Sample complexity of partition identification using multi-armed bandits

Sandeep Juneja JUNEJA@TIFR.RES.IN
TIFR, Mumbai

Subhashini Krishnasamy SUBHASHINI.KB@UTEXAS.EDU
TIFR Mumbai

Editors: Alina Beygelzimer and Daniel Hsu

Abstract

Given a vector of probability distributions, or arms, each of which can be sampled independently, we consider the problem of identifying the partition to which this vector belongs from a finitely partitioned universe of such vector of distributions. We study this as a pure exploration problem in multi-armed bandit settings and develop sample complexity bounds on the total mean number of samples required for identifying the correct partition with high probability. This framework subsumes well studied problems such as finding the best arm or the best few arms. We consider distributions belonging to the single parameter exponential family and primarily consider partitions where the vector of means of arms lie either in a given set or its complement. The sets considered correspond to distributions where there exists a mean above a specified threshold, where the set is a half space and where either the set or its complement is a polytope, or more generally, a convex set. In these settings, we characterize the lower bounds on mean number of samples for each arm highlighting their dependence on the problem geometry. Further, inspired by the lower bounds, we propose algorithms that can match these bounds asymptotically with decreasing probability of error. Applications of this framework may be diverse. We briefly discuss a few associated with simulation in finance.

Keywords: multi-armed bandits, best arm identification, pure exploration, partition identification

1. Introduction

Suppose that $\mathcal{Q}$ denotes a collection of vectors $\nu = (\nu_1, \dots, \nu_K)$ where each ν_i is a probability distribution on $\mathbb{R}$. Further, $\mathcal{Q} = \cup_{i=1}^m \mathcal{A}_i$ where the component sets $\mathcal{A}_i$ are disjoint, and thus partition $\mathcal{Q}$. In this set-up, given $\mu = (\mu_1, \dots, \mu_K) \in \mathcal{Q}$, we consider the problem of identifying the correct component $\mathcal{A}_i$ that contains μ. The distributions $(\mu_i : i \leq K)$ are not known to us, however, it is possible to generate independent samples from each μ_i. We call this the partition identification or $\mathcal{PI}$ problem.

We consider algorithms that sequentially and adaptively generate samples from each distribution in μ and then after generating finitely many samples, stop and announce a component of $\mathcal{Q}$ that is inferred to contain μ. Specifically, we study the *δ-correct* algorithms in the $\mathcal{PI}$ framework.

Definition 1 An algorithm is said to be δ-correct for the $\mathcal{PI}$ problem $\mathcal{Q} = \cup_{i=1}^m \mathcal{A}_i$, if, for every $\mu \in \mathcal{Q}$, for any specified $\delta \in (0, 1)$, it restricts the probability of announcing an incorrect component to at most δ.

More generally, in similar sequential decision making problems, algorithms are said to provide δ-correct guarantees if the probability of incorrect decision is bounded from above by δ for each $\delta \in (0,1)$.

In multi-armed bandit (MAB) literature, for any $\nu \in \mathcal{Q}$, generating a sample from distribution ν_i is referred to as sampling from, or pulling, an arm i. The $\mathcal{PI}$ framework is quite general and captures popular pure exploration problems studied in the MAB literature. For instance, the problem of finding the best arm, that is, the arm with the highest mean, is well studied and fits $\mathcal{PI}$ framework (see, e.g., in learning theory 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 theory literature - Glynn and Juneja (2004), Kim and Nelson (2001), Chen et al. (2000), Dai (1996), Ho et al. (1992)).

More generally, identifying r arms (for some $r < K$) with the the largest r means amongst K distributions also is a $\mathcal{PI}$ problem (see, e.g., Kaufmann and Kalyanakrishnan (2013), Kalyanakrishnan et al. (2012)).

Sample complexity of an algorithm is defined as the expected total number of arms pulled by the algorithm before it terminates. Further, δ-correct guarantees provided by algorithms impose constraints on expected number of times each arm must be pulled. These constraints are made explicit using the 'transportation inequality' developed by Garivier and Kaufmann (2016). (Their work in turn is built upon 'change of measure' based earlier analysis that goes back at least to Lai and Robbins (1985). See also Mannor and Tsitsiklis (2004), Burnetas and Katehakis (1996)). The transportation inequality allows us to formulate the problem of arriving at *efficient* lower bounds on sample complexity in the $\mathcal{PI}$ framework as an optimization problem - a linear program with infinitely many constraints; this also has an equivalent max-min formulation. We refer to the resulting optimization problem as the *lower bound problem*.

The advantage of $\mathcal{PI}$ framework is that it provides a unified approach to tackle a large class of problems, both in developing efficient lower bounds on the sample complexity of δ-correct algorithms, as well as in arriving at δ-correct algorithms with sample complexity that asymptotically (as $\delta \to 0$) matches the developed lower bounds under certain distributional restrictions on the arms.

To further analyze the lower bound problem, we assume that each arm distribution belongs to a single parameter exponential family (SPEF). See, e.g., Cappé et al. (2013), Garivier and Kaufmann (2016), Kaufmann et al. (2016), where similar distributional restrictions are imposed (see Glynn and Juneja (2018) for need for distributional restrictions). Examples of SPEF distributions include Binomial, Poisson, Gaussian with known variance, Gamma with known shape parameter. See, Cappé et al. (2013) for an elaborate discussion on SPEF distributions. Any member of SPEF distribution can be uniquely represented by its mean. This allows us to consider the partition problem in the parameter space (i.e., $\mathcal{Q} \subset \mathbb{R}^K$) instead of the distribution space. This further allows us to highlight the geometrical structure of the lower bound problem in a relatively simple manner.

We solve the lower bound problem for SPEF distributions, so that $\mathcal{Q} \subset \mathbb{R}^K$. Our focus is primarily on $\mathcal{Q} = \mathcal{A}_1 \cup \mathcal{A}_2$, where we consider the following settings:

Threshold crossing problem: For a threshold $u \in \mathbb{R}$, $\mathcal{A}_1 = \{\nu \in \mathbb{R}^K : \max_{i \leq K} \nu_i > u\}$ and $\mathcal{A}_2 = \{\nu \in \mathbb{R}^K : \max_{i \leq K} \nu_i < u\}$, we explicitly solve the lower bound problem for each $\mu \in \mathcal{Q}$. We refer to this as the threshold crossing problem, and point to the elegant asymmetry in the lower bounds depending upon whether $\mu \in \mathcal{A}_1$ or $\mu \in \mathcal{A}_2$. This problem was also analysed in-depth in Kaufmann et al. (2018). In Appendix A, we briefly discuss how this problem arises naturally in financial portfolio risk measurement involving nested simulations.

Half-space problem: For specified $(a_1, \ldots, a_K, b) \in \mathbb{R}^{K+1}$, $\mathcal{A}_1 = \{\nu \in \mathbb{R}^K : \sum_{i=1}^K a_i \nu_i > b\}$ and $\mathcal{A}_2 = \{\nu \in \mathbb{R}^K : \sum_{i=1}^K a_i \nu_i < b\}$, we characterize the solution to the lower bound problem for each $\mu \in \mathcal{Q}$.

Convex set problem: When $\mathcal{A}_2$ is a closed convex set and $\mathcal{A}_1$ is its complement in $\mathbb{R}^K$, we characterize the solution to the lower bound problem for each $\mu \in \mathcal{A}_1$. We also do this for $\mu \in \mathcal{A}_1$ when $\mathcal{A}_1$ is a polytope, and $\mathcal{A}_2$, a complement of $\mathcal{A}_1$, is a union of half-spaces. In these settings we highlight the geometric structure of the problem and propose geometry based simple algorithms to compute the lower bound solutions. Further, we use the duality approach (through Sion's Minimax Theorem) to considerably simplify our analysis.

Applications of the half-space and the convex set problem are many. These include identifying acceptable combination of projects in capital budgeting where samples of profitability from each project is a random output of a simulation model, and there may be constraints on the overall expected profitability from the selected projects as well as on minimum expected profitability from each project. Applications also include selecting financial assets in an investments portfolio with similar constraints on expected portfolio returns, and on minimum expected returns from each security.

Garivier and Kaufmann (2016) solve the lower bound problem in the best arm setting. They further use the solution to arrive at an adaptive δ-correct algorithm whose stopping rule is based on the generalized likelihood ratio test earlier proposed in Chernoff (1959). Also, see Albert (1961). The sample complexity of their proposed algorithm is shown to asymptotically match the lower bound solution (as $\delta \to 0$). Under mild conditions, we show that the solution to the $\mathcal{PI}$ lower bound problem is a continuous function of the underlying expectation of arms (see Lemma 13). This allows us to adapt their algorithm to the $\mathcal{PI}$ setting, to again arrive at an adaptive δ-correct algorithm whose sample complexity asymptotically matches the corresponding lower bound.

The $\mathcal{PI}$ framework was also considered in Chernoff (1959) where $\mathcal{Q}$ was restricted to be finite. Albert (1961) generalized this work to allow for $\mathcal{Q}$ with infinite elements. The key difference of our paper compared to these references is that we work in a δ-correct framework that provides explicit error guarantees. Their work involves guarantees with constants that are not explicitly available. Further, as mentioned earlier, we use duality methods, and exploit the geometry of the solution to the lower bound problem to solve it efficiently. These issues are not considered in Chernoff (1959) and Albert (1961).

Roadmap: In Section 2, we state the transportation inequality from Kaufmann et al. (2016) and state the resultant lower bound problem in $\mathcal{PI}$ framework as an optimization problem. We also spell out preliminaries such as SPEF distributions and related assumptions in this section. In Section 3, we characterize the solution to the lower bound problem for various special cases of partition of $\mathcal{Q}$ into disjoint sets $\mathcal{A}_1$ and $\mathcal{A}_2$. For the threshold crossing problem (Section 3.1), we give a closed form expression for the solution to the lower bound problem. For the half-space problem (Section 3.2), we give a simple characterization of the solution that allows for easy numerical evaluation. Similarly, for the problem where $\mathcal{Q}$ is partitioned into a convex set and its complement, we derive some useful properties of the solution to the lower bound problem (Sections 3.3, 3.4). In Section 4, we propose a δ-correct algorithm that in substantial generality achieves the derived lower bounds asymptotically as δ decreases to zero. The detailed proofs are given in the appendices.

2. Preliminaries and basic optimization problem

Recall that $\mathcal{Q}$ denotes a collection of vectors $\nu = (\nu_1, \ldots, \nu_K)$ where each ν_i is a probability distribution in $\mathbb{R}$. Further, $\mathcal{Q} = \cup_{i=1}^m \mathcal{A}_i$ where the $\mathcal{A}_i$ are disjoint, and thus partition $\mathcal{Q}$.

Let $KL(\mu_i||\nu_i) = \int \log(\frac{d\mu_i}{d\nu_i}(x))d\mu_i(x)$ denote the Kullback-Leibler divergence between distributions μ_i and ν_i. We further assume that for each $\nu, \tilde{\nu} \in \mathcal{Q}$, the components ν_i and $\tilde{\nu}_i$ for each i are mutually absolutely continuous. For $p, q \in (0,1)$, let

$$d(p,q) := p \log\left(\frac{p}{q}\right) + (1-p)\log\left(\frac{1-p}{1-q}\right),$$

that is, $d(p,q)$ denotes the KL-divergence between Bernoulli distributions with mean p and q, respectively. For any set $\mathcal{B}$, let $\mathcal{B}^c$ denote its complement, $\mathcal{B}^o$ its interior, $\bar{\mathcal{B}}$ its closure and $\partial\mathcal{B}$ its boundary.

Under a δ-PAC algorithm, and for $\mu \in \mathcal{A}_j$, the following transportation inequality follows from Kaufmann et al. (2016):

$$\sum_{i=1}^K E_\mu N_i \, KL(\mu_i||\nu_i) \geq d(\delta, 1-\delta) \geq \log\left(\frac{1}{2.4\delta}\right) \tag{1}$$

for any $\nu \in \mathcal{A}_j^c$, where N_i denotes the number of times arm i is pulled by the algorithm. Taking $t_i = E_\mu N_i / \log(\frac{1}{2.4\delta})$, our lower bound on sample complexity problem can be modelled as the following convex programming problem, when $\mu \in \mathcal{A}_j$ (call it **O1**):

$$\min_{\mathbf{t}=(t_1,\ldots,t_K)} \quad \textstyle\sum_{i=1}^K t_i$$

$$\text{s.t.} \quad \inf_{\nu \in \mathcal{A}_j^c} \sum_{i=1}^K t_i KL(\mu_i||\nu_i) \quad \geq 1, \quad t_i \geq 0 \;\; \forall i.$$

Letting $w_i = \frac{t_i}{\sum_j t_j}$ and $\mathcal{P}_K \triangleq \{w \in \mathbb{R}^k : w_i \geq 0 \; \forall i, \sum_{i=1}^K w_i = 1\}$ denote the K-dimensional probability simplex, **O1** maybe equivalently stated as

$$\max_{w \in \mathcal{P}_K} \inf_{\nu \in \mathcal{A}_j^c} \sum_{i=1}^K w_i KL(\mu_i||\nu_i). \tag{Problem LB}$$

Let $C^*(\mu)$ be the optimal value of the above problem. The lower bound on the total expected number of samples is then given by $\log(\frac{1}{2.4\delta})T^*(\mu)$ where $T^*(\mu) = 1/C^*(\mu)$.

Remark 2 While the optimization problem **O1** is equivalent to Problem LB, one advantage of the former is that it can be viewed as a linear program with infinitely many constraints, or a semi-infinite linear program (see, e.g., López and Still (2007)). Then linear programming duality provides a great deal of insight into the solution structure. However, we instead present our analysis using the max-min Problem LB, since Sion's minimax theorem can be applied on it to directly arrive at the solution.

Single Parameter Exponential Families (SPEF): In the remaining paper, we consider SPEF of distributions for each arm. For each $1 \leq i \leq K$, let ρ_i denote a reference measure on the real line, and let $\Lambda_i(\eta) \triangleq \log\left(\int_{x\in\mathbb{R}} \exp(\eta x) d\rho_i(x)\right)$. Λ_i is referred to as a cumulant or a log-partition function. Further, set $\mathcal{D}_i \triangleq \{\eta : \Lambda_i(\eta) < \infty\}$.

An SPEF distribution for arm i and $\eta \in \mathcal{D}_i$, $p_{i,\eta}$, has the form $dp_{i,\eta}(x) = \exp(\eta x - \Lambda_i(\eta)) d\rho_i(x)$. Note that Λ_i is $\mathcal{C}^\infty$ in $\mathcal{D}_i^o$ (see, e.g., 2.2.24 Dembo and Zeitouni (2011)). Further, $\Lambda_i(\eta)$ is a convex function of $\eta \in \mathcal{D}_i^o$, and if the underlying distribution is non-degenerate, then it is strictly convex.

Let Λ_i^* denote the Legendre-Fenchel transform of Λ_i, that is, $\Lambda_i^*(\theta) = \sup_{\eta\in\mathcal{D}_i}(\eta\theta - \Lambda_i(\eta))$.

Further, let μ_i denote the mean under p_{i,η_i}. Then, $\mu_i = \Lambda_i'(\eta_i)$ for $\eta_i \in \mathcal{D}_i^o$. In particular, μ_i is a strictly increasing function of η_i, and there is a one-to-one mapping between the two. Below we suppress the notational dependence of μ_i on η_i and vice-versa.

Let $\mathcal{U}_i \triangleq \{\Lambda_i'(\eta_i), \eta_i \in \mathcal{D}_i^o\}$. Since $\Lambda_i'(\eta_i)$ is strictly increasing for $\eta_i \in \mathcal{D}_i^o$, $\mathcal{U}_i$ is an open interval, and sans the boundary cases, denotes the value of means attainable for arm i. For $\eta_i \in \mathcal{D}_i^o$, the following are well known and easily checked: $\eta_i = \Lambda_i^{*\prime}(\mu_i)$, and

$$\Lambda_i^*(\mu_i) + \Lambda_i(\eta_i) \quad = \quad \mu_i\eta_i. \tag{2}$$

For $\eta_i, \beta_i \in \mathcal{D}_i^o$, It is easily seen that $KL(p_{i,\eta_i}||p_{i,\beta_i}) = \Lambda_i(\beta_i) - \Lambda_i(\eta_i) - \mu_i(\beta_i - \eta_i)$, where again $\mu_i = \Lambda_i'(\eta_i)$. We denote the above by $K_i(\mu_i|\nu_i)$ with $\nu_i = \Lambda_i'(\beta_i)$ emphasizing that when the two distributions are from the same SPEF, Kullback-Leibler divergence only depends on the mean values of the distributions. Using (2), we have

$$K_i(\mu_i|\nu_i) = \Lambda_i^*(\mu_i) - \Lambda_i^*(\nu_i) - \beta_i(\mu_i - \nu_i), \tag{3}$$

where $\beta_i = \Lambda_i^{*\prime}(\nu_i)$. Again, it can be shown that Λ_i^* is $\mathcal{C}^\infty$ in $\mathcal{U}_i$ (see, 2.2.24 Dembo and Zeitouni (2011)), and it is strictly convex if Λ_i is strictly convex. Thus, K_i is $\mathcal{C}^\infty$ in $\mathcal{U}_i$ with respect to each of its arguments.

In the remaining paper, Problem LB refers to

$$\max_{w\in\mathcal{P}_K} \inf_{\nu\in\mathcal{A}_j^c} \sum_{i=1}^{K} w_i KL(\mu_i|\nu_i), \tag{4}$$

each $\mathcal{A}_k$ a subset of $\mathbb{R}^K$, and again $\mu \in \mathcal{A}_j$.

Remark 3 For any $w \in \mathcal{P}_K$, the sub-problem in LB, $\inf_{\nu\in\mathcal{A}_j^c} \sum_{i=1}^K w_i KL(\mu_i|\nu_i)$, has an elegant geometrical interpretation. For $c > 0$, consider the sublevel set

$$S(\mu, w, c) \triangleq \left\{\nu : \sum_{i=1}^{K} w_i KL(\mu_i|\nu_i) \le c\right\}.$$

Then, for element-wise strictly positive w, $S(\mu, w, 0) = \{\mu\}$. The set $S(\mu, w, c)$ for some $c > 0$ intersects with $\mathcal{A}_j^c$. Further, the set shrinks as c reduces. We are looking for the smallest $c = c^*$ for which $S(\mu, w, c)$ has a non-empty intersection with $\bar{\mathcal{A}}_j^c$. Equivalently, we are looking for the first $c > 0$ for which the set grows beyond the interior of $\mathcal{A}_j$ and intersects with $\bar{\mathcal{A}}_j^c$. Thus,

$$\inf_{\nu\in\bar{\mathcal{A}}_j^c} \sum_{i=1}^{K} w_i KL(\mu_i|\nu_i) = \inf\{c : S(\mu, w, c) \cap \bar{\mathcal{A}}_j^c \neq \emptyset\}.$$

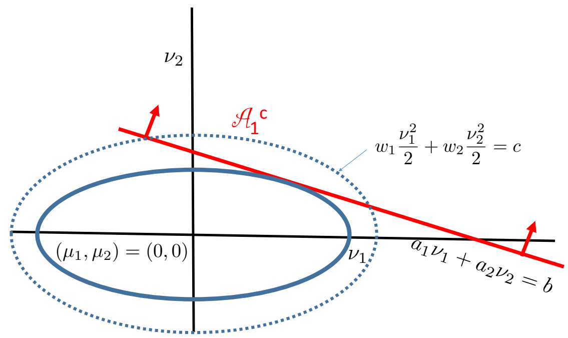

Figure 1: Geometrical view of the sub-problem in LB. Two arms with standard Gaussian distribution. $\mathcal{A}_1$ is a half-space in $\mathbb{R}^2$, and $\mu = (0, 0) \in \mathcal{A}_1$.

Figure 1 demonstrates this in a simple setting of two arms. Arm i, for $i = 1, 2$, is Gaussian distributed with mean $\mu_i = 0$ and variance 1. $\mathcal{A}_1 = \{(\nu_1, \nu_2) \in \mathbb{R}^2 : a_1\nu_1 + a_2\nu_2 < b\}$ for $a_1, a_2, b > 0$, and it contains $(\mu_1, \mu_2) = (0, 0)$. $KL(\mu_i|\nu_i) = \frac{\nu_i^2}{2}$ for $i = 1, 2$. The convex set $S(\mu, w, c)$ is tangential to $a_1\nu_1 + a_2\nu_2 = b$ at $c = c^*$.

Conditions on KL-Divergence: Since Λ_i^* is a convex function, we have that K_i is convex in its first argument. Since $K_i(\mu_i|\nu_i)$ decreases with ν_i for $\nu_i \leq \mu_i$, and it increases with ν_i for $\nu_i \geq \mu_i$, it is a quasi-convex function of ν_i. For many known SPEFs, including Bernoulli, Poisson, Gaussian with known variance and Gamma with known shape parameter, the KL-divergence is also strictly convex in the second argument. But there are also SPEFs for which it is not convex in the second argument, e.g., Rayleigh, centered Laplacian and negative Binomial (with number of failures fixed).

Our analysis is substantially simplified when $\sum_{i=1}^K w_i KL(\mu_i|\nu_i)$ is a strictly convex function of ν. This is ensured by Assumption 1:

Assumption 1 *For each i, $\mathcal{D}_i^o$ is non-empty and $\Lambda_i(\eta_i)$ is strictly convex for $\eta_i \in \mathcal{D}_i^o$. Further, for any $\mu_i \in \mathcal{U}_i$, $K_i(\mu_i|\nu_i)$ is a strictly convex function of $\nu_i \in \mathcal{U}_i$.*

We also make the following assumption to ease some technicalities. This assumption holds for most distributions encountered in practice.

Assumption 2 *For any $\mu_i \in \mathcal{U}_i$, $K_i(\mu_i|\nu_i) \to \infty$ as $\nu_i \to \partial\mathcal{U}_i$ with ν_i taking values in $\mathcal{U}_i$.*

3. Lower bounds for some $\mathcal{PI}$ problems

In this section we explore the structure of Problem LB in a number of settings. In each setting we specify $\mathcal{A}_1$ and $\mathcal{A}_2$, and $\mathcal{Q}$ is set to $\mathcal{A}_1 \cup \mathcal{A}_2$.

3.1 Threshold crossing problem: Let $\mathcal{U} = \times_{i=1}^K \mathcal{U}_i$, $\mathcal{A}_1 = \{\mu \in \mathcal{U} : \max_{i\leq K} \mu_i > u\}$, and $\mathcal{A}_2 = \{\mu \in \mathcal{U} : \max_{i\leq K} \mu_i < u\}$.

Theorem 4 below points to an interesting asymmetry that arises in the lower bound problem associated with threshold crossing as a function $\mu \in \mathcal{Q}$.

Theorem 4 *Suppose that* $(u,\dots,u)\in\mathcal{U}$. *Consider* $\mu\in\mathcal{A}_1$ *such that, w.l.o.g., for some* $i\geq 1$,

$$\mu_j > u \text{ for } j=2,\dots,i,\ \ \mu_j < u \ \text{ for } i+1\leq j\leq K,$$

and $K_1(\mu_1|u) > K_j(\mu_j|u)$ *for* $j=1,\dots,i$. *Then, Problem LB has a unique solution given by*

$$w_1^*=1, \text{ and } w_j^*=0 \text{ for } j=2,\dots,K. \tag{5}$$

The lower bound on expected total number of samples generated equals $\frac{1}{K_1(\mu_1|u)}\times\log(\frac{1}{2.4\delta})$.

When $\mu\in\mathcal{A}_2$, *Problem LB has a unique solution given by*

$$w_j^*\propto 1/K_j(\mu_j|u),\ 1\leq j\leq K, \tag{6}$$

and the lower bound on expected total number of samples generated equals $\sum_{j=1}^K\frac{1}{K_j(\mu_j|u)}\times\log(\frac{1}{2.4\delta})$.

Intuitive explanation for the lower bound asymmetry in the two cases $\mu\in\mathcal{A}_1$ and $\mu\in\mathcal{A}_2$ is as follows: When $\mu\in\mathcal{A}_1$, any algorithm has to establish with at least $1-\delta$ probability that there exists at least one arm above u. The lower bound is then achieved by focussing on the arm that is most separated from u. That is, arm i with $\mu_i>u$ and with the largest value of $K_i(\mu_i|u)$. On the other hand, when $\mu\in\mathcal{A}_2$, any algorithm would need to rule that each and every arm has mean less than u, again while controlling the probability of error for each arm.

In Appendix A, Example 1, we discuss how the threshold crossing problem arises naturally in nested simulation used in financial portfolio risk measurement.

3.2 Half-space problem: We consider the problem of identifying the half-space to which the mean vector belongs. Set $\mathcal{A}_1=\{\nu\in\mathbb{R}^K\cap\mathcal{U}:\sum_{k=1}^K a_k\nu_k<b\}$ and $\mathcal{A}_2=\{\nu\in\mathbb{R}^K\cap\mathcal{U}:\sum_{k=1}^K a_k\nu_k>b\}$. W.l.o.g. each a_i can be taken to be non-zero and $b>0$. Problem LB may be formulated as: For $\mu\in\mathcal{A}_1$, and non-empty $\mathcal{A}_2$,

$$\max_{w\in\mathcal{P}_K}\inf_{\nu\in\bar{\mathcal{A}}_2}\sum_{j=1}^K w_jK_j(\mu_j|\nu_j). \tag{7}$$

Theorem 5 *Under Assumptions 1, 2, and that* $\mathcal{A}_2$ *is non-empty, there is a unique optimal solution* (w^*,ν^*) *to Problem LB. Further,*

$$K_i(\mu_i|\nu_i^*)=K_1(\mu_1|\nu_1^*)\quad\forall i, \tag{8}$$

$$\sum_{k=1}^K a_k\nu_k^*=b, \tag{9}$$

$$\nu_i^*>\mu_i \text{ if } a_i>0, \text{ and } \nu_i^*<\mu_i \text{ if } a_i<0. \tag{10}$$

Relations (8), (9) and (10) uniquely specify $\nu^*\in\mathcal{U}$. *Moreover,*

$$\frac{w_i^*}{a_i}K_i'(\mu_i|\nu_i^*)=\frac{w_1^*}{a_1}K_1'(\mu_1|\nu_1^*)\quad\forall i, \tag{11}$$

where the derivatives are with respect to the second argument.

The proof details are given in the appendix. Ignoring technicalities, the intuition for (8) follows from Sion's Minimax Theorem, which, loosely speaking, implies that (7) equals

$$\inf_{\substack{\nu\in\mathbb{R}^K\cap\mathcal{U}:\\ \sum_{k=1}^K a_k\nu_k\geq b}} \max_{w\in\mathcal{P}_K} \sum_{j=1}^K w_j K_j(\mu_j|\nu_j) = \inf_{\substack{\nu\in\mathbb{R}^K\cap\mathcal{U}:\\ \sum_{k=1}^K a_k\nu_k\geq b}} \max_j K_j(\mu_j|\nu_j).$$

Relations (8), (9) and (10) then follow from KKT conditions applied to RHS above. Uniqueness of ν^* follows as $\max_j K_j(\mu_j|\nu_j)$ is a strictly convex function of ν. Equation (11) corresponds to the slope matching that occurs as the boundary of the sub-level set associated with w^* (see Remark 3) is tangential to the hyperplane $\sum_{k=1}^K a_k\nu_k = b$.

Algorithm to determine ν^* and w^*: Recall that $K_i(\mu_i|\nu_i)$ equals zero at $\nu_i = \mu_i$. It strictly increases with ν_i for $\nu_i \geq \mu_i$ and it strictly reduces with ν_i for $\nu_i \leq \mu_i$. Assume w.l.o.g. that $a_1 > 0$, and for $\nu_1 \geq \mu_1$, consider the function $\nu_i(\nu_1) = K_i^{-1}(K_1(\mu_1|\nu_1))$ where $\nu_i(\nu_1) \geq \mu_i$ if $a_i > 0$, and $\nu_i(\nu_1) \leq \mu_i$ if $a_i < 0$. Now, the function $h(\nu_1) \triangleq \sum_{i=1}^K a_i\nu_i(\nu_1) < b$ for $\nu_1 = \mu_1$ and it strictly increases with ν_1. Further, observe that as $\nu_1 \uparrow \overline{u}_1$, $\nu_i(\nu_1) \uparrow \overline{u}_i$ if $a_i > 0$, and $\nu_i(\nu_1) \downarrow \underline{u}_i$ if $a_i < 0$. Thus, $h(\nu_1) \uparrow \sum_{i=1}^K a_i\hat{u}_i$ and ν_1 can be increased to a unique $\nu^* \in \mathcal{U}$ so that $h(\nu_1^*) = b$, and (8) and (10) hold. (11) can then be used to compute w^*.

3.3 $\mathcal{A}_2$ is a convex set: To avoid undue technicalities, assume that $\mathcal{Q} \subset \mathcal{U}$. Suppose that $\mathcal{A}_2$ is a non-empty closed convex set and $\mu \in \mathcal{A}_1$. Let the associated lower bound problem be denoted by Problem CVX.

$$\max_{w\in\mathcal{P}_K} \inf_{\nu\in\mathcal{A}_2} \sum_{j=1}^K w_j K_j(\mu_j|\nu_j). \qquad \text{(Problem CVX)}$$

Recall that C^* denotes the optimal value for Problem CVX (it is easily seen to be finite). Under Assumption 1, $\sum_{j=1}^K w_j K_j(\mu_j|\cdot)$ is strictly convex and there is a unique $\nu \in \partial\mathcal{A}_2$ that achieves the minimum in the sub-problem $\inf_{\nu\in\mathcal{A}_2}\sum_{j=1}^K w_j K_j(\mu_j|\nu_j)$. Let $\nu(w)$ denote this unique solution for any $w \in \mathcal{P}_K$. Lemma 6 below shows that for every optimal solution to Problem CVX, the same ν achieves the minimum in the above sub-problem.

Lemma 6 *Under Assumption 1, for any w^*, s^* that are optimal for Problem CVX, $\nu(w^*) = \nu(s^*)$.*

Let ν^* be the unique value of ν which achieves the minimum in the sub-problem for every optimal solution. In Theorem 16, we provide an alternate characterization of ν^*, as well as a characterization of the solution of Problem CVX.

Some notation is needed to state Theorem 16. For any index set $\mathcal{J} \subseteq [K]$ and vector $\nu \in \mathbb{R}^K$, let $\nu_{\mathcal{J}}$ denote the projection of the vector ν on to the lower dimensional subspace with coordinate set given by $\mathcal{J}$. Similarly, for any set $\mathcal{B} \subseteq \mathbb{R}^K$, let $\mathcal{B}_{\mathcal{J}}$ denote its projection onto the subspace restricted to the coordinate set $\mathcal{J}$, i.e, $\mathcal{B}_{\mathcal{J}} = \{\nu_{\mathcal{J}} : \nu \in \mathcal{B}\}$. Note that if $\mathcal{B}$ is convex, then $\mathcal{B}_{\mathcal{J}}$ is also convex. If $\mathcal{B}$ is the c-sublevel set of a convex function f, then

$$\mathcal{B}_{\mathcal{J}} = \{\nu_{\mathcal{J}} : f(\nu_{\mathcal{J}},\nu_{\mathcal{J}^c}) \leq c \text{ for some } \nu_{\mathcal{J}^c} \in \mathbb{R}^{|\mathcal{J}^c|}\} = \{\nu_{\mathcal{J}} : \inf_{\nu_{\mathcal{J}^c}\in\mathbb{R}^{|\mathcal{J}^c|}} f(\nu_{\mathcal{J}},\nu_{\mathcal{J}^c}) \leq c\}.$$

In other words, $\mathcal{B}_{\mathcal{J}}$ is the c-sublevel set of the function $h_{\mathcal{J}} := \inf_{\nu_{\mathcal{J}^c}\in\mathbb{R}^{|\mathcal{J}^c|}} f(\nu_{\mathcal{J}},\nu_{\mathcal{J}^c})$.

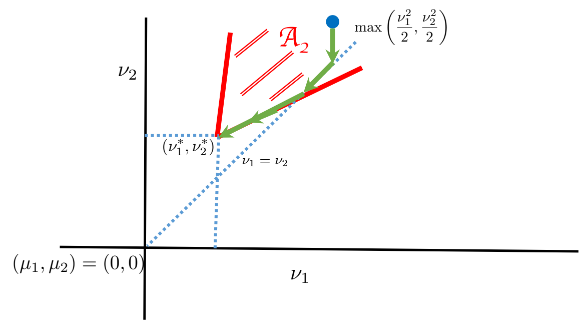

Figure 2: Algorithm to solve (12) in a simple setting of two arms with standard Gaussian distribution. $\mathcal{A}_2$ is a closed convex set and $\mu = (0,0) \in \mathcal{A}_1$. Since $\nu_2^* > \nu_1^*$, $\mathcal{I} = \{2\}$. This suggests that it is optimal to only sample arm 2 to separate $\mu = (0,0)$ from $\mathcal{A}_2$.

Theorem 7 *Suppose that $\mu \in \mathcal{A}_1$, $\mathcal{A}_2$ is non-empty, $\mathcal{Q} \subset \mathcal{U}$, and Assumptions 1 and 2 hold. Then, for any optimal solution (w^*, ν^*) to Problem CVX, the ν^* uniquely solves the min-max problem*

$$\inf_{\nu \in \mathcal{A}_2} \max_i K_i(\mu_i|\nu_i). \tag{12}$$

Further, the following are necessary and sufficient conditions for such an (w^*, ν^*). Let $\mathcal{I} = \operatorname{argmax}_i K_i(\mu_i|\nu_i^*)$. Then,

1. $w_i^* = 0 \quad \forall i \in \mathcal{I}^c$,
2. $\nu_{\mathcal{I}}^* \in \partial(\mathcal{A}_2)_{\mathcal{I}}$, *and*
3. *there exists a supporting hyperplane of $(\mathcal{A}_2)_{\mathcal{I}}$ at $\nu_{\mathcal{I}}^*$ given by $\sum_{i \in \mathcal{I}} a_i \nu_i = b$ such that*

$$\nu_i^* > \mu_i \text{ if } a_i > 0, \text{ and } \nu_i^* < \mu_i \text{ if } a_i < 0 \quad \forall i \in \mathcal{I}, \tag{13}$$

$$\frac{w_i^*}{a_i} K_i'(\mu_i|\nu_i^*) = \frac{w_j^*}{a_j} K_j'(\mu_j|\nu_j^*) \quad \forall i, j \in \mathcal{I}. \tag{14}$$

Problem CVX (and indeed Problem LB) may heuristically be viewed as a game between an optimal algorithm and nature. An algorithm picks a $w \in \mathcal{P}_K$ that provides a recipe for proportionate sampling of different arms. Nature then selects a $\nu \in \mathcal{A}_2$ that for a given w minimizes $\sum_{j=1}^K w_j K_j(\mu_j|\nu_j)$, and hence for the algorithm is the most difficult to separate from μ. The algorithm looks for a w that maximizes this minimum separation. *Theorem 7 makes an interesting observation that for convex $\mathcal{A}_2$, the algorithm has the option of not sampling some arms. Maximum separation may be obtained by focusing on a subset of arms and showing that they are well separated from the projection of $\mathcal{A}_2$ along the subspace associated with these arms.*

Condition (3) in Theorem 7 highlights the fact that along the projected space, finding a solution to Problem CVX is equivalent to finding a solution to an appropriate half-space problem that is tangential to the projected convex set.

Remark 8 Since $\max_i K_i(\mu_i|\nu_i)$ is a strictly convex function of ν, (12) shows that Problem CVX maybe solved for ν^* using any standard convex programming solver. Remark 9 below emphasizes the point that w^* is easily calculated once ν^* is known, if there is a unique supporting hyperplane in $\mathbb{R}^{|\mathcal{I}|}$, of $(\mathcal{A}_2)_{\mathcal{I}}$ at $\nu^*_{\mathcal{I}}$.

Figure 2 demonstrates how a steepest descent based procedure may work to solve (12) in a simple setting of two arms. Arm i, for $i = 1,2$, is Gaussian distributed with mean $\mu_i = 0$ and variance 1. $KL(\mu_i|\nu_i) = \frac{\nu_i^2}{2}$ for $i = 1,2$. The algorithm starts at a point $(\nu_1, \nu_2) \in \mathcal{A}_2$ with $\nu_2 > \nu_1$. Steepest descent direction to minimize $\max_{i=1,2} \frac{\nu_i^2}{2}$ corresponds to reducing ν_2 until $\nu_2 = \nu_1$. It then corresponds to decending along the direction $\nu_1 = \nu_2$, until boundary of $\mathcal{A}_2$ is hit. In Figure 2, the algorithm continues to descend along the boundary reducing the value of $\max_{i=1,2} \frac{\nu_i^2}{2}$ until the optimal point (ν_1^*, ν_2^*). Since $\nu_2^* > \nu_1^*$, we have $\mathcal{I} = \{2\}$. Thus, the lower bound analysis suggests that it is optimal to only sample arm 2 to separate $\mu = (0,0)$ from $\mathcal{A}_2$.

Remark 9 Condition 3 shows that the problem has a unique solution, i.e., the optimal w^* is a singleton, if there is a unique supporting hyperplane of $(\mathcal{A}_2)_{\mathcal{I}}$ at $\nu^*_{\mathcal{I}}$. Consider the case where $\mathcal{A}_2 = \{\nu : f(\nu) \leq c\}$ is the c-sublevel set of a convex function f. Then, $(\mathcal{A}_2)_{\mathcal{I}}$ is the c-sublevel set of the function $h : \Re^{|\mathcal{I}|} \to \Re, h(\nu_{\mathcal{I}}) := \inf_{\nu_{\mathcal{I}^c} \in \mathbb{R}^{|\mathcal{I}^c|}} f(\nu_{\mathcal{I}}, \nu_{\mathcal{I}^c})$. Further suppose that $h(\cdot)$ is a smooth function. Then, the unique tangential hyperplane at $\nu^*_{\mathcal{I}}$ is given by $\nabla h(\nu^*_{\mathcal{I}})^\intercal(\nu_{\mathcal{I}} - \nu^*_{\mathcal{I}}) = 0$. In particular, in this case for $i \in \mathcal{I}$, $w_i^* \propto \frac{\frac{\partial h}{\partial \nu_i}(\nu^*_{\mathcal{I}})}{K_i'(\mu_i|\nu_i^*)}$.

3.4 $\mathcal{A}_1$ is a polytope In Section 3.3, $\mathcal{A}_2$ is convex, while $\mathcal{A}_1$ need not be. This allowed us to explicitly characterize the solution to the lower bound problem. We now briefly consider the case where $\mathcal{A}_1$ is convex, and $\mathcal{A}_2$ need not be. Specifically, we examine the case where $\mathcal{A}_1$ is a polytope so that $\mathcal{A}_2$ is a union of half-spaces. Just as the single half-space problem was useful in studying the case where $\mathcal{A}_2$ is convex, analyzing $\mathcal{A}_2$ when it is a union of half-spaces, may provide insights to a more general problem where $\mathcal{A}_2$ is a union of convex sets. The latter may be an interesting area for future research.

Let

$$\mathcal{B}_j \triangleq \{\nu \in \mathbb{R}^K : \sum_{k=1}^{K} a_{j,k}\nu_k \geq b_j\}, \tag{15}$$

each $b_j \geq 0$, and $\mathcal{A}_2 = \cup_{j=1}^m \mathcal{B}_j$ be the union of these half-spaces. To ease technicalities, suppose that $\mathcal{U} = \mathbb{R}^K$. The lower bound problem may be expressed as

$$C^*(\mu) = \max_{w \in \mathcal{P}_K} \inf_{\nu \in \cup_{j=1}^m \mathcal{B}_j} \sum_{i=1}^{K} w_i K_i(\mu_i|\nu_i). \tag{16}$$

Let $\mathcal{W}(\mu)$ denote the optimal solution set. Lemma 10 shows that the optimization problem in (16) has a unique solution, that is, $\mathcal{W}(\mu)$ is a singleton.

Lemma 10 *There is a unique $w \in \mathcal{P}_K$ that achieves the maximum in* (16).

Remark 11 It is easy to see that the **best arm identification problem** is a special case of this problem. To see this, suppose arm 1 has the highest mean among the K arms, i.e., $\mu_1 > \mu_j \ \forall j \neq 1$. We then have $\mathcal{A}_2 = \cup_{j=2}^K \mathcal{B}_j$, where for any j, $\mathcal{B}_j = \{\nu \in \mathbb{R}^K : \nu_j \geq \nu_1\}$.

Observe that $\inf_{\nu \in \mathcal{A}_2} \sum_{i=1}^K w_i K_i(\mu_i|\nu_i)$ being an infimum of linear functions of w, is a concave function of w, for any $\mathcal{A}_2$. Thus, standard gradient descent methods can be used to solve (16), once an algorithm exists for solving $g(\mu, w) \triangleq \inf_{\nu \in \cup_{j=2}^m \mathcal{B}_j} \sum_{i=1}^K w_i K_i(\mu_i|\nu_i)$. This is straightforward as

$$\inf_{\nu \in \cup_{j=2}^m \mathcal{B}_j} \sum_{i=1}^K w_i K_i(\mu_i|\nu_i) = \min_{j \leq m} \inf_{\nu \in \mathcal{B}_j} \sum_{i=1}^K w_i K_i(\mu_i|\nu_i).$$

Thus, one may solve the strictly convex problem $g_j(\mu, w) \triangleq \inf_{\nu \in \mathcal{B}_j} \sum_{i=1}^K w_i K_i(\mu_i|\nu_i)$ for each j and set $g(\mu, w) = \min_{j \leq m} g_j(\mu, w)$. An algorithm for numerically solving for $g_j(\mu, w)$ is easily designed and is given in Appendix C. An outline of a simple algorithm to compute $C^*(\mu)$ is as follows:

(i) Given a w, for each j, solve the strictly convex optimization problem $\inf_{\nu \in \mathcal{B}_j} \sum_{i=1}^K w_i K_i(\mu_i|\nu_i)$ to determine $g_j(\mu, w)$.

(ii) Compute $g(\mu, w) = \min_{j \leq m} g_j(\mu, w)$. Use a numerical procedure to determine the gradient of $g(\mu, w)$ with respect to w. Update w using any version of gradient-descent, and repeat.

In Appendix C.0.1, we restrict ourselves to two arms, both having a Gaussian distribution with known and common variance. This simple setting lends itself to elegant comprehensive analysis and a graphical interpretation.

4. An asymptotically optimal algorithm

In this section, we outline a δ-correct algorithm (Algorithm 1) for the $\mathcal{PI}$ problem which, under mild conditions, achieves asymptotically optimal mean termination time as $\delta \to 0$. Both the algorithm and its analysis closely follow the best arm identification in Garivier and Kaufmann (2016). The sampling rule used in the algorithm (described below) is inspired by the lower bound Problem LB. The stopping rule follows from the generalised likelihood ratio method (see Chernoff (1959)).

In Problem LB, let $\mathcal{W}(\mu)$ and $C^*(\mu)$, respectively denote the optimal solution set and optimal value. Let $V(\mu, w)$ and $g(\mu, w)$, respectively denote the optimal solution set and optimal value of the inner sub-problem. We consider settings where Problem LB has a unique optimal solution. That is, $|\mathcal{W}(\mu)| = 1$. As seen in Section 3, for threshold crossing, half space problem and the polytope problem, Problem LB has a unique optimal solution. When $\mathcal{A}_2$ is a closed convex set and the associated $\nu^* \in \partial\mathcal{A}_2$ is a smooth point (with a unique supporting hyperplane), then Problem LB again has a unique optimal solution. Lemma 13 below, shows that solution to Problem LB, $\mathcal{W}(\mu)$, is continuous function of μ when $|\mathcal{W}(\mu)| = 1$, an important requirement in proving Theorem 12. Recall that in Problem LB, $\mu \in \mathcal{A}_j$.

Sampling Rule: The essential idea is to draw samples according to estimated optimal sampling ratios obtained by solving Problem LB with empirical means substituting the true means. In other words, if $\hat{\mu}(t)$ is the vector of empirical means of the arms at time t, an arm is chosen to bring the ratio of total number of samples for all the arms closer to an optimal ratio $\hat{w}(t) \in \mathcal{W}(\hat{\mu}(t))$. But this simple strategy may result in erroneously giving too few samples to an arm due to initial bad estimates preventing convergence to the correct value in subsequent sample allocations. This difficulty can be dealt with through forced exploration for each arm to ensure sufficiently fast convergence.

Garivier and Kaufmann (2016) propose a 'D-Tracking' rule along these lines for the best arm problem that ensures convergence to the correct sampling ratio. We also use this rule as the sampling

rule in our algorithm. The rule can be described as follows. Let $N_i(t)$ denote the number of samples of arm i at sampling step t for all i and let $\hat{w}(t) \in \mathcal{W}(\hat{\mu}(t))$. If there exists an arm i such that $N_i(t) < \sqrt{t} - K/2$, choose that arm. Otherwise, choose an arm that has the maximum difference between the estimated optimal ratio and the actual fraction of samples, i.e., an arm is chosen from $\mathrm{argmax}_i\, \hat{w}_i(t) - N_i(t)/t$. This sampling rule has the following properties:
(i) each arm gets $\Omega(\sqrt{t})$,
(ii) if the estimated sampling ratios $\mathcal{W}(\hat{\mu}(t))$ converge to an optimal ratio $\mathcal{W}(\mu)$, then the actual fraction of samples also converges to the same optimal ratio.

Stopping Rule: Let *threshold function* $\beta(t,\delta) = \log\left(\frac{c\, t^2 \log(1/\delta)^{2K+1}}{\delta}\right)$, where c is an appropriately chosen constant. The term $\log(1/\delta)^{2K+1}$ in this expression was not used by Garivier and Kaufmann (2016), as they ignore a technicality in proving δ-correctness of the algorithm (see proof of Theorem 12 in Appendix E).

The stopping rule uses a threshold rule that imitates the lower bound (1). It first finds the partition in which the empirical mean vector $\hat{\mu}(t)$ lies. Denote this partition after generating t samples by $\mathcal{A}(t)$. If $\inf_{\nu \in \mathcal{A}^c(t)} \sum_i N_i(t) K_i(\hat{\mu}_i(t)|\nu_i) \geq \beta(t,\delta)$, then it stops and declares $\mathcal{A}(t)$ as the partition containing μ. Else, it continues to sample arms according to the D-Tracking rule.

Algorithm 1 Algorithm for one parameter exponential families

Sample each arm once. Set $\hat{\mu}(0)$ to the observed sample average of each arm. Set $t = 1$
At sample t,
Compute weights $w(\hat{\mu}(t-1))$ and sample according to D-Tracking rule ▷ Sampling Rule
Let $\hat{\mu}(t) \in \mathcal{A}(t)$.
If $\inf_{\nu \in \mathcal{A}^c(t)} \sum_i N_i(t) K_i(\hat{\mu}_i(t)|\nu_i) \geq \beta(t,\delta)$ **then** ▷ Termination Rule
Declare $\mu \in \mathcal{A}(t)$.
end if
Else Increment t by 1 and continue.

Sample complexity analysis: Let $T_U(\delta)$ be the time at which Algorithm 1 terminates. Then we have the following guarantee.

Theorem 12 *Suppose that $\mathcal{Q} \subset \mathcal{U}$ and Assumptions 1 and 2 hold. If Problem LB has a unique optimal solution, i.e., if $|\mathcal{W}(\mu)| = 1$, then Algorithm 1 is a δ-correct algorithm with* $\limsup_{\delta \to 0} \frac{\mathbb{E}[T_U(\delta)]}{\log\left(\frac{1}{\delta}\right)} \leq C^*(\mu)^{-1}$.

The proof of Theorem 12 is along the lines of Garivier and Kaufmann (2016), and is given in Appendix D for completeness. The following continuity result is critical to the proof.

Lemma 13 *Under conditions of Theorem 12, the function g is continuous at (μ, w) for any $w \in \mathcal{P}_K$. Further, if Problem LB has a unique optimal solution, then this solution is continuous at μ.*

For notational ease, let $\mathcal{A}$ denote $\mathcal{A}_j$ and $\mathcal{A}^c$ denote $\mathcal{Q} - \mathcal{A}_j$. If $\bar{\mathcal{A}}^c$ is compact, then Theorem 2.1 in Fiacco and Ishizuka (1990) implies that g is continuous at (μ, w). Continuity of the optimal solution $\mathcal{W}(\mu)$ to Problem LB at μ when $\mathcal{W}(\mu)$ is a singleton also follows from Theorem 2.2 in Fiacco and Ishizuka (1990). The details for general $\bar{\mathcal{A}}^c$ are given in Appendix D. A small simulation experiment illustrating that the non-asymptotic performance of the proposed algorithm is not much worse than the asymptotic limit is given in Appendix E.

References

Albert, A.E., The sequential design of experiments for infinitely many states of nature *The Annals of Mathematical Statistics*, pp.774-799. 1961.

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.

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.

Mark Broadie, Yiping Du, and Ciamac C Moallemi. Efficient risk estimation via nested sequential simulation. *Management Science*, 57(6):1172–1194, 2011.

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.

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.

Amir Dembo and Ofer Zeitouni. Large deviations techniques and applications. 1998. *Applications of Mathematics*, 38, 2011.

Darrell Duffie. *Dynamic asset pricing theory*. Princeton University Press, 2010.

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, Dec 1990.

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

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 and multi-armed bandits. *arXiv:1507.04564v3*, 1, 2018.

Michael B Gordy and Sandeep Juneja. Nested simulation in portfolio risk measurement. *Management Science*, 56(10):1833–1848, 2010.

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

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.

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 and Shivaram Kalyanakrishnan. Information complexity in bandit subset selection. In *Conference on Learning Theory*, pages 228–251, 2013.

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.

Emilie Kaufmann, W. M. Koolen, and Aurélien Garivier. Sequential test for the lowest mean: From Thompson to Murphy sampling. In *Advances in Neural Information Processing Systems*, pages 6332–6342, 2018.

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.

Marco López and Georg Still. Semi-infinite programming. *European Journal of Operational Research*, 180(2):491–518, 2007.

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*, 5:623–648, 2004.

Edward Paulson. 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.

Walter Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Co., New York, third edition, 1976. International Series in Pure and Applied Mathematics.

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

S. E. Shreve. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer Science & Business Media, 2004.

Acknowledgments

We thank the referees for their helpful comments.

Appendix A. Threshold Crossing Problem

In this section first in Example 1, we discuss how the threshold crossing problem arises naturally in nested simulation used in financial portfolio risk measurement. We then prove Theorem 4.

Example 1 Consider the problem of measuring tail risk in a portfolio comprising financial derivatives. The key property of a financial derivative is that as a function of underlying stock prices or other financial instruments, it's value is a conditional expectation (see, e.g., Duffie (2010), Shreve (2004)). Thus, the value of a portfolio of financial securities that contains financial derivatives can also be expressed as a conditional expectation given the value of underlying financial instruments.

Suppose that $(X_1, \ldots, X_K)$, where each X_t is a vector in a Euclidean space, denote the macroeconomic variables and financial instruments at time t, such as prevailing interest rates, stock index value and stock prices, on which the value of a portfolio depends. For notational convenience we have assumed that times take integer values.

Portfolio loss amount at any time t is a function of $\mathcal{X}_t \triangleq (X_1, \ldots, X_t)$ and is given by $E(Y_t|\mathcal{X}_t)$ for some random variable Y_t (see, e.g. Gordy and Juneja (2010), Broadie et al. (2011) for further discussion on portfolio loss as a conditional expectation, and the need for nested simulation). The quantity $E(Y_t|\mathcal{X}_t)$ is not known, however, conditional on $\mathcal{X}_t$, independent samples of Y_t can be generated via simulation. Our interest is in estimating the probability that the portfolio loss by time K exceeds a large threshold u or

$$\gamma \triangleq P(\max_{1\leq t\leq K} Z_t \geq u), \tag{17}$$

where $Z_t = E(Y_t|\mathcal{X}_t)$.

These probabilities typically do not have a closed form expression and are estimated using Monte Carlo simulation. An algorithm to estimate this probability maybe nested and is given as follows:

1. Repeat the outer loop iterations for $1 \leq j \leq n$.
2. At outer loop iteration j, generate through Monte Carlo a sample of underlying factors $(X_{1,j}, \ldots, X_{K,j})$.
3. Given this sample, we need to ascertain whether

$$W_j \triangleq \max_{1\leq t\leq K} Z_{t,j} \geq u,$$

where $Z_{t,j} = E(Y_t|\mathcal{X}_{t,j})$. This fits our framework of threshold crossing problem where we may sequentially generate conditionally independent samples of Y_t for each t conditional on $(X_{1,j}, \ldots, X_{t,j})$ and arrive at an indicator $\hat{W}_j$ that equals W_j with probability $\geq 1 - \delta$.

Then,

$$\hat{\gamma}_n(\Delta) \triangleq \frac{1}{n} \sum_{j=1}^{n} \hat{W}_j$$

denotes our estimator for γ. There are interesting technical issues related to optimally distributing computational budget in deciding the number of samples in the outer loop, in the inner loop and the value of δ to be selected. These issues, however, are not addressed in the paper and may be a topic for future research.

Proof of Theorem 4: To see (5), first observe that due to continuity of each $K_j(\mu_j|\nu_j)$ as a function of $\nu_j \in \mathcal{U}_j$, we have

$$\inf_{\nu \in \mathcal{A}_2} \sum_{j=1}^{K} w_j K_j(\mu_j|\nu_j) = \inf_{\nu \in \bar{\mathcal{A}}_2} \sum_{j=1}^{K} w_j K_j(\mu_j|\nu_j),$$

where recall that for any set $\mathcal{A}$, $\bar{\mathcal{A}}$ denotes its closure. The RHS above is solved by

$$\nu = (u, \ldots, u, \mu_{i+1}, \ldots, \mu_k)$$

in the sense that for any other $\tilde{\nu} \in \bar{\mathcal{A}}_2$,

$$\sum_{j=1}^{K} w_j K_j(\mu_j|\tilde{\nu}_j) \geq \sum_{j=1}^{K} w_j K_j(\mu_j|\nu_j) = \sum_{j=1}^{i} w_j K_j(\mu_j|u).$$

Our lower bound problem reduces to

$$\max_{w \in \mathcal{P}_K} \sum_{j=1}^{i} w_j K_j(\mu_j|u).$$

This can easily be seen to be solved uniquely by $w_1^* = 1$, $w_j^* = 0$ for $j = 2, \ldots, K$, and the optimal value C^* is $K_1(\mu_1|u)$. The lower bound on the overall expected number of samples generated is then given by $\log(\frac{1}{2.4\delta})/C^*$.

To see (6), observe that to simplify $\inf_{\nu \in \bar{\mathcal{A}}_1} \sum_{j=1}^{K} w_j K_j(\mu_j|\nu_j)$, it suffices to consider $\nu(s) \in \bar{\mathcal{A}}_1$ for each s ($1 \leq s \leq K$) where

$$\nu(s) \triangleq (\mu_1, \ldots, \mu_{s-1}, u, \mu_{s+1}, \ldots, \mu_k),$$

in the sense that for any $\nu \in \bar{\mathcal{A}}_1$

$$\sum_{j=1}^{K} w_j K_j(\mu_j|\nu_j) \geq \min_{s=1,\ldots,K} \sum_{j=1}^{K} w_j K_j(\mu_j|\nu_j(s)) = \min_{s=1,\ldots,K} w_s K_s(\mu_s|u).$$

The lower bound problem then reduces to

$$\max_{w \in \mathcal{P}_K} \min_j w_j K_j(\mu_j|u).$$

The solution to this problem is given by

$$w_j^* \propto 1/K_j(\mu_j|u) \; \forall j,$$

and the optimal value C^* is $\left(\sum_{j=1}^K \frac{1}{K_j(\mu_j|u)}\right)^{-1}$. The lower bound on the overall expected number of samples generated is equal to $\log(\frac{1}{2.4\delta})/C^*$. □

Appendix B. The half space lower bound problem

In this section we restate (to aid readability) Theorem 5 as Theorem 14 and prove it. We also state and prove Lemma 15 needed for proof of Theorem 14.

Theorem 14 *Under Assumptions 1, 2, and that $\mathcal{A}_2$ is non-empty, there is a unique optimal solution (w^*, ν^*) to Problem LB. Further,*

$$K_i(\mu_i|\nu_i^*) = K_1(\mu_1|\nu_1^*) \quad \forall i, \tag{18}$$

$$\sum_{k=1}^K a_k \nu_k^* = b, \tag{19}$$

$$\nu_i^* > \mu_i \text{ if } a_i > 0, \text{ and } \nu_i^* < \mu_i \text{ if } a_i < 0. \tag{20}$$

Relations (18), (19) and (20) uniquely specify $\nu^* \in \mathcal{U}$. Moreover,

$$\frac{w_i^*}{a_i} K_i'(\mu_i|\nu_i^*) = \frac{w_1^*}{a_1} K_1'(\mu_1|\nu_1^*) \quad \forall i, \tag{21}$$

where the derivatives are with respect to the second argument.

Let $\overline{u}_i = \sup\{u \in \mathcal{U}_i\}$, and $\underline{u}_i = \inf\{u \in \mathcal{U}_i\}$. Further, set

$$\hat{u}_i = \overline{u}_i \text{ if } a_i > 0, \text{ and } \hat{u}_i = \underline{u}_i \text{ if } a_i < 0.$$

The following lemma is useful in proving Theorem 14.

Lemma 15 *Under Assumption 2, the following are equivalent*

1. *$\mathcal{A}_2 \neq \emptyset$.*
2. *$\sum_{i=1}^K a_i \hat{u}_i > b$.*
3. *There exists a unique $\nu^* \in \mathcal{U}$ such that (18), (19) and (20) hold.*

Proof [Proof of Lemma 15:]

Claim 1 implies existence of ν such that $\sum_{i=1}^K a_i \nu_i > b$ and $K_i(\mu_i|\nu_i) < \infty$ for all i. Claim 2 follows as

$$\sum_{i=1}^K a_i \nu_i < \sum_{i=1}^K a_i \hat{u}_i.$$

To see that Claim 2 implies Claim 3, recall that $K_i(\mu_i|\nu_i)$ equals zero at $\nu_i = \mu_i$. It strictly increases with ν_i for $\nu_i \geq \mu_i$ and it strictly reduces with ν_i for $\nu_i \leq \mu_i$.

Assume w.l.o.g. that $a_1 > 0$, and for $\nu_1 \geq \mu_1$, consider the function

$$\nu_i(\nu_1) = K_i^{-1}(K_1(\mu_1|\nu_1))$$

where $\nu_i(\nu_1) \geq \mu_i$ if $a_i > 0$, and $\nu_i(\nu_1) \leq \mu_i$ if $a_i < 0$. Now, the function

$$h(\nu_1) \triangleq \sum_{i=1}^{K} a_i \nu_i(\nu_1) < b$$

for $\nu_1 = \mu_1$ and it strictly increases with ν_1.

Further, observe that as $\nu_1 \uparrow \overline{u}_1$, $\nu_i(\nu_1) \uparrow \overline{u}_i$ if $a_i > 0$, and $\nu_i(\nu_1) \downarrow \underline{u}_i$ if $a_i > 0$. Thus, $h(\nu_1) \uparrow \sum_{i=1}^{K} a_i \hat{u}_i$ and thus there exists a unique $\nu^* \in \mathcal{U}$ so that $h(\nu_1^*) = b$, and (18) and (20) hold.

To see that Claim 3 implies Claim 1, observe that Claim 3 guarantees that

$$(\nu_1^*, \nu_2(\nu_1^*), \ldots, \nu_K(\nu_1^*)) \in \mathcal{U}$$

By selecting $\nu_1 > \nu_1^*$ and sufficiently small, Claim 1 follows. ■

Proof of Theorem 14:

Lemma 15 guarantees the existence of ν^*, w^* that solve (18), (19), (20) and (21). Here, observe that (21) defines w^*.

Note that ν^* is the solution to the optimization problem:

$$\inf_{\nu \in \bar{\mathcal{A}}_2} \sum_{j=1}^{K} w_j^* K_j(\mu_j|\nu_j).$$

This can be verified by observing that the first order KKT conditions for this convex programmimg problem are given by (19), (20) and (21). (recall that $\bar{\mathcal{A}}_2 = \{\nu : \sum_{i=1}^{K} a_i \nu_i \geq b\}$). Further, from (18), it follows that

$$\inf_{\nu \in \bar{\mathcal{A}}_2} \sum_{j=1}^{K} w_j^* K_j(\mu_j|\nu_j) = \sum_{j=1}^{K} w_j^* K_j(\mu_j|\nu_j^*) = K_1(\mu_1|\nu_1^*).$$

For any another feasible solution $\tilde{w}$, we have

$$\inf_{\nu \in \bar{\mathcal{A}}_2} \sum_{i=1}^{K} \tilde{w}_i K_i(\mu_i|\nu_i) \leq \sum_{i=1}^{K} \tilde{w}_i K_i(\mu_i|\nu_i^*) \leq K_1(\mu_1|\nu_1^*),$$

which shows that w^* is an optimal solution to the problem.

Uniqueness: It remains to show that above is a unique solution to (Problem LB). We skip the details as, in Section 3, Lemma 10, we prove uniqueness of the solution for a general case where $\bar{\mathcal{A}}_2$ is a union of half-spaces. □

3.3 Lower bounds when $\mathcal{A}_2$ is convex:

In this section we prove the results stated in Section 3.3. We first prove Lemma 6. We then restate Theorem 7 as Theorem 16 and prove it.

Proof of Lemma 6: First note that $\inf_{\nu\in\mathcal{A}^c}\sum_{j=1}^{K} w_j K_j(\mu_j|\nu_j)$ is a concave function of w. This shows that, if w^* and s^* are two optimal solutions, then $\alpha w^* + (1-\alpha)s^*$ for $\alpha \in (0,1)$ is another optimal solution. Since it is optimal, we have

$$\sum_{j=1}^{K}(\alpha w_j^* + (1-\alpha)s_j^*)K_j(\mu_j|\nu_j(\alpha w^* + (1-\alpha)s^*)) = C^*.$$

Now due to Assumption 1,

$$\sum_{j=1}^{K} w_j^* K_j(\mu_j|\nu_j(\alpha w^* + (1-\alpha)s^*)) > C^*$$

if $\nu(\alpha w^* + (1-\alpha)s^*) \neq \nu(w^*)$ and

$$\sum_{j=1}^{K} s_j^* K_j(\mu_j|\nu_j(\alpha w^* + (1-\alpha)s^*)) > C^*$$

if $\nu(\alpha w^* + (1-\alpha)s^*) \neq \nu(s^*)$, it follows that $\nu(w^*) = \nu(\alpha w^* + (1-\alpha)s^*) = \nu(s^*)$. □

Theorem 16 *Suppose that $\mu \in \mathcal{A}_1$, $\mathcal{A}_2$ is non-empty, $\mathcal{Q} \subset \mathcal{U}$, and Assumptions 1 and 2 hold. Then, for any optimal solution (w^*, ν^*) to Problem CVX, the ν^* uniquely solves the min-max problem*

$$\inf_{\nu\in\mathcal{A}_2} \max_i K_i(\mu_i|\nu_i). \tag{22}$$

Further, the following are necessary and sufficient conditions for such an (w^*, ν^*). Let $\mathcal{I} = \operatorname{argmax}_i K_i(\mu_i|\nu_i^*)$. Then,

1. $w_i^* = 0 \quad \forall i \in \mathcal{I}^c$,
2. $\nu_{\mathcal{I}}^* \in \partial(\mathcal{A}_2)_{\mathcal{I}}$, *and*
3. *there exists a supporting hyperplane of $(\mathcal{A}_2)_{\mathcal{I}}$ at $\nu_{\mathcal{I}}^*$ given by $\sum_{i\in\mathcal{I}} a_i\nu_i = b$ such that*

$$\nu_i^* > \mu_i \text{ if } a_i > 0, \text{ and } \nu_i^* < \mu_i \text{ if } a_i < 0 \quad \forall i \in \mathcal{I}, \tag{23}$$

$$\frac{w_i^*}{a_i}K_i'(\mu_i|\nu_i^*) = \frac{w_j^*}{a_j}K_i'(\mu_j|\nu_j^*) \quad \forall i,j \in \mathcal{I}. \tag{24}$$

Proof [Proof of Theorem 16]

Let $\mathcal{B}_n$ denote a closed ball centered at μ with radius n. Consider n sufficiently large so that $\tilde{\nu}$ defined as the solution to (22) lies in $\mathcal{B}_n$ (since the objective function $\max_i K_i(\mu_i|\nu_i)$ is strictly convex in ν, such a $\tilde{\nu}$ is unique).

Since $\mathcal{A}_2 \cap \mathcal{B}_n$ is a compact set, and $\sum_{i=1}^K w_i K_i(\mu_i|\nu_i)$ is continuous in w and ν and concave in $w \in \mathcal{P}_K$ and convex in $\nu \in \mathcal{A}_2 \cap \mathcal{B}_n$, by Sion's Minimax Theorem

$$\begin{aligned}\max_{w\in\mathcal{P}_K} \inf_{\nu\in\mathcal{A}_2\cap\mathcal{B}_n} \sum_{i=1}^K w_i K_i(\mu_i|\nu_i) &= \inf_{\nu\in\mathcal{A}_2\cap\mathcal{B}_n} \max_{w\in\mathcal{P}_K} \textstyle\sum_{i=1}^K w_i K_i(\mu_i|\nu_i) \\ &= \inf_{\nu\in\mathcal{A}_2\cap\mathcal{B}_n} \max_i K_i(\mu_i|\nu_i) \\ &= \inf_{\nu\in\mathcal{A}_2} \max_i K_i(\mu_i|\nu_i). \end{aligned} \tag{25}$$

Observe that

$$r_n(w) \triangleq \inf_{\nu\in\mathcal{A}_2\cap\mathcal{B}_n} \sum_{i=1}^K w_i K_i(\mu_i|\nu_i)$$

is continuous in w (see Theorem 2.1 in Fiacco and Ishizuka (1990)) and decreases with n to $r(w) \triangleq \inf_{\nu\in\mathcal{A}_2} \sum_{i=1}^K w_i K_i(\mu_i|\nu_i)$. Thus, we have uniform convergence (see Theorem 7.13 in Rudin (1976))

$$\sup_{w\in\mathcal{P}_K} |r_n(w) - r(w)| \to 0.$$

This in turn implies that

$$\max_{w\in\mathcal{P}_K} r_n(w) \to \max_{w\in\mathcal{P}_K} r(w).$$

From (25) it follows that LHS above is independent of n. Therefore, the min-max relation

$$\max_{w\in\mathcal{P}_K} \inf_{\nu\in\mathcal{A}_2} \sum_{i=1}^K w_i K_i(\mu_i|\nu_i) = \inf_{\nu\in\mathcal{A}_2} \max_i K_i(\mu_i|\nu_i) \tag{26}$$

holds.

Now if (w^*, ν^*) is a saddlepoint of the min-max problem, and since ν^* is unique, it equals $\tilde{\nu}$.

Necessity of conditions on optimal (w^*, ν^*)**:** Let $\mathcal{I} = \operatorname{argmax}_i K_i(\mu_i|\nu_i^*)$. The minimax equality in (26) shows that (w^*, ν^*) is a saddle point, and therefore, w^* solves the optimization problem

$$\max_{(w_1,\ldots,w_K)\in\mathcal{P}_K} \sum_{j=1}^K w_j K_j(\mu_j|\nu_j^*). \tag{27}$$

From this, it is easy to see that $w_i^* = 0 \; \forall i \in \mathcal{I}^c$.

To see 2, note that ν^* uniquely solves the optimization problem

$$\min_{(\nu_1,\ldots,\nu_K)\in\mathcal{A}_2} \sum_{j=1}^K w_j^* K_j(\mu_j|\nu_j). \tag{28}$$

If $\nu_{\mathcal{I}}^*$ is in the interior of $(\mathcal{A}_2)_{\mathcal{I}}$, it is easy to come up with $\nu \neq \nu^*$ on $\partial\mathcal{A}_2$, with a smaller value of $\sum_{j=1}^K w_j^* K_j(\mu_j|\nu_j)$.

Now, consider the convex set

$$\mathcal{C} := \left\{ \nu_{\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}|} : \sum_{i\in\mathcal{I}} w_i^* K_i(\mu_i|\nu_i) < \sum_{i\in\mathcal{I}} w_i^* K_i(\mu_i|\nu_i^*) \right\}$$

(convexity of $\mathcal{C}$ follows from Assumption 1). By the separating hyperplane theorem, there exists a hyperplane $\sum_{i\in\mathcal{I}} a_i\nu_i = b$ that separates $\mathcal{C}$ and $(\mathcal{A}_2)_{\mathcal{I}}$. Since $\nu_{\mathcal{I}}^* \in \partial\mathcal{C} \cap \partial(\mathcal{A}_2)_{\mathcal{I}}$, this hyperplane passes through $\nu_{\mathcal{I}}^*$, and is a supporting hyperplane to both convex sets $\mathcal{C}$ and $(\mathcal{A}_2)_{\mathcal{I}}$. From the fact that it is a supporting hyperplane to $\mathcal{C}$ at $\nu_{\mathcal{I}}^*$, we have

$$\frac{w_i^*}{a_i} K_i'(\mu_i|\nu_i^*) = \frac{w_j^*}{a_j} K_i'(\mu_j|\nu_j^*) \quad \forall i,j \in \mathcal{I}.$$

This proves Condition 3.

Sufficiency: Let ν^* and w^* be such that Conditions 1, 2, 3 hold. Note that $\sum_{i\in\mathcal{I}} a_i\mu_i < b$ and $(\mathcal{A}_2)_{\mathcal{I}} \subseteq \{\nu_{\mathcal{I}} : \sum_{i\in\mathcal{I}} a_i\nu_i \geq b\}$. Then, from Theorem 14, $w_{\mathcal{I}}^*$ and $\nu_{\mathcal{I}}^*$ solve the following half space problem in the lower dimensional subspace restricted to coordinate set $\mathcal{I}$:

$$\max_{w_{\mathcal{I}}\in\mathcal{P}_{\mathcal{I}}} \inf_{\nu_{\mathcal{I}}:\sum_{i\in\mathcal{I}} a_i\nu_i\geq b} \sum_{i\in\mathcal{I}} w_i K_i(\mu_i|\nu_i).$$

In particular,

$$\inf_{\nu_{\mathcal{I}}:\sum_{i\in\mathcal{I}} a_i\nu_i\geq b} \sum_{i\in\mathcal{I}} w_i^* K_i(\mu_i|\nu_i) = \sum_{i\in\mathcal{I}} w_i^* K_i(\mu_i|\nu_i^*).$$

Further, for any $w_{\mathcal{I}}$, note that

$$\inf_{\nu_{\mathcal{I}}\in(\mathcal{A}_2)_{\mathcal{I}}} \sum_{i\in\mathcal{I}} w_i K_i(\mu_i|\nu_i) \geq \inf_{\nu_{\mathcal{I}}:\sum_{i\in\mathcal{I}} a_i\nu_i\geq b} \sum_{i\in\mathcal{I}} w_i K_i(\mu_i|\nu_i).$$

This shows that

$$\inf_{\nu\in\mathcal{A}_2} \sum_{j=1}^{K} w_j^* K_j(\mu_j|\nu_j) = \inf_{\nu_{\mathcal{I}}\in(\mathcal{A}_2)_{\mathcal{I}}} \sum_{i\in\mathcal{I}} w_i^* K_i(\mu_i|\nu_i) = \sum_{i\in\mathcal{I}} w_i^* K_i(\mu_i|\nu_i^*) = \max_i K_i(\mu_i|\nu_i^*).$$

Now, consider any $\tilde{w}$ which is a feasible solution of Problem CVX. Then,

$$\inf_{\nu\in\mathcal{A}_2} \sum_{i=1}^{K} \tilde{w}_i K_i(\mu_i|\nu_i) \leq \sum_{i=1}^{K} \tilde{w}_i K_i(\mu_i|\nu_i^*) \leq \max_i K_i(\mu_i|\nu_i^*).$$

This proves our claim that w^*, $\nu(w^*) = \nu^*$ form an optimal solution. ■

Appendix C. Lower bound analysis when $\mathcal{A}_1$ is a polytope

In this section we first outline a simple algorithm to solve the sub problem

$$g_j(\mu, w) = \inf_{\nu \in \mathcal{B}_j} \sum_{i=1}^{K} w_i K_i(\mu_i|\nu_i)$$

as discussed in Section 3.4. We then provide a proof of Lemma 10. In Appendix C.0.1, we consider two arms, both with Gaussian distribution and known and common variance. In this simple setting, we conduct a comprehensive analysis of the lower bound problem and provide a graphical interpretation of the solutions.

Solving $g_j(\mu, w)$: Observe that solving for $g_j(\mu, w)$ is equivalent to solving

$$\inf_{\nu : \sum_{i=1}^{K} a_i \nu_i \geq b} \sum_{i=1}^{K} w_i f_i(\nu_i), \tag{29}$$

for a given $w \in \mathcal{P}_K$, where each f_i is strictly convex, and $f_i(\mu_i) = f_i'(\mu_i) = 0$, and without loss of generality $b > 0$. Again, without loss of generality, we assume that $w_i > 0$ for each i. The existence of a unique solution is best seen from the graphical interpretation in Remark 3. We now discuss how this may be efficiently computed.

Observe that f_i' is a strictly increasing function. Let h_i denote the inverse function of f_i'. h_i is also strictly increasing.

The first order conditions applied to (29) imply that the optimal solution ν^* satisfies

$$\nu_i^* = h_i\left(\frac{\lambda a_i}{w_i}\right)$$

for a non-negative λ such that

$$\sum_{i=1}^{K} a_i h_i\left(\frac{\lambda a_i}{w_i}\right) = b. \tag{30}$$

Observe that $\sum_{i=1}^{K} a_i h_i(\frac{\lambda a_i}{w_i})$ equals 0 for $\lambda = 0$, and it strictly increases with increase in λ. Thus one may use any line search method to find λ that solves (30).

Proof of Lemma 10: Denote the optimal value of (16) by C^*. We first show that if $q, s \in \mathcal{P}_K$ are two distinct optimal solutions and $\nu(q), \nu(s) \in \mathcal{A}^c$, respectively achieve the minimum in the sub-problem, then $\nu(q) \neq \nu(s)$. To see this, suppose $\nu(q) = \nu(s) = \nu \in \partial \mathcal{B}_j$ for some $1 \leq j \leq m$. Then ν achieves the minimum in the subproblem $\inf_{\nu \in \mathcal{B}_j} \sum_{i=1}^{K} w_i K_i(\mu_i|\nu_i)$ for both $w = q$ and $w = s$. Hence, both q, s solve the following equations:

$$\sum_{i=1}^{K} w_i K_i(\mu_i|\nu_i) = C^*, \tag{31}$$

$$\frac{w_i}{a_{j,i}} K_i'(\mu_i|\nu_i) = \frac{w_1}{a_{j,1}} K_1'(\mu_1|\nu_1) \quad \forall i. \tag{32}$$

This is a contradiction as the above set of equations has a unique solution.

Now, suppose $q, s \in \mathcal{P}_K$ are two distinct optimal solutions of the convex program (16). Then any convex combination $z = \alpha q + (1-\alpha)s$ is also an optimal solution. Let $\nu(z)$ achieve the minimum in the sub-problem for z. Then

$$C^* = \sum_{i=1}^{K} z_i K_i(\mu_i|\nu_i(z)) = \alpha \sum_{i=1}^{K} q_i K_i(\mu_i|\nu_i(z)) + (1-\alpha) \sum_{i=1}^{K} s_i K_i(\mu_i|\nu_i(z)).$$

In addition, for any ν, we have $\sum_{i=1}^{K} w_i K_i(\mu_i|\nu_i) \leq C^*$ for both $w = q$ and $w = s$. Then, the above equality is possible only if $\sum_{i=1}^{K} q_i K_i(\mu_i|\nu_i(z)) = \sum_{i=1}^{K} s_i K_i(\mu_i|\nu_i(z)) = C^*$. This in turn implies that $\nu(z)$ achieves the minimum in the sub-problem for both q, s, which is a contradiction to our earlier result. Hence proved. □

C.0.1. Two arms Gaussian setting

To illustrate the issues that arise with $\mathcal{A}_2$ being a union of half-spaces, consider a simple setting of two arms. Both are assumed to have a Gaussian distribution and the variance of each arm is assumed to be 1/2. W.l.o.g. mean of each arm is set to zero. Then, for $j = 1, 2$,

$$\mathcal{B}_j = \{\nu \in \mathbb{R}^2 : a_{j,1}\nu_1 + a_{j,1}\nu_2 \geq b_j\}, \tag{33}$$

and $\mathcal{A}_2 = \mathcal{B}_1 \cup \mathcal{B}_2$ be the union of the two half-spaces. To avoid degeneracies we assume that each $a_{j,k} \neq 0$. Further suppose that $\frac{a_{1,1}}{a_{1,2}} \neq \frac{a_{2,1}}{a_{2,2}}$ so that $\mathcal{A}_2$ is non-convex.

The lower bound problem is then given by

$$\max_{(w_1,w_2)\in\mathcal{P}_2} \inf_{\nu\in\mathcal{A}_2} \sum_{i=1}^{2} w_i \nu_i^2. \tag{34}$$

The following geometrical result provides useful insights towards solution of (34).

Proposition 17 *For $w_1, w_2, C > 0$, a necessary and sufficient condition for an ellipse of the form*

$$w_1\nu_1^2 + w_2\nu_2^2 = C \tag{35}$$

to be uniquely tangential to lines

$$a_{1,1}\nu_1 + a_{1,2}\nu_2 = b_1 \tag{36}$$

and

$$a_{2,1}\nu_1 + a_{2,2}\nu_2 = b_2 \tag{37}$$

is that

$$\min_{k=1,2} \left|\frac{a_{2,k}}{a_{1,k}}\right| < \frac{b_2}{b_1} < \max_{k=1,2} \left|\frac{a_{2,k}}{a_{1,k}}\right|. \tag{38}$$

Then, the tangential ellipse is specified by

$$\frac{w_1}{C} = \frac{(a_{1,2}a_{2,1})^2 - (a_{1,1}a_{2,2})^2}{(b_2 a_{1,2})^2 - (b_1 a_{2,2})^2} \tag{39}$$

and

$$\frac{w_2}{C} = \frac{(a_{1,2}a_{2,1})^2 - (a_{1,1}a_{2,2})^2}{(b_1 a_{2,1})^2 - (b_2 a_{1,1})^2}. \tag{40}$$

The ellipse (35) meets the line (36) at point

$$\left(\frac{Ca_{1,1}}{w_1 b_1}, \frac{Ca_{1,2}}{w_2 b_1}\right)$$

and it meets line (37) at point

$$\left(\frac{Ca_{2,1}}{w_1 b_2}, \frac{Ca_{2,2}}{w_2 b_2}\right).$$

Proof A necessary and sufficient condition for ellipse (35) to be tangential to line (36) at point (v_1^*, v_2^*) is for (v_1^*, v_2^*) to satisfy the two equations of ellipse and the line, respectively, and the slope matching condition

$$\frac{w_1 v_1^*}{a_{1,1}} = \frac{w_2 v_2^*}{a_{1,2}}. \tag{41}$$

The fact that (v_1^*, v_2^*) satisfies (35) and (36) implies that (41) equals C/b_2. Plugging (v_1^*, v_2^*) from (41) into (35), we observe,

$$\frac{a_{1,1}^2}{w_1} + \frac{a_{1,2}^2}{w_2} = \frac{b_1^2}{C}.$$

Similarly, considering the other half-space, we get

$$\frac{a_{2,1}^2}{w_1} + \frac{a_{2,2}^2}{w_2} = \frac{b_2^2}{C}.$$

The result follows by solving the two equations. ■

Theorem 18 *The solution to (34) depends in the following way on the underlying parameters*
Case 1:

$$\left(\frac{b_2}{b_1}\right)^2 \geq \left(\frac{a_{2,1}^2}{|a_{1,1}|} + \frac{a_{2,2}^2}{|a_{1,2}|}\right)(|a_{1,1}| + |a_{1,2}|)^{-1}. \tag{42}$$

In this case, (34) reduces to the half-space problem where $\mathcal{A}_2 = \mathcal{B}_1$ so that the optimal solution to (34) is given by

$$w_i^* = \frac{|a_{1,i}|}{|a_{1,1}| + |a_{1,2}|}, \quad i = 1, 2, \tag{43}$$

and the optimal value $C^* = \frac{b_1^2}{(|a_{1,1}|+|a_{1,2}|)^2}$.
Case 2:

$$\left(\frac{b_2}{b_1}\right)^2 \leq \left(\frac{a_{1,1}^2}{|a_{2,1}|} + \frac{a_{1,2}^2}{|a_{2,2}|}\right)^{-1}(|a_{2,1}| + |a_{2,2}|).$$

This simply corresponds to Case 1, with the $(a_{1,1}, a_{1,2}, b_1)$ interchanged with $(a_{2,1}, a_{2,2}, b_2)$.
Case 3:

$$\left(\frac{a_{1,1}^2}{|a_{2,1}|} + \frac{a_{1,2}^2}{|a_{2,2}|}\right)^{-1}(|a_{2,1}| + |a_{2,2}|) < \left(\frac{b_2}{b_1}\right)^2 < \left(\frac{a_{2,1}^2}{|a_{1,1}|} + \frac{a_{2,2}^2}{|a_{1,2}|}\right)(|a_{1,1}| + |a_{1,2}|)^{-1}. \tag{44}$$

Here (38) holds, and the optimal w_1^* and w_2^* are given by (39) and (40), respectively.

Proof Case 1:

First consider the half-space problem where $\mathcal{A}_2 = \mathcal{B}_1$. Our analysis in Section 3.2 shows that there is a unique (w_1^*, w_2^*) and (ν_1^*, ν_2^*) that solves the resulting problem, and

$$\text{sign}(a_{1,1})\nu_1^* = |\nu_1^*| = \text{sign}(a_{1,2})\nu_2^* = |\nu_2^*|,$$

$a_{1,1}\nu_1^* + a_{1,2}\nu_2^* = b_1$ so that

$$|\nu_1^*| = |\nu_2^*| = \frac{b_1}{|a_{1,1}| + |a_{1,2}|}.$$

Further, from

$$\frac{w_1^*\nu_1^*}{a_{1,1}} = \frac{w_2^*\nu_2^*}{a_{1,2}}, \tag{45}$$

it follows that for the half-space problem, $w_i^* \propto |a_{1,i}|$ is the optimal solution and the optimal value $C^* = \frac{b_1^2}{(|a_{1,1}|+|a_{1,2}|)^2}$.

Returning to (34), we show that when (42) is true and and $w_i^* \propto |a_{1,i}|$,

$$\inf_{\nu\in\mathcal{B}_2} \sum_{i=1}^{2} w_i^* K_i(\mu_i|\nu_i) = \inf_{\nu: a_{2,1}\nu_1 + a_{2,2}\nu_2 \geq b_2} w_1^*\nu_1^2 + w_2^*\nu_2^2 \geq C^*$$

and hence $w_i^* \propto |a_{1,i}|$ continues to be optimal for (34).

We first find the point $(\kappa_1^*, \kappa_2^*) \in \mathcal{B}_2$ that achieves the minimum in the above optimization problem. We know that (κ_1^*, κ_2^*) satisfies

$$a_{2,1}\kappa_1^* + a_{2,2}\kappa_2^* = b_2,$$

and the slope matching condition

$$\frac{w_1^*\kappa_1^*}{a_{2,1}} = \frac{w_2^*\kappa_2^*}{a_{2,2}}.$$

It follows from easy calculations that

$$\inf_{\nu: a_{2,1}\nu_1 + a_{2,2}\nu_2 \geq b_2} w_1^*\nu_1^2 + w_2^*\nu_2^2 = \frac{b_2^2}{\frac{a_{2,1}^2}{w_1^*} + \frac{a_{2,2}^2}{w_2^*}} = \frac{b_2^2}{\left(\frac{a_{2,1}^2}{|a_{1,1}|} + \frac{a_{2,2}^2}{|a_{1,2}|}\right)(|a_{1,1}| + |a_{1,2}|)}.$$

The above expression is greater than $\frac{b_1^2}{(|a_{1,1}|+|a_{1,2}|)^2}$ when (42) is true, which gives us the required result.

Case 2: Case 2 follows similarly as Case 1.

Case 3:

It is easy to see that (44) implies (38).

Let (w_1^*, w_2^*) denote the optimal solution to (34). It is clear that the corresponding ellipse must be tangential to both the half lines $a_{1,1}\nu_1 + a_{1,2}\nu_2 = b_1$ and $a_{2,1}\nu_1 + a_{2,2}\nu_2 = b_2$, since if it does not touch one of these half lines, then the associated constraint can be ignored in solving (34). However, that violates (44).

Therefore, the solution is provided by Proposition 17. ■

Appendix D. Analysis related to the proposed algorithm

In this section, we first prove Lemma 13. Then, in Lemma 19 we summarize results from Garivier and Kaufmann (2016) on the D-tracking rule that we use in our proof of Theorem 12. This proof is more or less identical to that in Garivier and Kaufmann (2016). We keep it here for completeness.

Proof of Lemma 13: First suppose that $\bar{\mathcal{A}}^c$ is compact. The fact that g is continuous at (μ, w) follows from the continuity results for non-linear programs. Specifically, since the objective function is continuous in ν and $\bar{\mathcal{A}}^c$ is compact, Theorem 2.1 in Fiacco and Ishizuka (1990) implies that g is continuous at (μ, w).

Now consider non-compact $\bar{\mathcal{A}}^c$ and define

$$g_n(\mu, w) = \inf_{\nu \in \bar{\mathcal{A}}^c \cap \mathcal{B}_n} \sum_{i=1}^{K} w_i K_i(\mu_i | \nu_i)$$

for each n where $\mathcal{B}_n$ is an Euclidean closed ball of radius n centred at μ. n is taken to be sufficiently large so that $\bar{\mathcal{A}}^c \cap \mathcal{B}_n$ is non-empty.

Then, $g_n(\mu, w)$ is continuous in (μ, w) and decreases with n to $g(\mu, w)$. Since this convergence is uniform, it follows that $g(\mu, w)$ is continuous in (μ, w).

To see that the optimal solution to Problem LB is continuous at μ if $\mathcal{W}(\mu)$ is a singleton, note that the problem is equivalent to $\max_{w \in \mathcal{P}_K} g(\mu, w)$. Since $g(\mu, \cdot)$ is continuous on $\mathcal{P}_K$ and $\mathcal{W}(\mu)$ is a singleton, from Theorem 2.2 in Fiacco and Ishizuka (1990), we conclude that the optimal solution $\mathcal{W}(\mu)$ is continuous at μ. □

Lemma 19 *The D-tracking rule ensures that* $\min_i N_i(t) \geq \left(\sqrt{t} - K/2\right)^+ - 1$ *and that for all* $\epsilon > 0$*, for all* t_0*, there exists* $t_\epsilon \geq t_0$ *such that if* $\sup_{t \geq t_0} \max_i |\hat{w}_i(t) - w_i| \leq \epsilon$ *for some* $w \in \mathcal{P}_K$*, then*

$$\sup_{t \geq t_\epsilon} \max_i \left| \frac{N_i(t)}{t} - w_i \right| \leq 3(K-1)\epsilon.$$

Proof of Theorem 12: Recall that $\mu \in \mathcal{A}$. We first prove that the probability of error is at most δ.

$$\begin{aligned}
\mathbb{P}_\mu[\text{error}] &\leq \mathbb{P}_\mu\left[\exists t \geq 1 : \inf_{\nu \in \mathcal{A}} \sum_i N_i(t) K_i(\hat{\mu}_i(t) | \nu_i) \geq \beta(t, \delta); \hat{\mu}_t \in \mathcal{A}^c\right] \\
&\leq \sum_{t=1}^{\infty} \mathbb{P}_\mu\left[\sum_i N_i(t) K_i(\hat{\mu}_i(t) | \mu_i) \geq \beta(t, \delta)\right] \\
&\leq \sum_{t=1}^{\infty} e^{K+1} \left(\frac{\beta(t, \delta)^2 \log t}{K}\right)^K e^{-\beta(t, \delta)}
\end{aligned}$$

The last inequality above follows from Magureanu et al. (2014) extended from Bernoulli family to SPEF.

Recall that $\beta(t,\delta) = \log\left(\frac{c\,t^2 \log(1/\delta)^{2K+1}}{\delta}\right)$. If c is chosen large enough s.t.

$$\sum_{t=1}^{\infty} \frac{e^{K+1}}{c\,t^2 \log(1/\delta)^{2K+1}} \left(\frac{[\log(c\,t^2) + \log((2K+1)\log(1/\delta)) + \log(1/\delta)]^2 \log t}{K}\right)^K \leq 1, \quad (46)$$

then

$$\mathbb{P}_\mu[\text{error}] \leq \delta.$$

(Above in (46), we need to restrict δ to less than 1 to avoid LHS blowing up to infinity.)

Next, we prove the upper bound on the mean termination time. Let $\mathcal{B}_\infty^y(x)$ denote a Euclidean ball around x of length y under the max norm.

Fix an $\epsilon > 0$. From the continuity of w at μ, there exists $\xi > 0$ such that for any $\mu' \in \mathcal{B}_\infty^\xi(\mu)$ we have $w(\mu') \in \mathcal{B}_\infty^\epsilon(w(\mu))$. For any $T \in \mathbb{N}$, define the event $\mathcal{E}_T := \bigcap_{t=h(T)}^T \{\hat{\mu}(t) \in \mathcal{B}_\infty^\xi(\mu)\}$. It is easy to show that (see Lemma 19 of Garivier and Kaufmann (2016)) there exist constants B, C depending on ϵ and μ such that

$$\mathbb{P}_\mu\left[\mathcal{E}_T^c\right] \leq B \exp\left(-CT^{1/8}\right).$$

Note that ξ, $\mathcal{E}_T$, B, C are all functions of ϵ and μ.

Now, for every $\epsilon > 0$, define

$$C_\epsilon^*(\mu) = \inf_{\substack{\mu' \in \mathcal{B}_\infty^{\xi(\epsilon)}(\mu), \\ w' \in \mathcal{B}_\infty^{3(K-1)\epsilon}(w(\mu))}} g(\mu', w').$$

By the continuity of w and g, we have

$$\lim_{\epsilon \to 0} C_\epsilon^*(\mu) = C^*(\mu) = (T^*(\mu))^{-1}.$$

From Lemma 19, for any $\epsilon > 0$, we have for every $T \geq T_\epsilon$ that on $\mathcal{E}_T(\epsilon)$,

$$\left\|\frac{N(t)}{t} - w(\mu)\right\|_\infty \leq 3(K-1)\epsilon \quad \forall t > \sqrt{T},$$

which in turn implies that

$$g\left(\hat{\mu}(t), \frac{N(t)}{t}\right) \geq C_\epsilon^*(\mu) \quad \forall t > \sqrt{T}.$$

Since the termination rule in the algorithm is given by

$$g\left(\hat{\mu}(t), \frac{N(t)}{t}\right) \geq \frac{\beta(t,\delta)}{t},$$

for $T \geq T_\epsilon$, on $\mathcal{E}_T(\epsilon)$, we have

$$\begin{aligned} \min(T_U(\delta), T) &\leq \sqrt{T} + \textstyle\sum_{t=\sqrt{T}}^T \left\{C_\epsilon^*(\mu) < \frac{\beta(t,\delta)}{t}\right\} \\ &\leq \sqrt{T} + \frac{\beta(T,\delta)}{C_\epsilon^*(\mu)}. \end{aligned}$$

Now, let

$$T_0(\delta) := \inf\left\{T \in \mathbb{N} : \sqrt{T} + \frac{\beta(T,\delta)}{C^*_\epsilon(\mu)} < T\right\}.$$

Therefore, for any $T \geq \max\{T_\epsilon, T_0(\delta)\}$, on $\mathcal{E}_T(\epsilon)$, we have $T_U(\delta) < T$, which gives us

$$\mathbb{E}[T_U(\delta)] \leq \sum_{T=1}^{\infty} \mathbb{P}[T_U(\delta) > T] \leq \max\{T_\epsilon, T_0(\delta)\} + \sum_{T=1}^{\infty} \mathbb{P}[\mathcal{E}_T(\epsilon)^c].$$

It easily follows that

$$T_0(\delta) = \frac{1}{C^*_\epsilon(\mu)} \left(O\left(\log\left(1/\delta\right)\right) + O\left(\log\log\left(1/\delta\right)\right)\right).$$

This gives us

$$\limsup_{\delta\to 0} \frac{\mathbb{E}[T_U(\delta)]}{\log\left(\frac{1}{\delta}\right)} \leq \frac{1}{C^*_\epsilon(\mu)}.$$

Now, letting ϵ go to zero, we get

$$\limsup_{\delta\to 0} \frac{\mathbb{E}[T_U(\delta)]}{\log\left(\frac{1}{\delta}\right)} \leq \lim_{\epsilon\to 0} \frac{1}{C^*_\epsilon(\mu)} = T^*(\mu).$$

□

Appendix E. Simulation experiment

Our proof for the proposed algorithm is asymptotic as $\delta \to 0$. To test the efficacy of the algorithm for practically relevant values of δ and to compare its performance to the case where δ is small, we conduct a small simulation experiment. We consider the half-space problem in the four Bernoulli arms setting. The associated hyper-plane is given by

$$x_1 + x_2 + x_3 + x_4 = 2.6,$$

and mean vector $\mu = (0.5, 0.5, 0.5, 0.5)$. We consider δ ranging from 0.05 to 0.001. Figure 3 plots the average computational effort of the proposed algorithm (2,000 independent trials were conducted for each δ), and compares it to the theoretical lower bound for this half-space problem. The key conclusion is that while the relative performance improves somewhat as $\delta \to 0$, it is not that much worse than the limit even for practically reasonable values of δ such as 0.05 or 0.01.

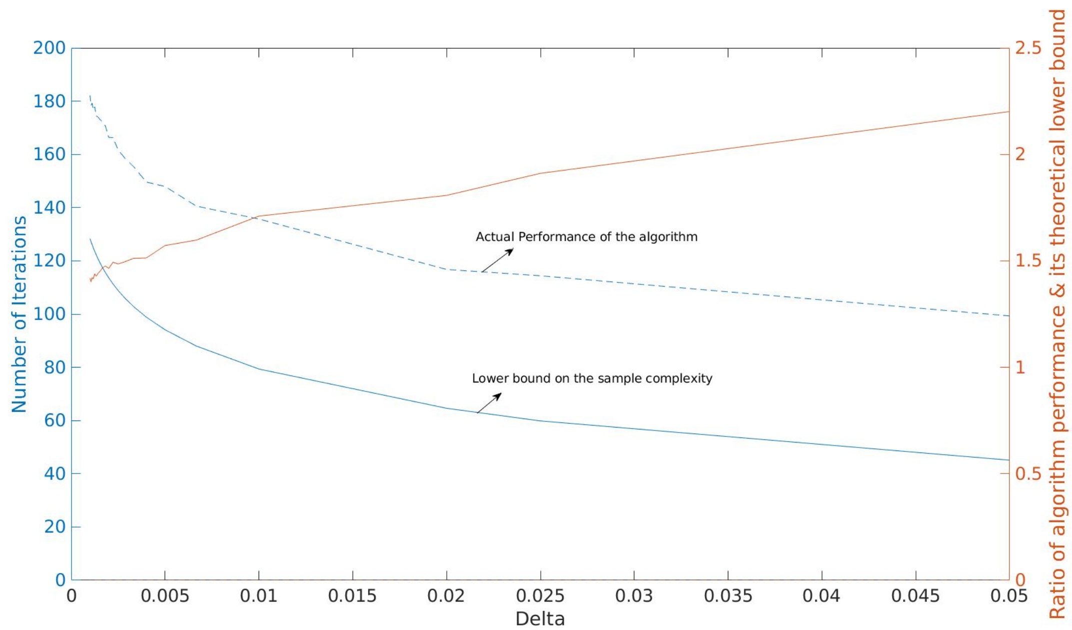

Figure 3: The four arm half-space problem.The relative performance of the algorithm compared to the lower bound does not deteriorate much with increasing δ.