_a}\right\}.\end{aligned}$$

■

C.5.1. PROOF OF THEOREM 13

Recall that $\tilde{e} > 0$, $D = \lceil \log(n)/\log(1+\tilde{e}) \rceil$ and $\mathcal{D} = \left\{ d \in \mathbb{N}^K \; s.t., \forall a \;\; 1 \leq d_a \leq D \right\}$. For $d \in \mathcal{D}$,

$$\mathcal{C}^d = \bigcap_{a=1}^{K} \left\{ (1+\tilde{e})^{d_a - 1} \leq N_a(n) \leq (1+\tilde{e})^{d_a} \right\} = \bigcap_{a=1}^{K} \mathcal{C}^d_a,$$

where $\mathcal{C}^d_a = \left\{ (1+\tilde{e})^{d_a-1} \leq N_a(n) \leq (1+\tilde{e})^{d_a} \right\}$. Clearly,

$$\begin{aligned} &\mathbb{P}\left(\sum_{a=1}^{K} N_a(n)\, \mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \Gamma \right) \\ &\leq \sum_{d \in \mathcal{D}} \mathbb{P}_\mu \left(\sum_{a=1}^{K} N_a(n)\, \mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \Gamma, \; \mathcal{C}^d \right), \end{aligned}$$

which can be further upper-bounded by

$$D^K \max_{d \in \mathcal{D}} \mathbb{P}_\mu \left(\mathbb{1}_{\mathcal{C}^d} \sum_{a=1}^{K} N_a(n)\, \mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \Gamma \right). \tag{60}$$

Lemma 30 bounds the term in the right hand side in the above expression. Using this, choosing the free constant $\tilde{e}$ appropriately and making the necessary approximations, we get the upper bound in Theorem 13.

Using (60) and Lemma 30,

$$\mathbb{P}\left(\sum_{a=1}^{K} N_a(n)\, \mathrm{KL}_{\mathrm{inf}}(\hat{\mu}_a(n), m(\mu_a)) \geq \Gamma \right) \leq e^{-\frac{\Gamma}{(1+\tilde{e})}} \left(\frac{De\Gamma}{K(1+\tilde{e})} \right)^K \prod_{a=1}^{K} \left\{ (n+1)^2 e^{\tilde{B}_a} \right\}. \tag{61}$$

Set

$$\tilde{e} = \frac{1}{\Gamma - 1},$$

and bound $\log(1+\tilde{e}) = -\log(1/(1+\tilde{e})) \geq 1/\Gamma$ to get an upper bound for D. Using these in (61), the expression above can be further bounded by

$$e^{-\Gamma+1} \prod_{a=1}^{K} \left\{ \frac{(e\Gamma \log(n) + 1)(\Gamma - 1)}{K} (n+1)^2 e^{\tilde{B}_a} \right\}.$$

Upper bounding $(n+1)$ by $2n$, and using $(\Gamma - 1)(\Gamma \log(n) + 1) \leq \Gamma^2 \log(n)$, we get the desired bound.

C.5.2. WHEN UNDERLYING DISTRIBUTIONS HAVE BOUNDED SUPPORT

If it is known apriori that the underlying distributions have bounded support, say $[a, b]$, such that $\max\{f(|a|), f(|b|)\} < B$, then it is sufficient to consider a sub-class $\underline{\mathcal{L}}$ of all the probability distributions with support in $[a, b]$, instead of class $\mathcal{L}$. In this case, given a probability measure

$\eta \in \mathcal{P}(\Re)$, and x such that $m(\eta) \leq x \leq b$, it can be shown that $\mathrm{KL}_{\mathrm{inf}}(\eta, x)$ has a simpler formulation as a solution to a 1-dimensional convex optimization problem, which is maximization of $\mathbb{E}_\eta \log\left(1-(X-x)\lambda\right)$ over λ in interval $\left[0, \frac{1}{b-x}\right]$. We also get an exponentially decaying probability of $\mathrm{KL}_{\mathrm{inf}}(\hat{\eta}, x)$ taking positive values for $m(\hat{\eta}) \leq x \leq m(\eta)$, where $\hat{\eta}$ is an empirical distribution of samples from η (see Honda and Takemura (2015)). Furthermore, the algorithm $\mathbf{AL_1}$, with an appropriate choice of β, acting on a vector of K distributions, each coming from class $\underline{\mathcal{L}}$, is δ-correct and matches the lower bound asymptotically, as δ approaches 0.

C.6. Proofs related to sample complexity part of Theorem 10

In this section, we formally prove that the algorithm $\mathbf{AL_1}$ is asymptotically optimal, i.e., the ratio of expected number of samples needed by the algorithm to stop and $\log(1/\delta)$ equals the lower bound of the quantity, asymptotically as $\delta \to 0$.

To this end, we first show that the fraction of times $\mathbf{AL_1}$ pulls arm a is close to its optimal proportion suggested by the lower bound, if the algorithm runs for sufficient time.

C.6.1. PROOF OF LEMMA 14

In order to show that $N_a(lm)/lm \to t_a^*(\mu)$ as $l \to \infty$, for $n \in \mathbb{N}$, let M_n denote the set of indices in $[1, 2, \ldots, n]$ where $\mathbf{AL_1}$ flipped the coins to decide which arm to sample from. Then, for $l \in \mathbb{N}$, from Lemma 8, $lm - |M_{lm}| \leq K\left(\sqrt{lm} - 1\right)$, so that

$$\frac{|M_{lm}|}{lm} \xrightarrow{a.s.} 1, \text{ as } l \to \infty. \tag{62}$$

Further, let $I_a(i) = 1$ if arm a was sampled under $\mathbf{AL_1}$ at step i. Then, by law of large numbers for Bernoulli random variables

$$\frac{1}{|M_n|} \sum_{i \in M_n} \left(I_a(i) - t_a^*(\hat{\mu}(i))\right) \xrightarrow{a.s.} 0, \text{ as } n \to \infty, \tag{63}$$

where we set $\hat{\mu}(i) = \hat{\mu}(lm)$ for $i \in [lm, lm+1, \ldots, (l+1)m - 1]$ for each l.

Further,

$$\frac{1}{|M_n|} \sum_{i \in M_n} \left(t_a^*(\hat{\mu}(i)) - t_a^*(\mu)\right) \xrightarrow{a.s.} 0, \text{ as } n \to \infty, \tag{64}$$

since $\hat{\mu}(n) \to \mu$ as $n \to \infty$, and t^* is a continuous function (Theorem 5).

Furthermore,

$$\begin{aligned} \frac{N_a(lm)}{lm} &= \frac{\sum_{i \in M_{lm}} I_a(i)}{lm} + \frac{\sum_{i \in [lm] \setminus M_{lm}} I_a(i)}{lm} \\ &= \frac{\sum_{i \in M_{lm}} I_a(i)}{|M_{lm}|} \frac{|M_{lm}|}{lm} + \frac{\sum_{i \in [lm] \setminus M_{lm}} I_a(i)}{lm - |M_{lm}|} \frac{lm - |M_{lm}|}{lm}. \end{aligned}$$

From above,

$$\frac{N_a(lm)}{lm} \xrightarrow{a.s.} t_a^*(\mu), \text{ as } l \to \infty.$$

C.6.2. PROOF OF SAMPLE COMPLEXITY

We now prove that for algorithm $\mathbf{AL_1}$, $\limsup_{\delta\to 0} \mathbb{E}_\mu(\tau_\delta)/\left(\log\left(1/\delta\right)\right) \leq V(\mu)^{-1}$. Recall that $\mu \in \mathcal{M}_\mathcal{L}$ is such that $m(\mu_1) > \max_{j\neq 1} m(\mu_j)$. Let $\epsilon > 0$. By continuity of the optimal proportions, $t^*(\mu)$, in μ (Theorem 5), $\exists \zeta(\epsilon) \leq \left(m(\mu_1) - \max_{j\neq 1} m(\mu_j)\right)/4$ (denoted by ζ) such that

$$\forall \mu' \in \mathcal{I}_\epsilon,\ \max_{a\in[K]} \left|t_a^*(\mu') - t_a^*(\mu)\right| \leq \epsilon, \tag{65}$$

for $\mathcal{I}_\epsilon$ defined as follows:

$$\mathcal{I}_\epsilon \triangleq B_\zeta(\mu_1) \times B_\zeta(\mu_2) \times \ldots \times B_\zeta(\mu_K),$$

where

$$B_\zeta(\mu_i) = \{\kappa \in \mathcal{L} : d_W(\kappa, \mu_i) \leq \zeta \text{ and } m(\kappa) \in [m(\mu_i) - \zeta, m(\mu_i) + \zeta]\},$$

and $d_W(\kappa, \mu_i)$ is the Wasserstein metric on $\mathcal{L}$. In particular, whenever $\hat{\mu}(n) \in \mathcal{I}_\epsilon$, the empirical best arm ($\hat{a}_n$) is arm 1. For $T \geq m, T \in \mathbb{N}$, set

$$l_2(T) \triangleq \left\lfloor \frac{T}{m} \right\rfloor, l_1(T) \triangleq \max\left\{1, \left\lfloor \frac{T^{3/4}}{m} \right\rfloor\right\}, \text{ and } l_0(T) \triangleq \max\left\{1, \left\lfloor \frac{T^{1/4}}{m} \right\rfloor\right\},$$

and define

$$\mathcal{G}_T(\epsilon) = \bigcap_{l=l_0(T)}^{l_2(T)} \{\hat{\mu}(lm) \in \mathcal{I}_\epsilon\} \bigcap_{l=l_1(T)}^{l_2(T)} \left\{\max_{a\leq K} \left|\frac{N_a(lm)}{lm} - t_a^*(\mu)\right| \leq 4\epsilon\right\}.$$

Let μ' be a vector of K, 1-dimensional distributions such that the 1^{st} distribution has the maximum mean, and let $t' \in \Sigma_K$. Define the following:

$$g(\mu', t') \triangleq \min_{b\neq 1} \inf_{x\in[m(\mu'_b), m(\mu'_1)]} \left(t'_1 \operatorname{KL_{inf}}(\mu'_1, x) + t'_b \operatorname{KL_{inf}}(\mu'_b, x)\right). \tag{66}$$

Note from Berge's Theorem (reproduced in Appendix A.1) that $g(\mu, t)$ is a jointly continuous function of the (μ, t). Let $\|.\|_K$ be the maximum norm in $\Re^K$, and

$$C_\epsilon^*(\mu) \triangleq \inf_{\substack{\mu'\in\mathcal{I}_\epsilon \\ t':\|t'-t^*(\mu)\|\leq 4\epsilon}} g(\mu', t'). \tag{67}$$

Furthermore, set

$$T_0(\delta) = \inf\left\{T \in \mathbb{N} : l_1(T) \times m + \frac{\beta(T,\delta)}{C_\epsilon^*(\mu)} \leq T\right\}.$$

Since $\tau_\delta \geq 0$,

$$\mathbb{E}_\mu(\tau_\delta) = \sum_{T=0}^{\infty} \mathbb{P}_\mu(\tau_\delta \geq T) \leq T_0(\delta) + m + \sum_{T=T_0(\delta)+m+1}^{\infty} \mathbb{P}_\mu(\tau_\delta \geq T).$$

From Lemma 31 and 32 below,

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_\mu(\tau_\delta)}{\log\left(1/\delta\right)} \leq \frac{(1+\tilde{e})}{C_\epsilon^*(\mu)} + \limsup_{\delta\to 0} \frac{m}{\log\left(1/\delta\right)}. \tag{68}$$

From continuity of $g(\mu', t')$ in (μ', t'), it follows that $C^*_\epsilon(\mu) \xrightarrow{\epsilon \to 0} V(\mu)$. First letting $\tilde{\epsilon} \to 0$ and then letting $\epsilon \to 0$, we get

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\mu(\tau_\delta)}{\log(1/\delta)} \leq \frac{1}{V(\mu)} + \limsup_{\delta \to 0} \frac{m}{\log(1/\delta)}. \tag{69}$$

Since $m = o(\log(1/\delta))$, $\limsup\limits_{\delta \to 0} \frac{m}{\log 1/\delta} = 0$. Using this in (69), we get (13).

Lemma 31

$$\mathbb{E}_\mu(\tau_\delta) \leq T_0(\delta) + m + \sum_{T = T_0(\delta)+m+1}^{\infty} \mathbb{P}_\mu(\mathcal{G}^c_T). \tag{70}$$

Furthermore, for any $\tilde{\epsilon} > 0$,

$$\limsup_{\delta \to 0} \frac{T_0(\delta)}{\log(1/\delta)} \leq \frac{1 + \tilde{\epsilon}}{C^*_\epsilon(\mu)}. \tag{71}$$

Lemma 32

$$\limsup_{\delta \to 0} \frac{\sum\limits_{T=m+1}^{\infty} \mathbb{P}_\mu(\mathcal{G}^c_T(\epsilon))}{\log(1/\delta)} = 0.$$

C.6.3. PROOF OF LEMMA 31

Recall that on $\mathcal{G}_T(\epsilon)$, arm 1 has the highest empirical mean. This follows from choice of ζ and definition of $\mathcal{G}_T(\epsilon)$. Hence on $\mathcal{G}_T(\epsilon)$, for $t \geq l_0(T) \times m$, the log "generalized likelihood ratio" statistic, used in the stopping rule, is given by, $Z(t) = \min_{b \neq 1} Z_{1,b}(t)$ where,

$$Z_{1,b}(t) = t \inf_{x \in [m(\hat{\mu}_b(t)), m(\hat{\mu}_1(t))]} \left(\frac{N_1(t)}{t} \mathrm{KL_{inf}}(\hat{\mu}_1(t), x) + \frac{N_b(t)}{t} \mathrm{KL_{inf}}(\hat{\mu}_b(t), x) \right). \tag{72}$$

In particular, for $T \geq m$ and $l \geq l_1(T)$, on $\mathcal{G}_T(\epsilon)$

$$\begin{aligned} Z(lm) &= \min_{b \neq 1} \inf_{x \in [m(\hat{\mu}_b(lm)), m(\hat{\mu}_1(lm))]} N_1(lm)\, \mathrm{KL_{inf}}(\hat{\mu}_1(lm), x) + N_b(lm)\, \mathrm{KL_{inf}}(\hat{\mu}_b(lm), x) \\ &= lm \times \min_{b \neq 1} \inf_x \left(\frac{N_1(lm)}{lm} \mathrm{KL_{inf}}(\hat{\mu}_1(lm), x) + \frac{N_b(lm)}{lm} \mathrm{KL_{inf}}(\hat{\mu}_b(lm), x) \right) \\ &= lm \times g\left(\hat{\mu}(lm), \left\{ \frac{N_1(lm)}{lm}, \dots, \frac{N_K(lm)}{lm} \right\} \right) \\ &\geq lm \times C^*_\epsilon(\mu). \end{aligned} \tag{73}$$

Furthermore, the stopping time is at most $m \times \inf\{l \geq l_1(T) : Z(lm) \geq \beta(lm, \delta), l \in \mathbb{N}\}$. On $\mathcal{G}_T(\epsilon)$,

$$\begin{aligned}
\min\{\tau_\delta, T\} &\leq l_1(T) \times m + m \sum_{l=l_1(T)+1}^{l_2(T)} \mathbb{1}\left(lm < \tau_\delta\right) \\
&\leq l_1(T) \times m + m \sum_{l=l_1(T)+1}^{l_2(T)} \mathbb{1}\left(Z\left(lm\right) < \beta\left(lm, \delta\right)\right) \\
&\leq l_1(T) \times m + m \sum_{l=l_1(T)+1}^{l_2(T)} \mathbb{1}\left(l < \frac{\beta\left(lm, \delta\right)}{mC^*_\epsilon(\mu)}\right) \\
&\leq l_1(T) \times m + \frac{\beta(T, \delta)}{C^*_\epsilon(\mu)}.
\end{aligned} \tag{74}$$

Recall,

$$T_0(\delta) = \inf\left\{t \in \mathbb{N} : l_1(t) \times m + \frac{\beta(t, \delta)}{C^*_\epsilon(\mu)} \leq t\right\}.$$

On $\mathcal{G}_T$, for $T \geq \max\{m, T_0(\delta)\}$, from (74) and definition of $T_0(\delta)$,

$$\min\{\tau_\delta, T\} \leq l_1(T) \times m + \frac{\beta(T, \delta)}{C^*_\epsilon(\mu)} \leq T,$$

which gives that for such a T, $\tau_\delta \leq T$. Thus, for $T \geq \max\{m, T_0(\delta)\}$, we have $\mathcal{G}_T(\epsilon) \subset \{\tau_\delta \leq T\}$ and hence, $\mathbb{P}_\mu\left(\tau_\delta > T\right) \leq \mathbb{P}_\mu(\mathcal{G}^c_T)$. Since $\tau_\delta \geq 0$,

$$\mathbb{E}_\mu(\tau_\delta) \leq T_0(\delta) + m + \sum_{T=m+1}^{\infty} \mathbb{P}_\mu\left(\mathcal{G}^c_T(\epsilon)\right). \tag{75}$$

Now, to bound $\frac{T_0(\delta)}{\log(1/\delta)}$ as $\delta \to 0$, let $\tilde{e} > 0$ and define

$$C(\tilde{e}) \triangleq \inf\left\{T \in \mathbb{N} : T - ml_1(T) \geq \frac{T}{1+\tilde{e}}\right\} \text{ and } T_2(\delta) \triangleq \inf\left\{T \in \mathbb{N} : \frac{T}{1+\tilde{e}} \geq \frac{\beta(T, \delta)}{C^*_\epsilon(\mu)}\right\}.$$

Then,

$$T_0(\delta) \leq \inf\left\{T \in \mathbb{N} : T - l_1(T) \times m \geq \frac{T}{1+\tilde{e}} \geq \frac{\beta(T, \delta)}{C^*_\epsilon(\mu)}\right\} \leq C(\tilde{e}) + T_2(\delta). \tag{76}$$

From the definition of $T_2(\delta)$ above and using the expression for β from (10),

$$T_2\left(\delta\right) = (1+\tilde{e}) \frac{\log\left(\frac{C}{\delta}\left(\log\frac{C}{\delta}\right)^\alpha\right)}{C^*_\epsilon(\mu)} + O\left(\log\left(\log\frac{1}{\delta}\right)\right). \tag{77}$$

Clearly,

$$\limsup_{\delta \to 0} \frac{T_2(\delta)}{\log 1/\delta} = \frac{1+\tilde{e}}{C^*_\epsilon(\mu)} \text{ and } \limsup_{\delta \to 0} \frac{C(\tilde{e})}{\log 1/\delta} = 0.$$

Taking limits in (76),

$$\limsup_{\delta \longrightarrow 0} \frac{T_0(\delta)}{\log\left(1/\delta\right)} \leq \frac{(1+\tilde{e})}{C^*_\epsilon(\mu)}. \qquad \square$$

Remark 33 Using (76), and (77) in (75), for small δ,

$$\mathbb{E}_\mu(\tau_\delta) \le (1+\tilde{e})\frac{\log\left(\frac{C}{\delta}\left(\log\frac{C}{\delta}\right)^\alpha\right)}{C^*_\epsilon(\mu)} + m + o(\log(1/\delta)).$$

Since $\tilde{e} > 0$ is arbitrary, $\mathbb{E}_\mu(\tau_\delta)$ is bounded by $\frac{\log\left(\frac{C}{\delta}\left(\log\frac{C}{\delta}\right)^\alpha\right)}{C^*_\epsilon(\mu)} + m + o(\log(1/\delta))$. Now, letting ϵ decrease to 0,

$$\mathbb{E}_\mu(\tau_\delta) \le \frac{\log\left(\frac{C}{\delta}\left(\log\frac{C}{\delta}\right)^\alpha\right)}{V(\mu)} + m. \tag{78}$$

We use the rhs in (78) as a proxy for $\mathbb{E}_\mu(\tau_\delta)$ in our numerical experiments.

C.6.4. PROOF OF LEMMA 32

Fix $T \ge m+1$. Let

$$\mathcal{G}^1_T \triangleq \left\{ \bigcap_{l'=l_0(T)}^{l_2(T)} \hat{\mu}(l'm) \in \mathcal{I}_\epsilon \right\}.$$

Using union bounds,

$$\mathbb{P}_\mu(\mathcal{G}^c_T(\epsilon)) \le \sum_{l=l_0(T)}^{l_2(T)} \mathbb{P}_\mu\left(\hat{\mu}(lm) \notin \mathcal{I}_\epsilon\right) + \sum_{l=l_1(T)}^{l_2(T)} \sum_{i=1}^{K} \mathbb{P}\left(\left|\frac{N_i(lm)}{lm} - t^*_i(\mu)\right| \ge 4\epsilon,\ \mathcal{G}^1_T\right).$$

The first term above can be bounded as:

$$\begin{aligned}\mathbb{P}_\mu\left(\hat{\mu}(lm) \notin \mathcal{I}_\epsilon\right) \le& \sum_{i=1}^{K} \mathbb{P}_\mu\left(d_W(\hat{\mu}_i(lm), \mu_i) \ge \zeta\right) \\ &+ \sum_{i=1}^{K} \mathbb{P}_\mu\left(m(\hat{\mu}_i(lm)) \notin [m(\mu_i)-\zeta, m(\mu_i)+\zeta]\right).\end{aligned} \tag{79}$$

For $l \ge 1$, by Lemma 8, $N_a(lm) \ge \sqrt{lm}-1$ for each arm a. Let $\hat{\mu}_{(a,s)}$ denote the empirical distribution corresponding to s samples from arm a. Using union bound,

$$\begin{aligned}\mathbb{P}_\mu\left(d_W(\hat{\mu}_i(lm), \mu_i) \ge \zeta\right) &= \mathbb{P}_\mu\left(d_W(\hat{\mu}_i(lm), \mu_i) \ge \zeta,\ N_i(lm) \ge \sqrt{lm}-1\right) \\ &\le \sum_{s=\sqrt{lm}-1}^{T} \mathbb{P}_\mu\left(d_W(\hat{\mu}_{(i,s)}, \mu_i) \ge \zeta\right).\end{aligned}$$

By Sanov's Theorem, there exist non-decreasing functions $f_i : \Re \to \Re$ and $g_i : \Re \to \Re$ which satisfy $\lim_{s\to\infty} \log f_i(s)/s \to 0$ and $\lim_{s\to\infty} g_i(s)/s \to c_i$ for some constants, c_i and that

$$\mathbb{P}_\mu\left(d_W(\hat{\mu}_{(i,s)}, \mu_i) \ge \zeta\right) \le f_i(s)e^{-g_i(s)}.$$

Let

$$\tilde{f}_i(T) = \max_{j=-1,0,1,\dots,T-\sqrt{lm}} f_i\left(\left\lceil\sqrt{lm}\right\rceil + j\right) \text{ and } \tilde{g}_i(T) = \min_{j=-1,0,1,\dots,T-\sqrt{lm}} g_i\left(\left\lceil\sqrt{lm}\right\rceil + j\right).$$

Clearly, both $\tilde{f}_i$ and $\tilde{g}_i$ are non-decreasing and satisfy $\lim_{s\to\infty} \log \tilde{f}_i(s)/s \to 0$ and $\lim_{s\to\infty} \tilde{g}_i(s)/s \to \tilde{c}_i$ for some constant $\tilde{c}_i$. Further upper bounding (79) using these,

$$\mathbb{P}_\mu\left(d_W(\hat{\mu}_i(lm),\mu_i) \geq \zeta\right) \leq \sum_{s=\sqrt{lm}-1}^{T} f_i(s)e^{-g_i(s)} \leq T\tilde{f}_i(T)e^{-\tilde{g}_i(T)}. \tag{80}$$

Recall that $\rho(x,y)$ denotes the Kullback-Leibler divergence between Bernoulli random variables with mean x and y. Using union bound and Chernoff's bound,

$$\begin{aligned}\mathbb{P}_\mu\left(m(\hat{\mu}_i(lm)) \leq m(\mu_i)-\zeta\right) &= \mathbb{P}_\mu\left(m(\hat{\mu}_i(lm)) \leq m(\mu_i)-\zeta,\ N_i(lm) \geq \sqrt{lm}-1\right)\\ &\leq \sum_{s=\sqrt{lm}-1}^{T} e^{-sd(m(\mu_i)-\zeta,m(\mu_i))}\\ &\leq \frac{e^{-(\sqrt{lm}-1)\times d(m(\mu_i)-\zeta,m(\mu_i))}}{1-e^{-d(m(\mu_i)-\zeta,m(\mu_i))}}.\end{aligned} \tag{81}$$

Similarly we have the other inequality:

$$\mathbb{P}_\mu\left(m(\hat{\mu}_i(lm)) \geq m(\mu_i)+\zeta\right) \leq \frac{e^{-(\sqrt{lm}-1)\times d(m(\mu_i)+\zeta,m(\mu_i))}}{1-e^{-d(m(\mu_i)+\zeta,m(\mu_i))}}. \tag{82}$$

Define the following constants:

$$E_2 := \min_i\left(\min\left\{d(m(\mu_i)-\zeta,m(\mu_i)), d(m(\mu_i)+\zeta,m(\mu_i))\right\}\right),$$

$$F(T) := \max_a \tilde{f}_a(T), \quad G(T) := \min_a \tilde{g}_a(T),$$

and

$$E_1 := \sum_{i=1}^{K}\left(\frac{e^{d(m(\mu_i)-\zeta,m(\mu_i))}}{1-e^{-d(m(\mu_i)-\zeta,m(\mu_i))}} + \frac{e^{d(m(\mu_i)+\zeta,m(\mu_i))}}{1-e^{-d(m(\mu_i)+\zeta,m(\mu_i))}}\right).$$

Note that E_1 and E_2 are non-negative constants and $F(T)$ and $G(T)$ are also non-negative, non-decreasing and satisfy $\lim_{s\to\infty} \log F(s)/s \to 0$ and $\lim_{s\to\infty} G(s)/s \to c$ for some non-negative constant c. Then using (79), (80), (81) and (82) with the constants defined above,

$$\begin{aligned}\sum_{l=l_0(T)}^{l_2(T)} \mathbb{P}_\mu\left(\hat{\mu}(lm) \notin \mathcal{I}_\epsilon\right) &\leq \sum_{l=l_0(T)}^{l_2(T)} E_1 \exp\left(-E_2\sqrt{lm}\right) + \sum_{l=l_0(T)}^{l_2(T)}\sum_{a=1}^{K} T\tilde{f}_a(T)\exp\left\{-\tilde{g}_a(T)\right\}\\ &\leq \frac{E_1 T}{m}\exp\left(-E_2 T^{1/8}\right) + \frac{T^2 K}{m}F(T)\exp\left\{-G(T)\right\}.\end{aligned} \tag{83}$$

To bound the other term in the probability of complement of good set, for $l_2(T) \geq l \geq l_1(T)$, let

$$A_2 := \frac{1}{lm}\sum_{j\in M_{lm}} \left|t_i^*(\hat{\mu}(j)) - t_i^*(\mu)\right|, \quad \text{and} \quad A_3 := \frac{1}{lm}\sum_{j\notin M_{lm}} \left|I_i(j) - t_i^*(\mu)\right|.$$

Observe that

$$\mathbb{P}\left(\left|\frac{N_i(lm)}{lm} - t_i^*(\mu)\right| \geq 4\epsilon,\ \mathcal{G}_T^1\right) \leq \mathbb{P}\left(\underbrace{\frac{1}{lm}\left|\sum_{j\in M_{lm}} (I_i(j) - t_i^*(\hat{\mu}(j)))\right|}_{:=A_1} + A_2 + A_3 \geq 4\epsilon,\ \mathcal{G}_T^1\right).$$

Since $|I_i(j) - t_i^*(\mu)| \leq 1$, and from Lemma 8, the sampling algorithm ensures that $lm - |M_{lm}| \leq K\sqrt{lm}$, the term A_3 in the above expression can be bounded from above as,

$$A_3 \leq \frac{lm - |M_{lm}|}{lm} \leq \frac{K\sqrt{lm}}{lm} = \frac{K}{\sqrt{lm}},$$

If batch size m is proportional to $\log(1/\delta)$ (see (21) and the associated discussion for the choice of batch size) and decreases with increasing δ. Since we are only interested in values of δ close to 0, $A_3 \leq \epsilon$ for all T. Next,

$$A_2 = \frac{1}{lm}\sum_{\substack{j\in M_{lm}\\ j<l_0(T)m}} |t_i^*(\hat{\mu}(j)) - t_i^*(\mu)| + \frac{1}{lm}\sum_{\substack{j\in M_{lm}\\ j\geq l_0(T)m}} |t_i^*(\hat{\mu}(j) - t_i^*(\mu)|.$$

Observe that if $l_0(T) = 1$, then the first term above is 0 since for $j < m$, the algorithm does not flip any coins to decide the allocation of samples, and hence $|M_m| = 0$. On the other hand, if $l_0(T) = \left\lfloor \frac{T^{1/4}}{m} \right\rfloor$, then $l_1(T) = \left\lfloor \frac{T^{3/4}}{m} \right\rfloor$ and the first term being at most $\frac{l_0(T)m}{l_1(T)m}$, is bounded by $\frac{1}{T^{1/2}}$. However, since $T \geq m+1$ and $m \propto \log(1/\delta)$, for δ close to 0, $1/m \leq \epsilon$. Thus, the first term is less than ϵ for all $T \geq m$.

For the second term, for $j \geq l_0(T) \times m$, $\hat{\mu}(j)$ lies in $\mathcal{I}_\epsilon$, and hence this term is bounded by ϵ. This gives that $A_2 \leq 2\epsilon$.

Thus, for $T \geq m,$ and $l \geq l_1(T)$:

$$\begin{aligned}\mathbb{P}\left(\left|\frac{N_i(lm)}{lm} - t_i^*(\mu)\right| \geq 4\epsilon,\ \mathcal{G}_T^1\right) &\leq \mathbb{P}\left(\frac{1}{lm}\left|\sum_{j\in M_{lm}} (I_i(j) - t_i^*(\hat{\mu}(j)))\right| + 2\epsilon + \epsilon \geq 4\epsilon,\ \mathcal{G}_T^1\right)\\ &\leq \mathbb{P}\left(\left|\sum_{j\in M_{lm}} (I_i(j) - t_i^*(\hat{\mu}(j)))\right| \geq lm\epsilon\right).\end{aligned}$$

Let $S_n = \sum_{j\in M_n} (I_i(j) - t_i^*(\hat{\mu}(j)))$. Clearly, S_n being sum of zero-mean random variables, is a martingale. Further, $|S_{n+1} - S_n| \leq 1$. Thus using Azuma-Hoeffding inequality,

$$\mathbb{P}\left(\left|\frac{N_i(lm)}{lm} - t_i^*(\mu)\right| \geq 4\epsilon,\ \mathcal{G}_T^1\right) \leq 2\exp\left(-\frac{l^2m^2\epsilon^2}{2|M_{lm}|}\right) \leq 2\exp\left(-\frac{lm\epsilon^2}{2}\right).$$

Summing over l and i, the above bounded from above by

$$\sum_{l=l_1(T)}^{l_2(T)} 2K\exp\left(-\frac{lm\epsilon^2}{2}\right) \leq \frac{2KT}{m}\exp\left(-\frac{l_1(T)\times m\epsilon^2}{2}\right). \tag{84}$$

Combining (83) and (84), we get the desired result.

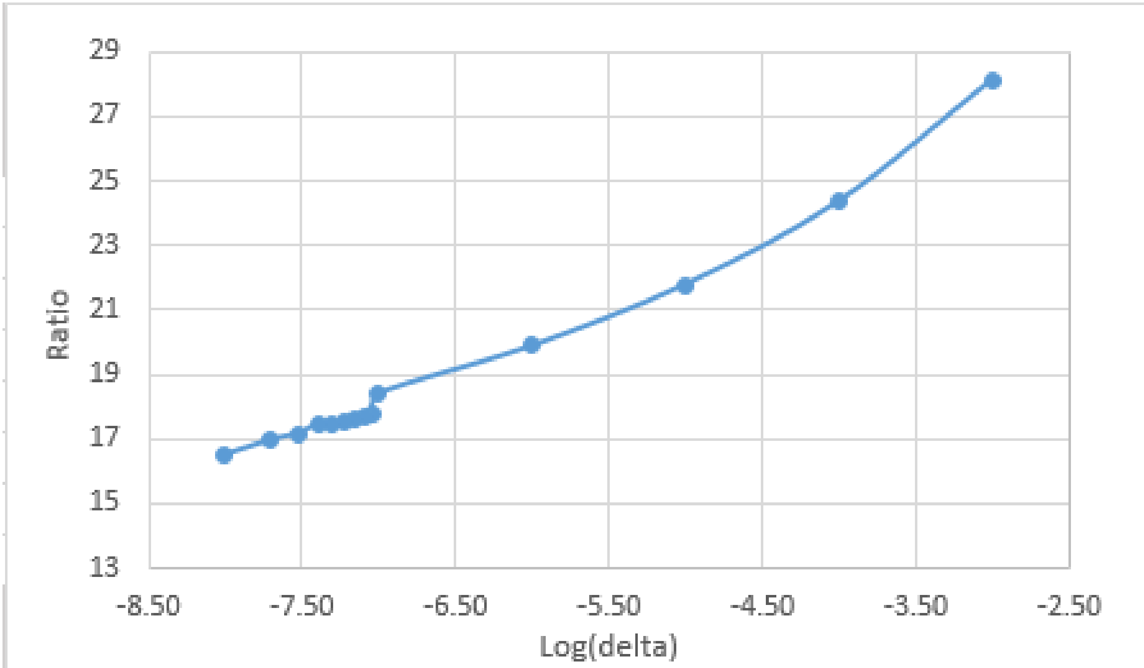

Figure 1: Ratio of average number of samples needed by the algorithm to stop and the lower bound as a function of $\log(\delta)$.

C.7. Numerical experiment

In this section we give the experimental results for the algorithm $\mathbf{AL_1}$ on the example discussed in Section 5. Recall that we consider a 4-armed bandit, with each arm having a Pareto distribution with parameters (α, β), that is supported on $[\beta, \infty)$ and has pdf $f_{\alpha,\beta}(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}$. The four arms have parameters set to $(4, 1.875), (4, 1.5), (4, 1.25)$, and $(4, 0.75)$. Also, recall that we choose $\mathcal{L}$ with $f(y) = y^2$ and $B = 9$.

We first test numerically that the expected number of samples until termination needed by $\mathbf{AL_1}$, when the underlying distributions are unknown, approaches (as δ decreases to 0) the lower bound on this quantity, which is computed assuming that the underlying distributions are known. To this end, Figure 1 plots the ratio of average number of samples needed by $\mathbf{AL_1}$ to stop, and the lower bound on this quantity, as a function of $\log(\delta)$. As can be seen from the figure, as δ reduces from 10^{-3} to 10^{-8}, this ratio decreases from 28 to 16.

Let c_1 denote the average cost of generating a sample from the arms (measured in seconds per sample). Let *computational cost* be the cost incurred by $\mathbf{AL_1}$ in solving the max-min optimization problem (5) at the end of each batch until termination (again, measured in seconds). Then, we call the sum of cost of generating all samples ($c_1 \times$ number of samples) and the computational cost as the *total cost* of $\mathbf{AL_1}$ measured in seconds.

In Figure 2, we demonstrate the total cost of $\mathbf{AL_1}$ as a function of the batch size, for a fixed δ (set to 0.01). The horizontal axis in the figure denotes the batch size in thousands and both the vertical axes correspond to average total cost, averaged across 20 independent experiments, performed on

Figure 2: Total computational cost for different values of c_1, as a function of batch size.

a standard desktop with 8 GB RAM running Linux operating system, using Python language for simulations. Solving the max-min optimization problem once after generating 2000, 5000, 10000 and 20000 samples takes $35, 69, 135,$ and 192 minutes, respectively on this system. We plot the total cost corresponding to two different values of c_1. The left vertical axis corresponds to the green curve, i.e., for $c_1 = 3$ seconds. The right one denotes the value of average total cost when $c_1 = 0.0001$ seconds (orange curve).

The figure shows that the total cost initially comes down with an increase in the batch size. This is because as the batch size increases, number of batches required for $\mathbf{AL_1}$ to terminate reduces, decreasing the computational cost. Left most points in both the curves of Figure 2 correspond to the batch size of 2000. On increasing the batch size from 2000 to 8000, the average number of batches until termination reduces from 5.7 to 2, reducing the average total cost.

With a further increase in the batch size up to a point, the figure indicates an increase in the total cost. This can be explained as follows: on increasing the batch size up to 17000, we observe that $\mathbf{AL_1}$ still requires 2 batches (in all the 15 independent experiments) to terminate. However, due to increase in batch size (from 8000 to 17000), there is an increase in delay in stopping, i.e., since $\mathbf{AL_1}$ checks for the stopping condition only at the end of each batch, the algorithm samples more than required number of samples in the last batch, thus increasing the overall sampling cost ($c_1 \times$ number of samples until termination) for the algorithm. When c_1 is high, this increase contributes significantly to the total cost (green curve). There is another phenomenon that explains the increase in total cost, especially when c_1 is small (orange curve). Because of increase in the number of samples per batch, at the end of each batch, $\mathbf{AL_1}$ solves the max-min problem for distributions that have bigger support sizes. Since solution of the max-min problem involves computing KL_{inf} whose computation time increases linearly with the support size of the distributions, computational cost of $\mathbf{AL_1}$ increases.

Increasing the batch size beyond this point reduces the number of batches required and hence a reduction in the computational cost and total cost is observed upto the point for which 1 batch is sufficient (batch size of 30000). As the batch size increases beyond this, an increase in the total cost is seen due to exactly the same reasons outlined in a previous paragraph.

In the figure, batch sizes of 8000, 17000 and 30000 correspond to the points of local minima, maxima, and minima (from left to right), respectively. Figure 2 also shows that when the cost of sampling is low (e.g., in online recommender systems), the optimal batch size is large (30000 for orange curve) and one batch is sufficient for the algorithm to stop. This suggests that in such cases, sampling uniformly from each arm minimizes the the total cost, even if the total number of samples needed are high. However, when c_1 is high (e.g., in the setting of clinical trials, where each sample is very costly), optimal batch size is small (4000 for green curve) to minimize the over-sampling in last batch.