Submitted to *INFORMS Journal on Computing*
manuscript (Please, provide the manuscript number!)

Practical Nonparametric Sampling Strategies for Quantile-based Ordinal Optimization

Dongwook Shin
School of Business and Management, Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon, Hong Kong
{dwshin@ust.hk}

Mark Broadie, Assaf Zeevi
Graduate School of Business, Columbia University, New York, NY 10025
{mnb2@columbia.edu, assaf@gsb.columbia.edu}

Given a finite number of stochastic systems, the goal of our problem is to dynamically allocate a finite sampling budget to maximize the probability of selecting the "best" system. Systems are encoded with the probability distributions that govern sample observations, which are unkown and only assumed to belong to a broad family of distributions that need not admit any parametric representation. The "best" system is defined as the one with the highest quantile value. The objective of maximizing the probability of selecting this "best" system is not analytically tractable. In lieu of that we use the rate function for the probability of error relying on large deviations theory. Our point of departure is an algorithm that naively combines sequential estimation and myopic optimization. This algorithm is shown to be aymptotitcally optimal, however, it exhibits poor finite-time performance and does not lead itself to implemention in settings with a large number of systems. To address this we propose practically implementable variants that retain the asymptotic performance of the former, while dramatically improving its finite-time performance.

Key words: Quantile, ordinal optimization, tractable procedures, large deviations theory
History:

1. Introduction

Given a finite number of alternative distributions, henceforth referred to as *systems*, we are concerned with the problem of selecting the "best" one, where performance of the systems is initially unknown but it is possible to sequentially sample from each of them. A critical assumption made in most academic studies is that a decision maker is primarily interested in the mean performances of the systems, i.e., the means of underlying distributions. However, such mean-based approaches do not accommodate various risk preferences of the decision maker. In this paper, we explore cases where downside or upside risk is more crucial than the average performance. In particular, the systems are compared based on pth quantiles of their distributions, with the value of p chosen according to the decision maker's risk preference.

In most cases, the underlying probability distributions rarely admit a parametric representation. Therefore, we focus on nonparametric settings where the probability distributions belong to a broad family that cannot be characterized by one or more parameters. While nonparametric approaches have an extensive range of applications, the main difficulty is loss of tractability. Hence, we are motivated to develop nonparametric sampling strategies that are implementable and perform well in practice.

The problem described above arises in a variety of applications. A widely used example is value-at-risk (VaR) that measures the pth quantile of a portfolio's value, with p typically chosen to be 1% or 5%. For large portfolios of complex derivative securities,[1] value-at-risk can be evaluated through Monte Carlo methods, which typically entails substantial computational burden. The quantile-based approach is quite relevant in the case where a portfolio manager needs to identify the portfolio with largest value-at-risk out of many alternatives; in particular, a portfolio manager needs to sequence the simulation trials in an efficient manner given a certain amount of time to increase the likelihood of identifying the portfolio with largest VaR.

Another example is selecting the best design of a telephone call center, where a typical measure of interest is the quantile of waiting time. For instance, if a conservative decision maker uses the quantiles for $p = 0.95$ to compare different designs, then the best design will have the smallest value of the 95% quantile of waiting times. The waiting time is a complex function of design parameters, such as staffing and routing policies, and is often evaluated through simulation. The decision maker's job is to configure a finite number of designs and to spend a certain amount of time testing different designs before deciding which is best.

The main objective of this paper is to design a dynamic sampling algorithm that minimizes the probability of selecting suboptimal systems subject to a given sampling budget.

[1] In practice, there can be some correlation in performance between portfolios if they consist of similar assets. However, following most of papers in the area of ordinal optimization and ranking and selection, our model assumes independent systems. (Our model serves as a good approximation when the degree of correlation is not significant.)

Since the aforementioned objective is not analytically tractable, our departure point will be an asymptotic benchmark characterized by a large sampling budget. In this regime, the rate function (Dembo and Zeitouni 2009) of the probability of false selection, hereafter simply referred to as the rate function, can be written in tractable form.

Exact evaluation of the rate function, however, is possible only when the decision maker knows the underlying probability distributions in advance, and the absence of perfect prior information introduces an important new component into the optimization problem. Namely, one needs to allocate samples to maximize the rate function (exploitation) but at the same time take sufficient samples from each system to learn the underlying distributions (exploration).

As a benchmark for further analysis, we propose a naive algorithm that estimates unknown values of the rate function from the history of sample observations, and then allocates a sample in each subsequent stage as if the unknown values are equal to the sample estimates; this is often referred to as a certainty-equivalence approach. As will be explained later, this algorithm maximizes the objective asymptotically, however, its finite-time performance is poor since it spends too much of the sampling budget on exploration, which leaves less budget to exploit that knowledge and optimize the objective. Another drawback of this approach is the heavy computational burden since it requires solving a difficult nested optimization problem repeatedly, which makes the algorithm not practically implementable in many applications with a large number of systems.

The main contribution of this paper is to introduce sampling strategies that are practical to implement in applications with large number of systems, and improve their finite-time performance relative to the aforementioned class of algorithms. In more detail, our contributions are summarized as follows.

(i) We introduce an alternative performance metric and show that it is closely aligned with the rate function when the gap between the quantiles of the best and second-best systems is sufficiently small. The alternative performance metric is structured around the quantiles and local density behavior, rather than global distribution functions, which lends itself to some key structural insights about near-optimal allocations.

(ii) Building on the structural properties of the alternative performance metric, we propose a family of nonparametric dynamic sampling algorithms, hereafter referred to as Quantile Divergence (QD), with the aim of improving the finite-time performance of the aforementioned naive algorithm. We rigorously show that this class of algorithms is near-optimal in a precise mathematical sense.

(iii) We also propose variants of QD, hereafter referred to as Adaptive QD (AQD), that are computationally efficient when the number of systems is large. We show via numerical testing that these variants preserve the asymptotic performance of the QD algorithms and improve the computation time dramatically.

Our proposed algorithms are *nonparametric* in the sense that they do not postulate any parametric structure on underlying probability distributions. Despite having almost no prior information, these algorithms allocate samples in a way that is close to the best ex ante allocation under full information. When the underlying distributions are known to belong to a parametric family, more efficient algorithms can be designed, albeit at the cost of misspeciation; in particular, if the assumed parametric structure is not "close enough" to that of the true underlying family of distributions, such parametric algorithms may fail to identify the best system.

The remainder of this paper is organized as follows: In §2 we survey related literature on ordinal optimization, and ranking and selection problems, and relevant studies from the multi-armed bandit literature. In §3 we derive a tractable objective function using large deviations theory and formulate a dynamic optimization problem. In §4 we suggest dynamic algorithms and provide theoretical analyses on their performances. In §5 we propose adaptive variants of the algorithms that are practically implementable in large problem instances. In §6 we test the suggested policies numerically and compare with several benchmark policies. This paper has an online supplement with three parts. Appendix A states additional theoretical results; Appendix B contains the proofs of main theoretical results and auxiliary lemmas; and Appendix C contains additional numerical results to provide practical guideline on the implementation of proposed algorithms.[2]

2. Literature Review

Mean-based R&S Procedures. An area closely related to ordinal optimization is that of ranking and selection (R&S). While the goal of this paper is to minimize the probability of

[2] A very preliminary version of this paper appeared as Shin et al. (2016). In particular, §4 contains much stronger theoretical results in §3 of the prior paper, and Appendix A provides further theoretical results. New algorithms to improve those in the prior paper are given in §5. Lastly, §6 exhibits numerical experiments in more extensive settings than those in §4 of the previous paper.

false selection given a *fixed* sampling budget, the goal of the R&S problem is to take as few samples as possible to satisfy a desired guarantee on the probability of correctly selecting the best system; most of this literature considers "best" to be the largest mean. The traditional R&S approach traces back to the work of Bechhofer (1954) who established the Indifference Zone (IZ) formulation. (See the survey paper by Kim and Nelson (2006).) Most IZ procedures rely on the assumption that sample observations are normally distributed, and hence raise the risk of misspecification alluded to earlier.

Quantile-based R&S Procedures. Despite the wide usage of quantiles as a performance metric, the topic of quantile-based R&S procedure has not received much attention in the literature. Bekki et al. (2007) modify the traditional two-stage IZ procedure by Rinott (1978) to suggest a heuristic technique that addresses the issue of non-normality of quantile estimates, by averaging them over batches. (Thus the batch behavior would be nearly normal by the central limit theorem.) Batur and Choobineh (2010) suggest a two-stage procedure based on Rinott (1978), where a set of quantile values is compared between two systems. Lastly, Lee and Nelson (2014) suggest an R&S procedure based on bootstrapping, which can be applied to general performance measures including quantile, albeit with a heavy computational load. In contrast to these papers, we show that the (asymptotically) optimal allocation is unaffected by the use of batches, so that the allocation based on the normal approximation may give substantially suboptimal allocations.

Mean-based Ordinal Optimization Procedures. Several studies consider the probability of selecting the best system when a finite sampling budget is given. Notable recent examples include the Optimal Computing Budget Allocation (OCBA) rule by Chen et al. (2000), the Knowledge Gradient (KG) Bayesian framework of Frazier et al. (2008). These formulations provide an attractive, stylized analysis but they rely heavily on normality assumptions. In contrast, the "frequentist" approach of Glynn and Juneja (2004) based on large deviations theory allows for much broader scope but is difficult to implement due to (i) the challenges in estimating the moment generating function, as noted by Glynn and Juneja (2015), and (ii) the computational burden in optimizing the large deviations rate function, as noted by Pasupathy et al. (2015). Regarding (i), Glynn and Juneja show that this issue can be mitigated in restricted settings where upper bounds on suitable moments of underlying distributions are known (such information is rarely available in most practical applications). Regarding (ii), Pasupathy et al. show that the optimal allocation for the rate function becomes tractable in a certain asymptotic limit characterized by the large number of systems. In the context of our work, the large deviations rate function also entails issues (i) and (ii); see § 4.1. However, our work is significantly different from previous papers in that we circumvent these issues by introducing an approximation to the rate function in a certain asymptotic regime characterized by the quantiles of the systems.

The recent work of Shin et al. (2018) also addresses the issues with the large deviations rate function approach of Glynn and Juneja (2004) by proposing a two moment approximation and leveraging it to design a well-performing dynamic procedure called Welch Divergence (WD). This paper shares an important common theme with Shin et al. (2018) in that we too use the large deviations rate function to construct competitive sampling procedures. On the other hand, the key conclusions from the two papers are fundamentally different: the two-moment approximation of Shin et al. (2018) suggests that when systems are compared based on means, the probability of false selection can be approximated by the rate function corresponding to Gaussian distribution if the probability distributions of the best and second-best systems are "close" to each other. In contrast, when systems are compared based on quantiles, we show that the allocation based on the two-moment approximation can be significantly sub-optimal, as the (asymptotically) near-optimal allocation rule requires density information rather than just two moments. This calls into further questions the Gaussian assumptions that are prevalent in the literature.

Quantile-based Ordinal Optimization Procedures. Pasupathy et al. (2010) characterize the rate function associated with the probability of false selection using large deviations theory. However to build on this one requires knowledge of the underlying distribution functions. This paper extends their work along three dimensions: we strengthen their theoretical results on the rate function characterization by relaxing conditions on the underlying distributions; we derive an approximation to the rate function using only quantiles and local density estimates at particular points, which is simpler than having to estimate an entire distribution function; and we propose sampling algorithms that judiciously manage the issue of the lack of prior information on the underlying distributions.

Peng et al. (2019) propose Bayesian sampling algorithms for selecting the optimal quantile based on the premise that underlying probability distributions are normal. They rigorously show the (in)consistency properties of two algorithms and develop a switching strategy between the two in order to provide balanced performance in small- and large-sample scenarios. They show via numerical testing the performance of the switching strategy in cases with non-normal probability distributions, though the theoretical guarantee hinges significantly on the normality assumption. Our focus is to design and analyze algorithms that do not rely on any parametric distributional assumption, providing rigorous justification for the efficacy of the proposed algorithms in the non-normal case.

Quantile-based Best-arm Identification. Our research is closely related to that of pure exploration in the multi-armed bandit (MAB) problem, often referred to as best-arm identification; see Bubeck and Cesa-Bianchi (2012) for a comprehensive overview. The best-arm identification procedures seek the same goal of selecting the best arm (i.e., system, in the language of our paper). In the context of quantile-based pure exploration, Yu and Nikolova (2013) consider the problem where arms are compared based on value-at-risk. Szorenyi et al. (2015) proposes quantile-based online learning algorithms when rewards from the arms are not necessarily continuous real-valued. Our work differs significantly from these papers in that we deal with the setting with a fixed sampling budget. Further, from an analytical perspective, the two antecedent papers primarily use the concentration property of quantile estimators to determine the number of samples taken to satisfy a desired guarantee. In contrast, our main concern is to minimize the probability of false selection given a (sufficiently large) number of samples, which gives rise to the question of what is the behavior of the probability distributions of quantile estimators in the extreme tails. This question is addressed in § 3.2.

3. Formulation

3.1. Model Primitives

Consider k stochastic systems, each of which is characterized by a distribution function $F_j(\cdot)$, $j=1,\dots,k$, with its support denoted as $\mathscr{H}_j$. We define the system configuration as $\mathbf{F}=\{F_1,\dots,F_k\}$; we denote $\mathbf{F}\in\mathscr{C}$ if each distribution is continuous on its domain; and $\mathbf{F}\in\mathscr{D}$ if each distribution is discrete and, without loss of generality, has a support in the set of nonnegative integers. Fix $p\in(0,1)$ that represents the quantile of interest and define the pth quantile of $F_j(\cdot)$ as

$$\xi_j=\inf\{x: F_j(x)\geq p\}. \tag{1}$$

Denote $\boldsymbol{\xi}=(\xi_1,\dots,\xi_k)$ as the k-dimensional vector of the pth quantiles. Without loss of generality, we take $\xi_1>\xi_2\geq\cdots\geq\xi_k$. We make the following assumptions throughout the paper.

(F1) ξ_j is the unique solution x of $F_j(x-) \le p \le F_j(x)$ for each system j

(F2) $[\xi_k, \xi_1] \subset \cap_{j=1}^k \mathscr{H}_j$

Note that (F1) is a mild assumption that ensures $F_j(\cdot)$ is not flat around the pth quantile. Assumption (F2) ensures that each quantile estimated from sample observations can take any value in the interval $[\xi_k, \xi_1]$, in order to avoid trivial cases where the probability of false selection is zero. In our model the decision maker does not know the system configuration $\mathbf{F}$, but is able to sequentially take T independent samples from the systems to infer quantiles of the distributions, where T is exogenously given. The goal is to correctly identify the system with the largest pth quantile, i.e., system 1.

Let π denote an algorithm, which is a sequence of random variables, $\pi_1, \pi_2, \ldots$, taking values in the set $\{1, \ldots, k\}$; the event $\{\pi_t = j\}$ means a sample from system j is taken at stage t. Define X_{jt}, $t = 1, \ldots, T$, as a random sample from system j in stage t. The set of non-anticipating policies is denoted as Π, in which the sampling decision in stage t is determined by all the sampling decisions and samples observed in previous stages.

Let N_{jt}^π be the cumulative number of samples up to stage t from system j induced by algorithm π, and define $\alpha_{jt}^\pi := N_{jt}^\pi / t$ as the sampling rate for system j at stage t. The sample distribution function for system j is defined as

$$\hat{F}_{jt}^\pi(x) = \frac{1}{N_{jt}^\pi} \sum_{\substack{\tau=1 \\ \pi_\tau = j}}^{t} \mathbf{I}\{X_{j\tau} \le x\}, \tag{2}$$

where $\mathbf{I}\{A\}$ is one if A is true and zero otherwise. Denote $\hat{\xi}_{jt}^\pi$ as the quantile of the sample distribution function, i.e.,

$$\hat{\xi}_{jt}^\pi = \inf\{x : \hat{F}_{jt}^\pi(x) \ge p\}. \tag{3}$$

It is useful to define a static algorithm denoted by $\pi(\boldsymbol{\alpha}) \in \Pi$, where $\boldsymbol{\alpha} \in \Delta$ with

$$\Delta = \left\{ (\alpha_1, \ldots, \alpha_k) \in \mathcal{R}^k : \sum_{j=1}^k \alpha_j = 1 \text{ and } \alpha_j \ge 0 \text{ for all } j \right\}. \tag{4}$$

The static algorithm is summarized in Algorithm 1. Note that for static policies, $\pi_t(\boldsymbol{\alpha})$ is fixed prior to stage 1, independent of sample observations, and hence, the order of sampling does not affect the performance in the final stage. Also, under $\pi(\boldsymbol{\alpha})$ the number of total samples from system j is $\alpha_j T$, ignoring non-integrality. We let Δ^0 denote the interior of Δ.

ALGORITHM 1: Static($\boldsymbol{\alpha}$)

Set $\alpha_{j0} = 0$ for $j = 1, \dots, k$ and set $t = 1$.

repeat

Take a sample from system π_t, where

$$\pi_t = \arg\max_{j=1,\dots,k} \{\alpha_j - \alpha_{jt}^{\pi}\}, \tag{5}$$

with ties broken arbitrarily. Let $t = t + 1$.

until $t \le T$;

Return: $\arg\max_j \{\hat{\xi}_{jT}\}$.

Definition 1 (Consistency). An algorithm $\pi \in \Pi$ is consistent if $N_{jt}^{\pi} \to \infty$ in probability for each j as $t \to \infty$.[3]

In the optimization problem we consider, we further restrict attention to a set of consistent policies, denoted as $\bar{\Pi} \subset \Pi$. Under such policies the sample quantiles are consistent estimators of the population counterparts, as formalized in the following proposition.

Proposition 1 **(Consistency of quantile estimators)**. *Under (F1),*

$$\hat{\xi}_{jt}^{\pi} \to \xi_j \text{ as } t \to \infty \tag{6}$$

in probability for any consistent algorithm $\pi \in \bar{\Pi}$.

Note that consistent algorithms ensure that the probability of false selection eventually converges to zero as the sampling budget grows to infinity. We remark that the property of consistency is not straightforward to verify for a dynamic algorithm, although any static algorithm $\pi^{\boldsymbol{\alpha}}$ with $\boldsymbol{\alpha} \in \Delta^0$ is consistent since $N_{jt} = \alpha_j t \to \infty$ as $t \to \infty$. The following is a simple example of a dynamic algorithm that is not consistent.[4]

Example 1 (A dynamic algorithm that is not consistent). Suppose $\mathbf{F} \in \mathscr{C}$ and each distribution has infinite support. Define a dynamic algorithm π in which n_0 samples are initially taken from each system and $\pi_t = \arg\max_j\{\hat{\xi}_{jt}\}$ for $t \ge kn_0$. Suppose $k = 2$ and $(\xi_1, \xi_2) = (1, 0)$. At stage $2n_0$, it can be easily seen that the event $A = \{\hat{\xi}_{1,2n_0} \in$

[3] In this paper, we consider the weak consistency but note that some of the results can be strengthened to strong consistency, albeit with a stronger condition; the analysis of this case is available in Appendix A of the online supplement.

[4] See also the MAP algorithm in Peng et al. (2019), which is dynamic but not consistent if $p < 1/2$.

$(-\infty,-1), \hat{\xi}_{2,2n_0} \in (1,\infty)\}$ occurs with positive probability. Conditional on this event, system 1 would not be sampled in subsequent stages if $\hat{\xi}_{2t} \geq 0$ for all $t \geq 2n_0$. The latter event occurs with positive probability because

$$\mathsf{P}\left(\sup_{t\geq 2n_0} |\hat{\xi}_{2t} - \xi_2| \leq 1\right) \geq \frac{2}{1-\gamma}\gamma^{n_0}, \tag{7}$$

where $\gamma = \exp(-2\min(F_2(\xi_2+1)-p, p-F_2(\xi_2-1)))$ (see Section 2.3.2 of Serfling 2009). Combined with the fact that $\mathsf{P}(A) > 0$, system 1 is not sampled infinitely often with positive probability, and hence, the algorithm is not consistent.

Notational conventions. Throughout the paper, we use $\hat{x}_t$ to denote sample estimate of an unknown value x in stage t. For brevity, the superscript π may be dropped when it is clear from the context. We use boldface letters for k-dimensional vectors; e.g., $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_k)$ and $\boldsymbol{\alpha}_t := (\alpha_{1t}, \ldots, \alpha_{kt})$.

3.2. Large Deviations Preliminaries

The probability of false selection, denoted $\mathsf{P}(\mathrm{FS}_t^\pi)$ with $\mathrm{FS}_t^\pi := \{\hat{\xi}_{1t}^\pi < \max_{j\neq 1} \hat{\xi}_{jt}^\pi\}$, is a widely used criterion for the efficiency of a sampling algorithm (see, e.g., the survey paper by Kim and Nelson (2006)). However, the exact evaluation of $\mathsf{P}(\mathrm{FS}_t^\pi)$ under dynamic sampling policies is not analytically tractable. In this subsection, following Glynn and Juneja (2004) we build a tractable objective associated with $\mathsf{P}(\mathrm{FS}_t^\pi)$ based on large deviations theory. In particular, we fix a static algorithm $\pi = \pi(\boldsymbol{\alpha})$ for some $\boldsymbol{\alpha} \in \Delta$ and characterize how fast $\mathsf{P}(\mathrm{FS}_t^{\pi(\boldsymbol{\alpha})})$ converges to 0 as a function of $\boldsymbol{\alpha} \in \Delta$. We eliminate $\pi(\boldsymbol{\alpha})$ in the superscripts in order to improve clarity.

To begin, observe that N_{jt} is deterministic under the static algorithm $\pi(\boldsymbol{\alpha})$ and that

$$\mathsf{P}(\hat{\xi}_{jt} > x) = \mathsf{P}\left(\sum_{s=1}^{N_{jt}} \mathbf{I}\{X_{j\tau_j(s)} < x\} < \lfloor pN_{jt} \rfloor\right), \tag{8}$$

where $\tau_j(s) = \inf\{t : N_{jt} \geq s\}$ denotes the first time that system j is sampled s times and $\lfloor y \rfloor$ is the greatest integer less than y. Note that $\tau_j(s)$ is deterministic under the static algorithm so that $\mathbf{I}\{X_{j\tau_j(s)} < x\}$ are an independent Bernoulli random variables for $s = 1, \ldots, N_{jt}$. Applying Cramer's theorem (Dembo and Zeitouni 2009), the large deviation probability for $\hat{\xi}_{jt}$, $j = 1, \ldots, k$, can be characterized as follows (see Lemma B.1 in the online supplement):

$$\begin{aligned} \lim_{t\to\infty} \frac{1}{t} \log \mathsf{P}(\hat{\xi}_{jt} \geq x) &= -\alpha_j I_j(x) \quad \text{for } x > \xi_j \\ \lim_{t\to\infty} \frac{1}{t} \log \mathsf{P}(\hat{\xi}_{jt} \leq x') &= -\alpha_j I_j(x') \quad \text{for } x' < \xi_j, \end{aligned} \tag{9}$$

where

$$I_j(x) = p\log\left(\frac{p}{F_j(x)}\right) + (1-p)\log\left(\frac{1-p}{1-F_j(x)}\right). \tag{10}$$

PROPOSITION 2 **(Rate function)**. *Suppose (F1) and (F2) hold. For a static algorithm $\pi(\boldsymbol{\alpha})$ for some $\boldsymbol{\alpha} \in \Delta^0$,*

$$\lim_{t\to\infty} \frac{1}{t} \log \mathsf{P}(FS_t^{\pi(\boldsymbol{\alpha})}) = -\rho(\boldsymbol{\alpha}), \tag{11}$$

with $\rho(\boldsymbol{\alpha}) = \min_{j\neq 1}\{G_j(\boldsymbol{\alpha})\}$, where

$$G_j(\boldsymbol{\alpha}) = \inf_x \{\alpha_1 I_1(x) + \alpha_j I_j(x)\}. \tag{12}$$

We remark that the result of Proposition 2 generalizes the results in Pasupathy et al. (2010). Specifically, the analogous results in Pasupathy et al. (2010) require that $F_j(\cdot)$ is twice differentiable in the continuous case, since their proof relies on a Taylor expansion of the logarithmic moment generating function up to second order terms.

An important implication of Proposition 2 is that $\mathsf{P}(\mathrm{FS}_t^{\pi(\boldsymbol{\alpha})})$ behaves roughly like $\exp(-\rho(\boldsymbol{\alpha})t)$ for large values of t. Hence, it follows that $\rho(\cdot)$ is an appropriate measure of asymptotic efficiency that is closely associated with $\mathsf{P}(\mathrm{FS}_t^{\pi})$. Note that for each $x \in \mathcal{R}$, $\alpha_1 I_1(x) + \alpha_j I_j(x)$ is a continuous, linear (hence, concave) function of (α_1, α_j). Since $G_j(\boldsymbol{\alpha})$ in (12) is a point-wise infimum thereof, it is also concave. Therefore, $\rho(\boldsymbol{\alpha})$, being a minimum of $G_j(\boldsymbol{\alpha})$ for $j = 2, \ldots, k$, is concave for $\boldsymbol{\alpha} \in \Delta$. We define $\rho^* = \max_{\boldsymbol{\alpha}\in\Delta}\{\rho(\boldsymbol{\alpha})\}$.

3.3. Problem Formulation

Based on the relationship between $\rho(\boldsymbol{\alpha})$ and $\mathsf{P}(\mathrm{FS}_t^{\pi(\boldsymbol{\alpha})})$ provided in Proposition 2, we define the *relative efficiency* $\mathscr{R}_t^\pi$ for any given algorithm $\pi \in \Pi$ in stage t to be

$$\mathscr{R}_t^\pi = \frac{\rho(\boldsymbol{\alpha}_t^\pi)}{\rho^*}. \tag{13}$$

By definition, the value of $\mathscr{R}_t^\pi$ lies in the interval $[0,1]$.

We are interested in designing an algorithm that maximizes the expected relative efficiency with the sampling budget T:

$$\sup_{\pi\in\bar{\Pi}} \mathsf{E}(\mathscr{R}_T^\pi). \tag{14}$$

Note that an algorithm π is near optimal if $\boldsymbol{\alpha}_T^\pi$ is close to $\boldsymbol{\alpha}^* \in \arg\max_{\boldsymbol{\alpha}\in\Delta}\{\rho(\boldsymbol{\alpha})\}$ with high probability. However, the underlying distribution functions are not known a priori, hence neither is the function $\rho(\cdot)$ nor its maximizer $\boldsymbol{\alpha}^*$. Therefore, it is not tractable to obtain the optimal sampling algorithm for finite T using a rigorous dynamic programming approach. Alternatively, we focus on the following asymptotic criteria.

Definition 2 (Asymptotic optimality). An algorithm $\pi \in \bar{\Pi}$ is asymptotically optimal if

$$\mathsf{E}\left(\mathscr{R}_t^{\pi}\right) \to 1 \text{ as } t \to \infty. \tag{15}$$

It is not difficult to check that an asymptotically optimal algorithm is consistent; otherwise, there exists some system j for which $\alpha_{jt} \to 0$ as $t \to \infty$ with positive probability, which in turn implies that the limit of $\mathsf{E}(\mathscr{R}_t^{\pi})$ is less than one as $t \to \infty$.

4. Proposed Algorithms and Main Theoretical Results

4.1. A Naive Algorithm

We first suggest a naive algorithm that iteratively estimates the maximizer of the rate function by utilizing the history of sample observations. Define $\hat{\rho}_t(\boldsymbol{\alpha}) = \min_{j \neq b}\{\hat{G}_{jt}(\boldsymbol{\alpha})\}$ with $b = \arg\max_j\{\hat{\xi}_{jt}\}$,

$$\hat{G}_{jt}(\boldsymbol{\alpha}) = \inf_x \left\{\alpha_b \hat{I}_{bt}(x) + \alpha_j \hat{I}_{jt}(x)\right\} \text{ for } j \neq b, \tag{16}$$

and

$$\hat{I}_{jt}(x) = p \log\left(\frac{p}{\hat{F}_{jt}(x)}\right) + (1-p)\log\left(\frac{1-p}{1-\hat{F}_{jt}(x)}\right) \text{ for } j = 1, \ldots, k. \tag{17}$$

Note that $\hat{\rho}_t(\boldsymbol{\alpha})$ is an estimator of $\rho(\boldsymbol{\alpha})$ in stage t with each $F_j(\cdot)$ replaced with its empirical counterpart, $\hat{F}_{jt}(\cdot)$. Define $\hat{\boldsymbol{\alpha}}_t \in \arg\max_{\boldsymbol{\alpha} \in \Delta}\{\hat{\rho}_t(\boldsymbol{\alpha})\}$ as a maximizer of the function $\hat{\rho}_t(\boldsymbol{\alpha})$, with ties broken arbitrarily. The naive algorithm is summarized in Algorithm 2, with n_0, m, and c being tuning parameters; n_0 is the initial number of samples from each system, m is the batch size, and c controls the minimum sampling frequency. For ease of exposition, we assume T is a multiple of m.

The naive algorithm makes the current allocation ($\boldsymbol{\alpha}_t$) close to the target allocation ($\hat{\boldsymbol{\alpha}}_t$) in each stage; see equation (18). Essentially, it attempts to sample the system with greatest need of information; see, e.g., Chen and Lee (2011) for similar procedures of this kind. Note also that the naive algorithm (as well as those to be introduced in what follows) makes no use of the value of T, the problem horizon. A more forward looking approach typically requires dynamic programming, which may not be practically implementable and

ALGORITHM 2: Naive(n_0, m, c)

For each j, take n_0 samples and let $t = kn_0$.

repeat

If $N_{jt} \le c \log t$ for some j, let $\pi_{t+\ell} = \min_j\{N_{jt}\}$ for $\ell = 1,\ldots,m$, with ties broken arbitrarily.

Otherwise, solve for $\hat{\boldsymbol{\alpha}}_t \in \arg\max_{\boldsymbol{\alpha}\in\Delta}\{\hat{\rho}_t(\boldsymbol{\alpha})\}$, with ties broken arbitrarily, and let

$$\pi_{t+\ell} = \arg\max_j\{\hat{\alpha}_{jt} - \alpha_{jt}\} \tag{18}$$

for $\ell = 1,\ldots,m$. Let $t = t+m$.

until $t \le T$;

Return: $\arg\max_j\{\hat{\xi}_{jT}\}$.

hence will not be discussed in this paper. Readers interested in dynamic programming approaches are referred to Peng et al. (2018).

The following theorem shows that $\hat{\boldsymbol{\alpha}}_t$ eventually approaches the optimal allocation $\boldsymbol{\alpha}^*$, and therefore $\boldsymbol{\alpha}_t$ converges to $\boldsymbol{\alpha}^*$ as $t \to \infty$, implying the algorithm is asymptotically optimal.

Theorem 1 **(Asymptotic performance of the naive algorithm)**. *Suppose* $\mathbf{F} \in \{\mathscr{C}, \mathscr{D}\}$. *Under (F1)-(F2), the naive algorithm is consistent and asymptotically optimal.*

Remark 1 (Poor performance of the naive algorithm). From an implemention perspective, the naive algorithm requires a nested optimization in each stage. Specifically, the inner optimization loop involves solving for the infimum in (16) to find $\hat{G}_{jt}(\boldsymbol{\alpha})$ for given $\boldsymbol{\alpha}$, and the outer optimization task is to maximize $\hat{\rho}_t(\boldsymbol{\alpha}) = \min_{j\ne b}\{\hat{G}_{jt}(\boldsymbol{\alpha})\}$ over $\boldsymbol{\alpha} \in \Delta$. As we will see in Table 1 in § 5.1, this becomes a significant computational burden as the number of systems increases. From a performance standpoint, despite the theoretical guarantee on its asymptotic performance, the naive algorithm exhibits poor finite-time performance in terms of the probability of false selection. In particular, in the naive algorithm we need to estimate the distribution functions in order to evaluate the rate function, which results in relatively poor performance; see numerical results in §6.

Remark 2 (Forced sampling in the naive algorithm). An issue that can arise with the rate function estimator $\hat{\rho}_t(\cdot)$ based on the empirical distribution functions is a long period of no sampling from a particular system. In this case, the maximizer $\hat{\boldsymbol{\alpha}}_t \in \arg\max_{\boldsymbol{\alpha}\in\Delta}\{\hat{\rho}_t(\boldsymbol{\alpha})\}$ may lie at the boundary of Δ; that is, we may have $\hat{\alpha}_{jt} = 0$ for some j. In this case, the naive algorithm does not sample system j in stage t. If this problem persists in subsequent stages, the empirical distribution may not converge to its theoretical counterpart "fast enough". In order to prevent this issue, the naive algorithm ensures that each system is sampled at least $c \log t$ times in stage t. Note that the asymptotic result in Theorem 1 holds when $c \log t$ is replaced with any sublinear function of t, although such a change would affect the performance with finite t.

Remark 3 (Variants of the naive algorithm). In this paper we use the rate function estimator $\hat{\rho}_t(\cdot)$ based on empirical distribution functions. In cases with $\mathbf{F} \in \mathscr{C}$, one can also estimate the rate function based on "smooth" estimators; for example, see the kernel-type estimator in Reiss (1981). In cases with $\mathbf{F} \in \mathscr{D}$, one can also use discrete kernel estimators; for example, see Rajagopalan and Lall (1995). Smooth estimators can be useful when the sample observations are sparse, although the efficiency of these estimators depends on the choice of smoothing parameters whose optimal values are unknown a priori. It is not our intention to compare different types of distribution estimators, but we remark that the performance of the naive algorithm may vary with different choices of these estimators.

4.2. Alternative Algorithm for Continuous Distributions

We discuss here the case in which the underlying distributions are continuous with the following condition.

(F3) $F_j(\cdot)$ is twice continuously differentiable and possesses a positive continuous density $f_j(\cdot)$ over the interval $\mathscr{H}_j$.

It is trivial to check that the smoothness assumption (F3) implies (F1) and (F2). Note that $F_j(\cdot)$ must be twice continuously differentiable in order to ensure that $I_j(x)$ can be closely approximated, using a Taylor expansion, by a quadratic function of x in a neighborhood of ξ_j, which will be key to the theoretical results in this subsection.

Let $\delta := \xi_1 - \xi_2$ be the gap between the best and the second best systems and define $\rho^\delta(\boldsymbol{\alpha}) := \min_{j \neq 1}\{G_j^\delta(\boldsymbol{\alpha})\}$ with

$$G_j^\delta(\boldsymbol{\alpha}) = \frac{(\xi_1 - \xi_j)^2}{2p(1-p)\left(1/(\alpha_1 f_1^2(\xi_1)) + 1/(\alpha_j f_j^2(\xi_j))\right)}. \tag{19}$$

We provide some intuition behind the definition of $\rho^\delta(\boldsymbol{\alpha})$. Under (F3), the sample quantile from $\alpha_j T$ independent observations from system j is asymptotically normal with mean ξ_j and variance $p(1-p)/(\alpha_j T f_j^2(\xi_j))$ (see, e.g., pp. 77-79 of Serfling 2009). Hence, one may consider $G_j^\delta(\boldsymbol{\alpha})$ as a measure of divergence between two quantiles, measured in units of standard errors. The greater the value of $G_j^\delta(\boldsymbol{\alpha})$ the better we can distinguish whether $\xi_1 > \xi_j$ or not with greater confidence. This in turn suggests that $\rho^\delta(\boldsymbol{\alpha})$ is closely aligned with $\rho(\boldsymbol{\alpha})$, the rate function of the probability of false selection. This intuitive observation is formalized in the following proposition.

PROPOSITION 3 **(Characteristics of** ρ^δ**).** *Consider a set of system configurations in* $\mathscr{C}$*, where each configuration* $\mathbf{F}$ *satisfies (F1)-(F3) and* $f_j(\xi_j)$ *is in a compact set that does not include zero. Then,* $\rho^\delta(\boldsymbol{\alpha})$ *has a unique maximum* $\boldsymbol{\alpha}^\delta \in \Delta^0$ *and*

$$\frac{\rho(\boldsymbol{\alpha}^\delta)}{\rho^*} \to 1 \ \text{ as } \delta \to 0. \tag{20}$$

Proposition 3 states that one can achieve near-optimal performance with respect to $\rho(\boldsymbol{\alpha})$ by maximizing $\rho^\delta(\boldsymbol{\alpha})$ when δ is sufficiently close to 0. Note that the maximization of $\rho^\delta(\boldsymbol{\alpha})$ is much simpler than that of $\rho(\boldsymbol{\alpha})$ because the former is strongly concave in $\boldsymbol{\alpha}$ (Appendix A.1) and does not involve a nested optimization structure. In the preceding proposition, the condition that $f_j(\xi_j)$ lies in a compact set is a relatively mild restriction. We provide three examples to illustrate the nature of this condition.

EXAMPLE 2 (EXPONENTIAL DISTRIBUTIONS). First, consider two exponential systems with means $(\mu+\delta,\mu)$ for $\mu>0$, for which $(f_1(\xi_1), f_2(\xi_2)) \to ((1-p)/\mu,(1-p)/\mu)$ as $\delta \to 0$, satisfying the condition in Proposition 3. Second, consider two exponential systems with means $(\delta+1/\delta, 1/\delta)$, for which $(\xi_1,\xi_2) = (-(\delta+1/\delta)\log(1-p), -(1/\delta)\log(1-p))$. In this case, $(f_1(\xi_1), f_2(\xi_2)) \to (0,0)$ as $\delta \to 0$, violating the condition in Proposition 3. Third, consider two exponential systems with means $(2\delta,\delta)$, for which $(\xi_1,\xi_2) = (-2\delta\log(1-p), -\delta\log(1-p))$. In this case, $(f_1(\xi_1), f_2(\xi_2)) \to (\infty,\infty)$ as $\delta\to 0$, violating the condition of Proposition 3.

Based on the approximation $\rho^\delta(\cdot)$, we now propose an algorithm called Quantile Divergence (QD) for continuous distributions that iteratively estimates $\boldsymbol{\alpha}^\delta = \arg\max_{\boldsymbol{\alpha}\in\Delta}\{\rho^\delta(\boldsymbol{\alpha})\}$ from the history of sample observations. Specifically, denote $\hat{\boldsymbol{\alpha}}_t^\delta$ the estimator of $\boldsymbol{\alpha}^\delta$ in stage t. Formally, $\hat{\boldsymbol{\alpha}}_t^\delta = \arg\max_{\boldsymbol{\alpha}\in\Delta}\{\hat{\rho}_t^\delta(\boldsymbol{\alpha})\}$, where

$$\hat{\rho}_t^\delta(\boldsymbol{\alpha}) = \min_{j\neq b} \frac{(\hat{\xi}_{bt}-\hat{\xi}_{jt})^2}{2p(1-p)\left(1/(\alpha_b \hat{f}_{bt}^2(\hat{\xi}_{bt})) + 1/(\alpha_j \hat{f}_{jt}^2(\hat{\xi}_{jt}))\right)}, \tag{21}$$

ALGORITHM 3: QD-C $(n_0, m, K(\cdot), h(\cdot))$

For each j, take n_0 samples and let $t = kn_0$.

repeat

Let $b = \arg\max_j\{\hat{\xi}_{jt}\}$. If $\hat{\xi}_{bt} = \hat{\xi}_{jt}$ for system $j \neq b$, then take a sample from system

$$\pi_{t+\ell} = \arg\min_{i=j,b}\{\alpha_{it}\} \tag{23}$$

for $\ell = 1, \ldots, m$, with ties broken arbitrarily. Otherwise, estimate the density estimators in (22) using the kernel $K(\cdot)$ and the bandwidth parameter $h(\cdot)$. Solve for $\hat{\boldsymbol{\alpha}}_t^\delta = \arg\max_{\boldsymbol{\alpha}\in\Delta}\{\hat{\rho}_t^\delta(\boldsymbol{\alpha})\}$ and let

$$\pi_{t+\ell} = \arg\max_j\{\hat{\alpha}_{jt}^\delta - \alpha_{jt}\} \tag{24}$$

for $\ell = 1, \ldots, m$. Let $t = t + m$.

until $t \leq T$;

Return: $\arg\max_j\{\hat{\xi}_{jT}\}$.

---

with $b = \arg\max_j\{\hat{\xi}_{jt}\}$ and

$$\hat{f}_{jt}(y) = \frac{1}{N_{jt}} \sum_{\substack{s=1 \\ \pi_s = j}}^{t} K_{h(N_{jt})}(y - X_{js}) \tag{22}$$

is the kernel-based estimator with kernel $K(\cdot)$ and bandwidth $h(t) \geq 0$ for each t. A kernel with subscript h is called a scaled kernel and defined as $K_h(x) = h^{-1}K(x/h)$. The optimal choices of the kernel function and the bandwidth parameter depends on the true density functions that are a priori unknown (see, e.g., Silverman (1986)), but we impose standard regularity conditions on $K(\cdot)$ and $h(t)$, which are satisfied by almost any conceivable kernel such as normal, uniform, triangular, and others:

(K1) $\int |K(x)|dx < \infty$ and $\int K(x)dx = 1$

(K2) $|xK(x)| \to 0$ as $|x| \to \infty$

(K3) $h(t) \to 0$ and $th(t) \to \infty$ as $t \to \infty$.

We propose an algorithm that matches $\boldsymbol{\alpha}_t$ with $\hat{\boldsymbol{\alpha}}_t^\delta$ in each stage, simultaneously ensuring that $\hat{\boldsymbol{\alpha}}_t^\delta$ approaches $\boldsymbol{\alpha}^\delta$ as $t \to \infty$. The procedure is summarized in Algorithm 3, with n_0 and m being tuning parameters.

THEOREM 2 **(Asymptotic performance of QD-C)**. *Suppose* $\mathbf{F} \in \mathscr{C}$. *Under (F1)-(F2), the QD-C algorithm is consistent, and if (F3) is further satisfied, then*

$$\mathsf{E}(\mathscr{R}_t^\pi) \to \frac{\rho(\boldsymbol{\alpha}^\delta)}{\rho^*} \text{ as } t \to \infty. \tag{25}$$

To rephrase (25), the QD algorithm eventually allocates the sampling budget so that $\rho^\delta(\boldsymbol{\alpha})$ is maximized, and the loss in asymptotic efficiency due to maximizing $\rho^\delta(\cdot)$ instead of $\rho(\cdot)$ decreases to 0 as $\delta \to 0$. Combining the preceding theorem with Proposition 3, it can be seen that $\rho(\boldsymbol{\alpha}^\delta)/\rho^* \to 1$ as $\delta \to 0$, implying that the QD-C algorithm is near-optimal in an asymptotic regime as $t \to \infty$ and $\delta \to 0$.

Remark 4 (Relation to existing algorithms). The QD-C algorithm shares the common theme with two sampling algorithms in the literature of mean-based ordinal optimization. First, QD-C is analogous to the WD algorithm proposed by Shin et al. (2018) in the sense that the former (respectively, the latter) repeatedly maximizes the divergence between sample quantiles (respectively, sample means). Second, QD-C can be related to a family of sampling algorithms called OCBA (Chen and Lee 2011). Specifically, from the first order conditions it can be seen that $\hat{\boldsymbol{\alpha}}_t^\delta$ of the QD-C algorithm satisfies

$$\frac{(\hat{\xi}_{bt}-\hat{\xi}_{it})^2}{(\hat{\alpha}_{bt}^\delta \hat{f}_{bt}^2(\hat{\xi}_{bt}))^{-1}+(\hat{\alpha}_{it}^\delta \hat{f}_{it}^2(\hat{\xi}_{it}))^{-1}} = \frac{(\hat{\xi}_{bt}-\hat{\xi}_{jt})^2}{(\hat{\alpha}_{bt}^\delta \hat{f}_{bt}^2(\hat{\xi}_{bt}))^{-1}+(\hat{\alpha}_{jt}^\delta \hat{f}_{jt}^2(\hat{\xi}_{jt}))^{-1}} \quad \text{for} \quad i,j=2,\ldots,k \tag{26}$$

$$\hat{\alpha}_{bt}^\delta = \sqrt{\sum_{j\neq b} \left(\hat{\alpha}_{jt}^\delta\right)^2 \frac{\hat{f}_{jt}^2(\hat{\xi}_{jt})}{\hat{f}_{bt}^2(\hat{\xi}_{bt})}}. \tag{27}$$

If we further assume that $\hat{\alpha}_{bt}^\delta \gg \hat{\alpha}_{jt}^\delta$, then the above formulas are equivalent to the OCBA formulas, except that the means and variances of sample means are replaced with those of sample quantiles. Both WD and OCBA are derived based on the premise that sample means are normal, and hence, QD-C can be considered as a comparable algorithm when sample quantiles are approximately normally distributed.

Remark 5 (Bias-variance tradeoff). The major advantage of QD-C over the naive algorithm is that the estimation of $\rho^\delta(\cdot)$ is more "localized" and does not require the estimation of the entire distribution functions; see Remark 1 for a discussion about the optimization error under the naive algorithm. In other words, exploration of the function $\rho^\delta(\cdot)$ requires less of the sampling budget than that of $\rho(\cdot)$, which allows us to exploit more of the budget to maximize $\rho^\delta(\cdot)$, with a sacrifice due to the gap between $\rho^\delta(\cdot)$ and $\rho(\cdot)$. We now rigorously analyze the bias and variance in this tradeoff. For ease of exposition, let $k=2$ and, with a slight abuse of notation, denote $\rho(\alpha_{1t}) = \rho(\boldsymbol{\alpha}_t)$ for $\boldsymbol{\alpha}_t = (\alpha_{1t}, \alpha_{2t})$. Using a second-order Taylor expansion of $\rho(\cdot)$ at α_1^*, observe that

$$\rho(\alpha_1^*) - \rho(\alpha_{1t}) = -\frac{\rho''(\alpha_1^*)}{2}(\alpha_{1t}-\alpha_1^*)^2 + o((\alpha_{1t}-\alpha_1^*)^2), \tag{28}$$

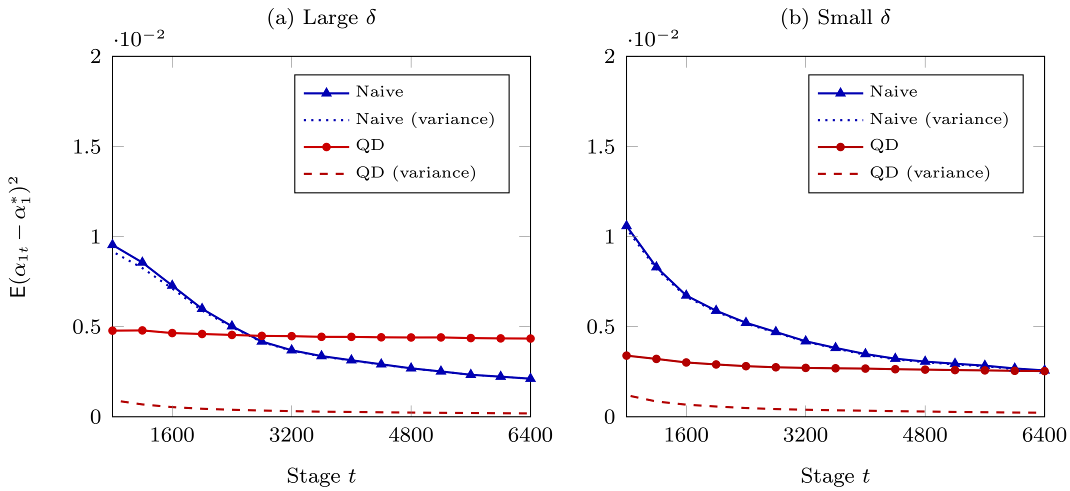

Figure 1 Bias-variance tradeoff. For each naive (blue) and QD-C (red) algorithm, the solid line is the mean squared error and the dotted (or dashed) line is its variance component. The gap between the solid and dotted (or dashed) lines corresponds to the bias component. The system configurations are (a) two normal systems with means $(0,0)$ and standard deviations $(1,3)$ and (b) two normal systems with the same means $(0,0)$ and standard deviations $(1,2)$. The 10% quantiles are $\xi = (0.13, 0.38)$ for (a) and $\xi = (0.13, 0.25)$ for (b). For both algorithms $(n_0, m) = (20, 10)$. Also, $c = 10$ for the naive algorithm and we use the normal kernel with the bandwidth parameter $h(n) = 1.06\sigma n^{-1/5}$ for QD-C, where σ is replaced with sample standard deviation.

where $\rho''(\alpha_1^*)$ is the second derivative of $\rho(\cdot)$ at α_1^*, and the term $o(\alpha_{1t} - \alpha_1^*)$ is sublinear in $(\alpha_{1t} - \alpha_1^*)$ as this difference tends to zero. Hence, the loss in the asymptotic efficiency, $\mathsf{E}(\rho(\alpha_1^*) - \rho(\alpha_{1t}))$, can be largely explained by the mean squared error (MSE), $\mathsf{E}(\alpha_{1t} - \alpha_1^*)^2$, which can be further decomposed as

$$\mathsf{E}(\alpha_{1t} - \alpha_1^*)^2 = \mathsf{E}(\alpha_{1t} - \mathsf{E}(\alpha_{1t}))^2 + (\mathsf{E}(\alpha_{1t}) - \alpha_1^*)^2. \tag{29}$$

The first term represents the variance component (due to noisy sample estimates) and the second represents the bias component (due to maximizing $\rho^\delta(\cdot)$ instead of $\rho(\cdot)$). As illustrated in Figure 1, most of the MSE contribution is due to variance under the naive algorithm. On the other hand, QD-C significantly reduces the variance, while introducing a small bias that decreases to zero as $\delta \to 0$. Overall, the QD-C algorithm has lower MSE's than the naive counterpart for small t, indicating it is more suitable in settings with small sampling budgets. Obviously, QD-C is even more attractive than the naive algorithm when δ is "small" because the loss due to the bias is "small" as well.

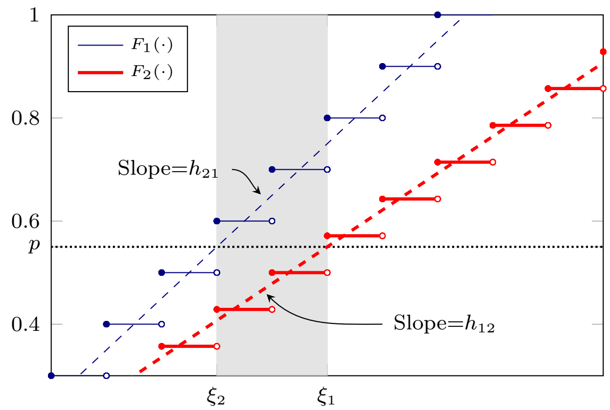

Figure 2 **Illustration of h_{12} and h_{21} defined in (30) for $p=0.55$. The step functions in thin and thick solid lines represent the distributions $F_1(\cdot)$ and $F_2(\cdot)$, respectively, and the dashed lines are the linear approximations.**

4.3. Alternative Algorithm for Discrete Distributions

We now develop an alternative algorithm for the case in which the underlying distributions are discrete, i.e., $\mathbf{F}\in\mathscr{D}$. To this end, define

$$h_{j1}=\frac{F_j(\xi_1)-F_j(\xi_j)}{\xi_1-\xi_j}$$
$$h_{1j}=\frac{F_1(\xi_1)-F_1(\xi_j)}{\xi_1-\xi_j} \tag{30}$$

for each $j\neq 1$ and let $\boldsymbol{h}=\{(h_{j1},h_{1j})\mid j\neq 1\}$. Figure 2 illustrates the definitions of h_{j1} and h_{1j} for $j=2$. Also define $\epsilon\in(0,1)$ as a constant such that, for each $j\neq 1$ and $x\in[\xi_j,\xi_1]$,

$$1-\epsilon\leq\frac{F_j(x)}{p+(x-\xi_j)h_{j1}},\frac{1-F_j(x)}{1-p-(x-\xi_j)h_{j1}}\leq 1+\epsilon$$
$$1-\epsilon\leq\frac{F_1(x)}{p+(x-\xi_1)h_{1j}},\frac{1-F_1(x)}{1-p-(x-\xi_1)h_{1j}}\leq 1+\epsilon. \tag{31}$$

The constant ϵ represents (multiplicative) errors when $F_j(x)$ is approximated by a linear function, $p+(x-\xi_j)h_{j1}$ or $p+(x-\xi_1)h_{1j}$.

We consider a set of discrete distributions that satisfies the following condition, which ensures that h_{j1} and h_{1j} defined in (30) are strictly positive for each $j\neq 1$.

(F3’) The probability mass function, $f_j(x)=F_j(x)-F_j(x-1)$, is positive for all integer $x\in\mathscr{H}_j$

Define $\rho^{\delta,\epsilon}(\boldsymbol{\alpha})=\min_{j\neq 1}\{G_j^{\delta,\epsilon}(\boldsymbol{\alpha})\}$ with

$$G_j^{\delta,\epsilon}(\boldsymbol{\alpha})=\frac{(\xi_1-\xi_j)^2}{2p(1-p)\left(1/(\alpha_1 h_{1j}^2)+1/(\alpha_j h_{j1}^2)\right)}, \tag{32}$$

which has a similar structure as (19) in the continuous case, except that $f_j(\xi_j)$ is replaced with h_{j1} or h_{1j}. To provide some intuition behind the definition of $G_j^{\delta,\epsilon}(\boldsymbol{\alpha})$, recall that the asymptotic variance of $\hat{\xi}_{jt}$ is inversely proportional to $f_j(\xi_j)$ in the continuous case. In other words, when h_{1j} or h_{j1} is high (low), samples from system j are denser (sparser) around ξ_j so that the asymptotic variance of the sample quantile is low (high). Therefore, $G_j^{\delta,\epsilon}(\boldsymbol{\alpha})$ can be seen as an appropriate measure of divergence between two quantiles. The following proposition validates this intuition rigorously. We use the following notation: $\bar{h} = \max_{j\neq 1}\{\max(h_{j1}, h_{1j})\}$, $r = \max_{j\neq 1}\{\xi_1 - \xi_j\}/\min_{j\neq 1}\{\xi_1 - \xi_j\}$, $\theta = r\bar{h}\delta$, and $u : \mathcal{R}_+ \to \mathcal{R}$ is a non-decreasing function of θ defined as

$$u(\theta) = \frac{2\theta p(1-p)(\theta^3 + 3\theta p(1-p) + p(1-p))}{3(p-\theta)^3(1-p-\theta)^3}. \tag{33}$$

Proposition 4 **(Characteristic of $\rho^{\delta,\epsilon}(\cdot)$)**. *Suppose* $\mathbf{F} \in \mathscr{D}$. *Under (F1)-(F2) and (F3'), for any* $\boldsymbol{\alpha} \in \Delta$

$$\frac{\rho(\boldsymbol{\alpha}^{\delta,\epsilon})}{\rho^*} \geq \frac{(1-u(\theta))(1-\epsilon)}{(1+u(\theta))(1+\epsilon)} \tag{34}$$

for $\theta = r\bar{h}\delta$ *sufficiently small so that* $u(\theta) < 1$.

Proposition 4 validates the approximation of $\rho(\cdot)$ by $\rho^{\delta,\epsilon}(\cdot)$ for small δ (equivalently, small θ) and ϵ. This in turn implies $\boldsymbol{\alpha}^*$ and $\boldsymbol{\alpha}^{\delta,\epsilon}$, the maximizers of $\rho(\cdot)$ and $\rho^{\delta,\epsilon}(\cdot)$, respectively, are close when δ and ϵ are sufficiently small.

Remark 6 (Tightness of the lower bound in Proposition 4). We note that $u(\theta)$ is a continuous, increasing function of $\theta \geq 0$ with $u(0) = 0$. Hence, when θ and ϵ are close to 0, the bound in (34) is tight. To see the tightness of the bound for different values of θ, we provide a numerical example in Figure 3 for the case of discrete uniform distributions. Note that the lower bound is tight for small values of θ, and while there is a noticeable gap for large values of θ, the actual values of $\rho(\boldsymbol{\alpha}^{\delta,\epsilon})/\rho^*$ are still close to one.

We now present the QD-D algorithm, which iteratively maximizes $\rho^{\delta,\epsilon}(\cdot)$ from the history of sample observations. Let $\hat{\boldsymbol{\alpha}}_t^{\delta,\epsilon} = \arg\max_{\boldsymbol{\alpha}\in\Delta}\{\hat{\rho}_t^{\delta,\epsilon}\}$ be the estimator of $\boldsymbol{\alpha}^{\delta,\epsilon *}$ in stage t, where

$$\hat{\rho}_t^{\delta,\epsilon} = \min_{j\neq b} \frac{(\hat{\xi}_{bt} - \hat{\xi}_{jt})^2}{2p(1-p)\left(1/(\alpha_b \hat{h}_{bjt}^2) + 1/(\alpha_j \hat{h}_{jbt}^2)\right)}, \tag{35}$$

ALGORITHM 4: QD-D (n_0, m)

For each j, take n_0 samples and let $t = kn_0$

repeat

Let $b = \arg\max_j\{\hat{\xi}_{jt}\}$. If $\hat{\xi}_{bt} = \hat{\xi}_{jt}$ for system $j \neq b$, then take a sample from system $\pi_{t+\ell} = \arg\min_{i=j,b}\{\alpha_{it}\}$ for $\ell = 1,\dots,m$. If there are multiple such j's, choose the one with the smallest j. Otherwise, solve for $\hat{\boldsymbol{\alpha}}_t^{\delta,\epsilon} = \arg\max_{\boldsymbol{\alpha}\in\Delta}\{\hat{\rho}_t^{\delta,\epsilon}(\boldsymbol{\alpha})\}$ and let

$$\pi_{t+1} = \arg\max_j\{\hat{\alpha}_{jt}^{\delta,\epsilon} - \alpha_{jt}\} \tag{37}$$

for $\ell = 1,\dots,m$. Let $t = t + m$

until $t \leq T$;

Return: $\arg\max_j\{\hat{\xi}_{jT}\}$.

with $b = \arg\max_j\{\hat{\xi}_{jt}\}$ and $\hat{h}_{jbt}$ and $\hat{h}_{bjt}$ defined as

$$\begin{aligned}
\hat{h}_{jbt} &= \begin{cases} (\hat{F}_{jt}(\hat{\xi}_{bt}) - \hat{F}_{jt}(\hat{\xi}_{jt}))/(\hat{\xi}_{bt} - \hat{\xi}_{jt}) & \text{if } \hat{F}_{jt}(\hat{\xi}_{bt}) - \hat{F}_{jt}(\hat{\xi}_{jt}) > 0 \\ (\hat{F}_{jt}(y_{jbt}) - \hat{F}_{jt}(\hat{\xi}_{jt}))/(y_{jbt} - \hat{\xi}_{jt}) & \text{otherwise} \end{cases} \\
\hat{h}_{bjt} &= \begin{cases} (\hat{F}_{bt}(\hat{\xi}_{bt}) - \hat{F}_{bt}(\hat{\xi}_{jt}))/(\hat{\xi}_{bt} - \hat{\xi}_{jt}) & \text{if } \hat{F}_{bt}(\hat{\xi}_{bt}) - \hat{F}_{bt}(\hat{\xi}_{jt}) > 0 \\ (\hat{F}_{bt}(\hat{\xi}_{bt}) - \hat{F}_{bt}(y_{bjt})/(\hat{\xi}_{bt} - y_{bjt}) & \text{otherwise,} \end{cases}
\end{aligned} \tag{36}$$

where $y_{jbt} := \inf\{y \geq \hat{\xi}_{bt} \mid \hat{F}_{jt}(y) - \hat{F}_{jt}(\hat{\xi}_{jt}) > 0\}$ is the smallest value of $y \geq \hat{\xi}_{bt}$ such that the empirical distribution $\hat{F}_{jt}(\cdot)$ has positive probability mass between $\hat{\xi}_{jt}$ and y. Similarly, $y_{bjt} := \sup\{y \leq \hat{\xi}_{jt} | \hat{F}_{bt}(\hat{\xi}_{bt}) - \hat{F}_{bt}(y_{bjt}) > 0\}$. This ensures $\hat{h}_{jbt}, \hat{h}_{bjt} > 0$ for each $j \neq b$. The QD-D algorithm for the discrete case is summarized in Algorithm 4, with n_0 and m being parameters of the algorithm.

We note that the event $\{\hat{\xi}_{bt} = \hat{\xi}_{jt}\}$ for some $j \neq b$ can occur with positive probability, in which case $\hat{\boldsymbol{\alpha}}_t^{\delta,\epsilon} = \arg\max_{\boldsymbol{\alpha}\in\Delta}\{\hat{\rho}_t^{\delta,\epsilon}\}$ may not be well defined. When these cases occur, the QD-D algorithm takes a sample from system j or b, whichever was sampled less. When there are multiple systems satisfying such conditions, choose j arbitrarily.

THEOREM 3 **(Asymptotic performance of QD-D)**. *Suppose* $\mathbf{F} \in \mathscr{D}$*. Under (F1)-(F2), the QD-D algorithm is consistent, and if (F3') is further satisfied, then*

$$\mathsf{E}(\mathscr{R}_t^\pi) \to \frac{\rho(\boldsymbol{\alpha}^{\delta,\epsilon})}{\rho^*} \text{ as } t \to \infty. \tag{38}$$

Note that the preceding theorem, combined with Proposition 4 implies that the QD-D algorithm is near-optimal in an asymptotic regime as $t \to \infty$ and $\theta \to 0$.

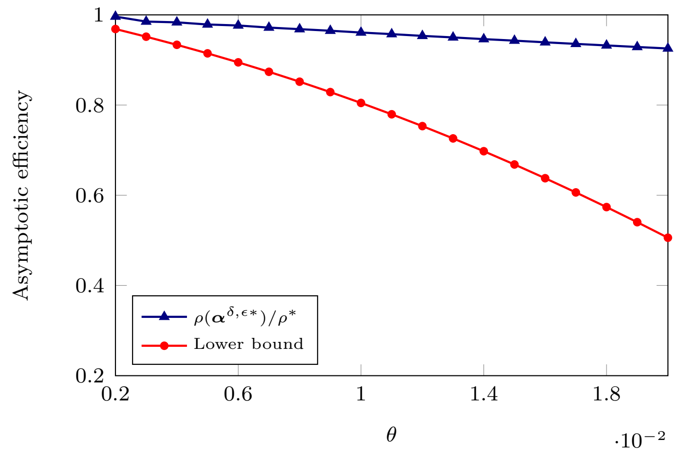

Figure 3 **Asymptotic performance of the QD-D algorithm, $\rho(\boldsymbol{\alpha}^{\delta,\epsilon})/\rho^*$, and its lower bound defined in (38), as a function of $\theta = r\bar{h}\delta$. The system configuration consists of three systems with discrete uniform distributions with 10% quantiles $(0, -\delta, -2\delta)$, where the values of δ range from 10 to 100 and the length of supports is 10^4 for all three systems. The parameters for the QD-D algorithm are $m = 10$ and $n_0 = 20$.**

5. Adaptive Heuristics for Fast Implementation

From an implementation standpoint, the QD algorithms (QD-C and QD-D) significantly improves the computation time of the naive algorithm by replacing the two-layer optimization problem with a single-layer one. The latter is a convex optimization problem that is easy to solve in small-scale problems ($k \leq 20$), however, there still exists a heavy computational burden for larger problems since the convex optimization problem should be solved repeatedly in every stage. (See Table 1 later in this section for computation times of the proposed algorithms as the number of systems increases.) Hence, we suggest variants of the QD algorithms for continuous and discrete distribution that are practically implementable with a large number of systems.

In the case with continuous distributions, the adaptive variant of the QD-C algorithm, AQD-C, is summarized in Algorithm 5. To provide some intuition behind AQD-C, recall the first order conditions for the maximizer of $\hat{\rho}_t^{\delta}(\boldsymbol{\alpha})$ in (26)-(27). While QD-C solves for (26)-(27) in every stage, AQD-C attempts to balance the left and right sides of (26)-(27) so that they are satisfied asymptotically as $t \to \infty$. The adaptive variant of the QD algorithm for discrete case (AQD-D) is similarly defined in Algorithm 6, with $\hat{f}_{jt}(\cdot)$ in AQD-C replaced with $\hat{h}_{jbt}$ or $\hat{h}_{bjt}$.

ALGORITHM 5: AQD-C $(n_0, m, K(\cdot), h(\cdot))$

For each j, take n_0 samples and let $t = kn_0$

repeat

Let $b = \arg\max_j\{\hat{\xi}_{jt}\}$. If $\hat{\xi}_{bt} = \hat{\xi}_{jt}$ for system $j \neq b$, then take a sample from system

$$\pi_{t+\ell} = \arg\min_{i=j,b}\{\alpha_{it}\} \tag{39}$$

for $\ell = 1, \ldots, m$, with ties broken arbitrarily. If

$$\alpha_{bt} < \sqrt{\sum_{j\neq b} \alpha_{jt}^2 \frac{\hat{f}_{jt}^2(\hat{\xi}_{jt})}{\hat{f}_{bt}^2(\hat{\xi}_{bt})}}, \tag{40}$$

set $\pi_{t+\ell} = b$ for $\ell = 1, \ldots, m$. Otherwise, estimate the density estimators in (22) using the kernel $K(\cdot)$ and the bandwidth parameter $h(\cdot)$. Set

$$\pi_{t+\ell} = \arg\min_{j\neq b}\left\{\frac{(\hat{\xi}_{bt} - \hat{\xi}_{jt})^2}{1/(\alpha_{bt}\hat{f}_{bt}^2(\hat{\xi}_{bt})) + 1/(\alpha_{jt}\hat{f}_{jt}^2(\hat{\xi}_{jt}))}\right\} \tag{41}$$

for $\ell = 1, \ldots, m$. Let $t = t + m$

until $t \leq T$;

Return: $\arg\max_j\{\hat{\xi}_{jT}\}$.

ALGORITHM 6: AQD-D (n_0, m)

For each j, take n_0 samples and let $t = kn_0$

repeat

Let $b = \arg\max_j\{\hat{\xi}_{jt}\}$. If $\hat{\xi}_{bt} = \hat{\xi}_{jt}$ for some $j \neq b$, then take a sample from system $\pi_{t+\ell} = \arg\min_{i=j,b}\{\alpha_{it}\}$ for $\ell = 1, \ldots, m$. Otherwise, set $\pi_{t+\ell} = b$ for $\ell = 1, \ldots, m$ if

$$\alpha_{bt} < \sqrt{\sum_{j\neq b} \alpha_{jt}^2 \frac{\hat{h}_{jbt}^2}{\hat{h}_{bjt}^2}}. \tag{42}$$

Otherwise, set

$$\pi_{t+\ell} = \arg\min_{j\neq b}\left\{\frac{(\hat{\xi}_{bt} - \hat{\xi}_{jt})^2}{1/(\alpha_{bt}\hat{h}_{bjt}^2) + 1/(\alpha_{jt}\hat{h}_{jbt}^2)}\right\} \tag{43}$$

for $\ell = 1, \ldots, m$. Let $t = t + m$

until $t \leq T$;

5.1. Discussion

Comparison to QD. In the case with $k = 2$ systems, the QD-C and AQD-C algorithms are identical. To see this, note that, when (40) is satisfied for some stage t, AQD-C takes a sample from system b. Further, (40) implies $\hat{\boldsymbol{\alpha}}_t^\delta = \arg\max_{\boldsymbol{\alpha}\in\Delta}\{\hat{\rho}_t^\delta\}$ satisfies $\hat{\alpha}_{bt}^\delta > \alpha_{bt}$ and $\hat{\alpha}_{jt}^\delta < \alpha_{jt}$ for $j \neq b$, and therefore, QD-C also samples system b. Since the preceding argument is true for each t, one can conclude that QD-C and AQD-C are identical when

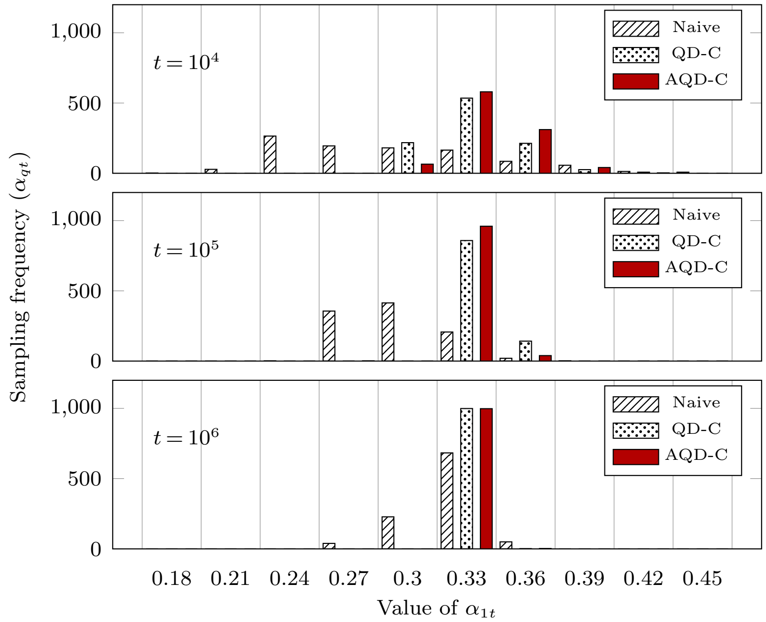

Figure 4 **The sampling frequencies (α_{1t}) under the naive, QD-C, and AQD-C algorithms. The configuration is characterized by $k = 4$ normally distributed systems with $\boldsymbol{\mu} = (0, \ldots, 0)$ and $\boldsymbol{\sigma} = (1, \ldots, 1.2)$. In this case, the 10% quantiles are $\xi = (0.13, 0.38, 0.38, 0.38)$ and $\alpha_1^* = 0.33$. For all algorithms $(n_0, m) = (20, 10)$. Also, $c = 10$ for the naive algorithm and we use the normal kernel with the bandwidth parameter $h(n) = 1.06\sigma n^{-1/5}$ for QD-C and AQD-C, where σ is replaced with sample standard deviation.**

$k = 2$. The case with $k > 2$ is less obvious, but we show via numerical example that the allocations under QD-C and AQD-C coincide eventually as $t \to \infty$ (Figure 4). Observe that the distributions of α_{1t} under the QD-C and AQD-C algorithms are almost identical. Also, Figure 4 illustrates that $\alpha_{1t} \to \alpha_1^* = 0.33$ as $t \to \infty$ for all three algorithms, but the convergence rate under the naive algorithm is slower than those of the QD-C and AQD-C, which is consistent with Remark 1. A similar argument applies between the QD-D and AQD-D algorithms.

Scalability. Compared to the naive and two QD algorithms, the AQD algorithms exhibit orders of magnitude decrease in CPU time for larger k. Table 1 summarizes the computation time of the algorithms as a function of the number of systems with normal distributions with zero means and 10% percentiles $\xi_1 - \xi_j = 0.13$ for all $j \neq 1$. The relative CPU times illustrate the dramatic difference between the AQD and the other algorithms, in particular scalability of the AQD algorithms. Indeed, we observe such a dramatic improvement in CPU time with different system configurations.

Generality. In our main theoretical results in §4, certain conditions on the underlying distributions, (F1)-(F3) or (F3'), are imposed in order to characterize the performance in

Table 1 CPU times (in seconds) per stage. CPU times are estimated by taking averages over 100 simulation trials. (Operating system: Windows 7; processor: Intel core i7 2.7GHz; memory: 32GB RAM; language: MATLAB.)

Algorithms	$k=4$	$k=40$	$k=400$	$k=4000$
Naive	0.057	0.064	9.287	2379
QD	0.012	0.043	4.567	206
AQD	0.003	0.003	0.010	0.135

terms of the relative efficiency defined in (13). However, it is important to note that the naive, QD-C, and QD-D algorithms are consistent so that $\mathsf{P}(\mathrm{FS}_t^\pi) \to 0$ as $t \to \infty$ under the mild conditions (F1)-(F2); see Theorems 1-3. We remark that the AQD-C and AQD-D algorithms can be applied in general settings where (F3) or (F3') is not necessarily satisfied.

6. Comparison with Benchmark Algorithms

6.1. Benchmark Sampling Algorithms

We compare the proposed algorithms with two benchmark algorithms: the equal allocation (EA) and a heuristic algorithm based on Hoeffding's inequality (HH). The HH algorithm is introduced as a simple benchmark that uses confidence intervals of sample quantiles; see Algorithm 7. It takes as input three tuning parameters: the number of initial samples n_0, the batch size m, and $\beta \in (0,1)$.

To provide some intuition behind the HH algorithm, let us denote ξ^p as the pth quantile and $\hat{\xi}^p$ as the sample quantile from N independent samples. From Hoeffding's inequality, observe that

$$\mathsf{P}(\hat{\xi}^{p+\zeta} < \xi^p) = \mathsf{P}\left(\sum_{i=1}^{N} (\mathbf{I}\{X_i \le \xi^p\} - p) \ge \zeta N\right) \le e^{-2N\zeta^2}. \tag{44}$$

Therefore, $\hat{\xi}^{p+\zeta}$ is greater than ξ^p with probability greater than or equal to $1-\exp(-2N\zeta^2)$. In other words, for a given value of ζ, the value of $1-\exp(-2N\zeta^2)$, or equivalently, $N\zeta^2$, can be considered as a proxy for the confidence level. The HH algorithm is designed to take a sample from the system with the least value of such a measure, $\arg\min_j \{N_{jt}\zeta_{jt}^2\}$, in stage t. The value of ζ_{jt} for each j depends on the parameter β. The optimal value of β may depend on underlying probability distributions, and hence, it may not be known a priori. In the following numerical experiments, we test the HH algorithm for different values of $\beta = 0.25, 0.5, 0.75$ but only show the results for $\beta = 0.5$ which gives a better overall performance in terms of the probability of false selection.

ALGORITHM 7: HH (n_0, m, β)

For each j, take n_0 samples and let $t = kn_0$.

repeat

Let $b = \arg\max_j \{\hat{\xi}_{jt}\}$ and $b' = \arg\max_{j \neq b} \{\hat{\xi}_{jt}\}$.

Fix $v_t = \beta \hat{\xi}_{bt} + (1-\beta)\hat{\xi}_{b't}$.

Define $\{\zeta_{jt}\}_{j=1}^{k}$ as follows

$$\zeta_{jt} = \begin{cases} p - \hat{F}_{bt}(v_t) & \text{for } j = b \\ \hat{F}_{jt}(v_t) - p & \text{for } j \neq b. \end{cases} \tag{45}$$

Set $\pi_{t+\ell} = \arg\min_j \{N_{jt}\zeta_{jt}^2\}$ for $\ell = 1, \ldots, m$ and let $t = t + m$.

until $t \leq T$;

Return: $\arg\max_j \{\hat{\xi}_{jT}\}$.

6.2. Numerical Experiments

We test our procedures for continuous (normal and Student's t) and discrete (uniform and Poisson) distributions. $\mathsf{P}(\mathrm{FS}_t^\pi)$ is estimated by counting the number of false selections out of v simulation trials, which is chosen so that:

$$\sqrt{\frac{P_t(1-P_t)}{v}} \leq \frac{P_t}{10}, \tag{46}$$

where P_t is the order of magnitude of $\mathsf{P}(\mathrm{FS}_t^\pi)$. This implies standard errors for the estimates of $\mathsf{P}(\mathrm{FS}_t^\pi)$ are at least ten times smaller than the value of $\mathsf{P}(\mathrm{FS}_t^\pi)$ so that we have sufficiently high confidence that the results are not driven by simulation error.

For the naive algorithm, we set $c = 10$ in all cases. For the QD-C and AQD-C algorithms for continuous distributions, we use the normal kernel with the bandwidth parameter $h(n) = 1.06\sigma n^{-1/5}$, where σ is replaced with the usual sample standard deviation. The tuning parameters chosen in our numerical experiments are by no means optimal but we provide numerical examples in Appendix C of the online supplement to show the sensitivity of performance with respect to these parameters.

In the following experiments, we test the naive and QD algorithms for cases with small number of systems ($k = 4$) since they are not practically implementable with larger problem instances; see CPU times of these algorithms in Table 1.

Normal distributions. We consider configurations where the quantiles are monotonically decreasing; in particular, we set $\mu_j = 0$ and $\sigma_j = 1 + \gamma(j-1)$ for $j = 1, \ldots, k$ for some $\gamma > 0$. We vary $\gamma = 0.1, 1$ which corresponds to $\delta = 0.128, 1.28$, the gap in quantile between the best and second best systems. Also, to show the performance of our procedure for

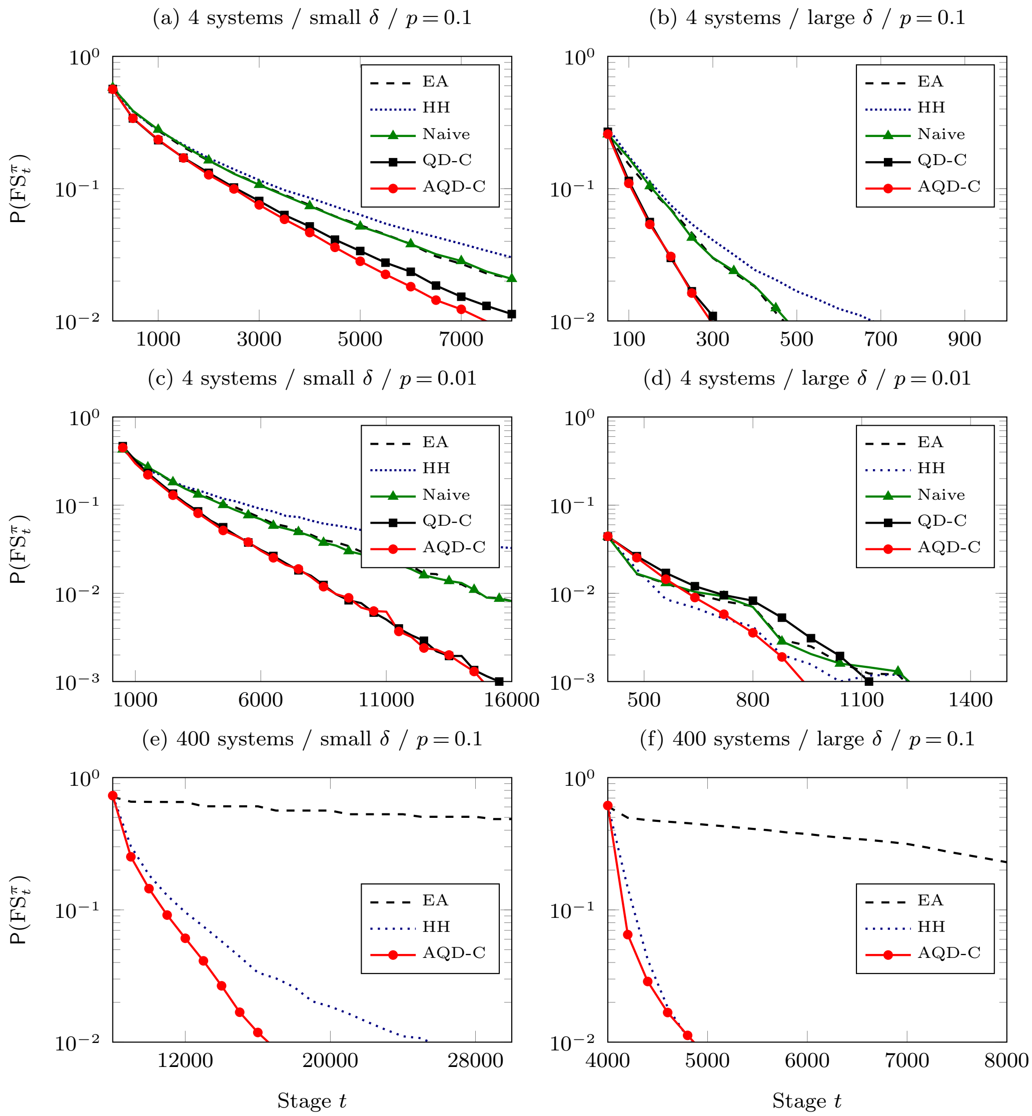

Figure 5 $\mathsf{P}(\mathbf{FS}_t^\pi)$ **as a function of stage** t **on a log-linear scale in the normal cases. In all cases, means of systems are equally set to zero and standard deviations are** $\sigma_j = 1 + \gamma(j-1)$ **for** $j = 1, \ldots, k$**, where we vary** $\gamma = 0.1, 1$ **for the cases with small and large values of** δ**, respectively. We set** $k = 4$ **and** $p = 0.1$ **in panels (a)-(b),** $k = 4$ **and** $p = 0.01$ **in panels (c)-(d), and** $k = 400$ **and** $p = 0.1$ **in panels (e)-(f). We take** $m = 10$ **samples per batch for all cases and set** $n_0 = 20$ **for (a)-(b),** $n_0 = 100$ **for (c)-(d), and** $n_0 = 10$ **for (e)-(f). For the HH algorithm, the parameter** β **is tuned to be** 0.5 **for both configurations.**

the cases with extreme quantiles and large number of systems, we vary $p = 0.1, 0.01$ and $k = 4, 400$.

In all cases of Figure 5, the QD-C and AQD-C algorithms outperform the others. In light of Theorem 2, the QD-C and AQD-C policies are near-optimal asymptotically when the gap δ in quantile between the best and second-best systems is sufficiently small. This is consistent with the observations from panels (a),(c), and (e) of Figure 5, where δ is sufficiently small and the finite-time performance of the two algorithms is competitive in terms of the probability of false selection. Somewhat surprisingly, the performance of QD-C and AQD-C are also preferable in cases with relatively large values of δ; see panels (b),(d), and (f) in Figure 5. These results indicate that the finite-time performance of the two algorithms is competitive for both small and large values of δ, which complements the asymptotic results of Theorem 2 as $\delta \to 0$. Further, note that AQD-C slightly outperforms QD-C in some cases in Figure 5.

Notice that the poor performance of the naive algorithm is anticipated by Remark 1. The performance of the HH algorithm varies significantly for different configurations; it performs competitively with the AQD-C algorithm in panel (f) of Figure 5, but it is even worse than EA in panels (a)-(c). This shows that the performance of the HH algorithm highly depends on the choice of parameter β, which is not known a priori. Lastly, note that the EA algorithm performs very poorly in the case with $k = 400$ since too many samples are allocated to systems that are far from the best one, while the AQD-C algorithm safely discards seemingly non-best systems in early stages.

Heavy tailed distributions. In Figure 6, we consider two system configurations, each with four Student t-distributed systems. In both cases, we set means $(0,0,0,0)$ and standard deviations $(1,2,3,4)$ with $p = 0.1$ so that the pth quantile values monotonically decrease with system index.

It is noteworthy that the performance of quantile-based algorithms is robust against the presence of heavy tails in the sense that $\mathsf{P}(\mathrm{FS}_t^\pi)$ converges to zero at an exponential rate (Proposition 2), as opposed to mean-based procedures under which the probability of false selection decreases to zero only at a polynomial rate (Broadie et al. 2007). This can be seen in Figure 6: On the left panel, the degrees of freedom is set to $\nu = 1$ so that the distributions are significantly heavy-tailed, while we set $\nu = 20$ for the less heavy-tailed distributions on the right panel of Figure 6. In both cases, the AQD-C and QD-C algorithms outperform the others, with AQD-C performing slightly better than QD-C on the left panel. Also, note that $\mathsf{P}(\mathrm{FS}_t^\pi) \approx c_1 \exp(-c_2 t)$ implies that $\log \mathsf{P}(\mathrm{FS}_t^\pi)$ is approximately linear in t. Hence, it can be seen that $\mathsf{P}(\mathrm{FS}_t^\pi)$ converges to zero at an exponential rate under QD-C and AQD-C, indicating these algorithms can be attractive alternatives to mean-based procedures in many applications where heavy tails are prevalent.

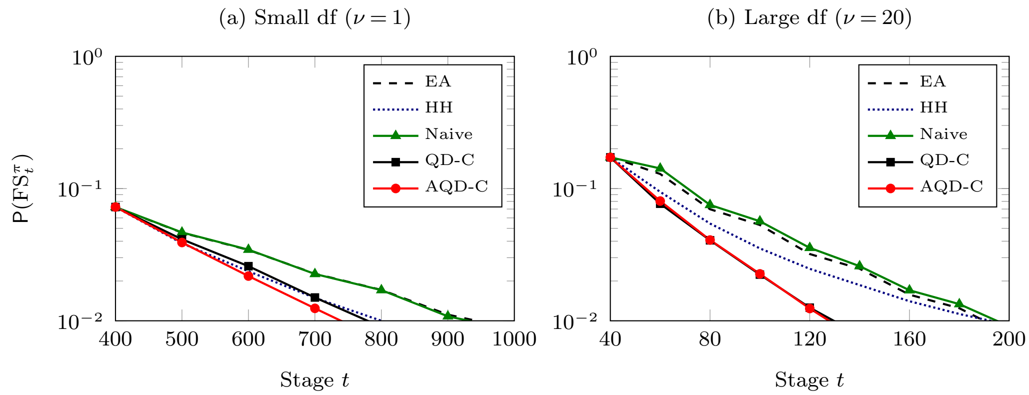

Figure 6 $\mathsf{P}(\mathbf{FS}_t^\pi)$ **as a function of stage** t **on a log-linear scale in the** t**-distribution cases. For both cases, there are four systems with means** $(0,0,0,0)$ **and standard deviations** $(1,2,3,4)$**, while the degrees of freedom are** $\nu = 1$ **on the left panel and** $\nu = 20$ **on the right panel. We set** $p = 0.1$, $n_0 = 10$, **and** $m = 10$ **for both cases. For the HH algorithm, the parameter** β **is tuned to be** 0.5.

Discrete cases: uniform and Poisson distributions. Recall from Theorem 3 that the asymptotic performance of the QD-D algorithm is characterized by $\rho(\boldsymbol{\alpha}^{\delta,\epsilon})/\rho^*$, which depends on the gap in means between the best and second-best systems (δ) and the error due to linear approximation of distribution functions (ϵ). To illustrate the effect of ϵ, we consider two families of discrete distributions: uniform distribution functions which can be closely approximated by a linear function (i.e., small ϵ) and Poisson distributions for which ϵ is relatively large. Also, to illustrate the effect of δ, we consider small and large values of δ for each of the two distributions. We fix $p = 0.1$ and $k = 4$ in all of these cases.

In the top panels of Figure 7, two configurations with uniform distributions are considered. In each case, the support of the distribution function is $\{1, \ldots, u_j\}$ for system j with $u_j = 1000 - \delta(j-1)$; we let $\delta = 50$ in (a) and $\delta = 200$ in (b). The values of ϵ are small (in the order of 10^{-3}) in both cases, while $\theta = 0.0064, 0.15$ for (a) and (b), respectively. Further, the asymptotic efficiencies characterized in Theorem 3, $\rho(\boldsymbol{\alpha}^{\delta,\epsilon*})/\rho^*$, are 0.89 and 0.61 for (a) and (b), respectively, because the gap in quantiles δ (or equivalently, θ) is smaller in the first configuration. It is interesting to observe that the performance of the AQD-D algorithm is still superior to the others in terms of the probability of false selection even with the relatively low asymptotic efficiency in Figure 7(b).

In the bottom panels of Figure 7, each system j is characterized by a Poisson distribution with parameter λ_j. In particular, we let $\lambda_j = 1000 - \delta(j-1)$ with $\delta = 5, 20$ in panels (c) and (d), respectively. As opposed to the uniform case, the values of ϵ are relatively large due to the non-linearity of Poisson distribution functions; in particular, $\epsilon = 0.101, 1.814$

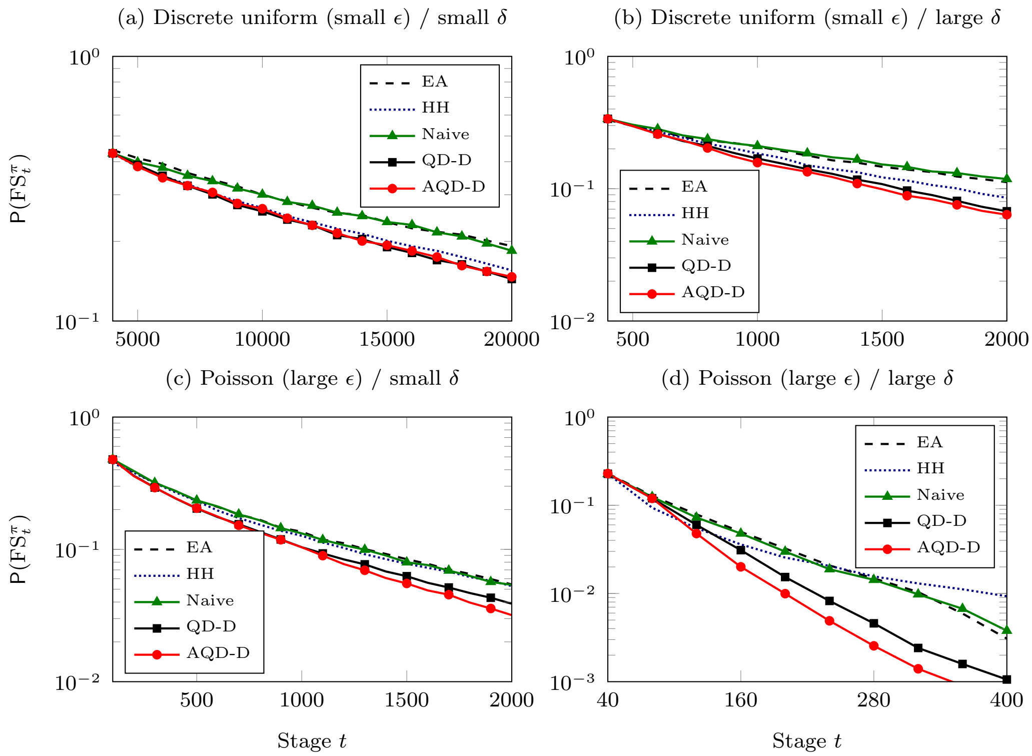

Figure 7 $\mathsf{P}(\mathsf{FS}_t^\pi)$ **as a function of stage** t **on a log-linear scale in the discrete cases. In (a) and (b), the system configuration is characterized by** $k=4$ **discrete uniform distributions, each on the range** $[0,u_j]$ **with** $u_j=1000-\delta(j-1)$ **for** $j=1,\ldots,k$**. We set** $\delta=50,200$ **for (a) and (b), respectively. In (c) and (d), the system configuration is characterized by** $k=4$ **Poisson distributions, each with parameter** $\lambda_j = 1000-\delta(j-1)$**. We set** $\delta=5,20$ **for (c) and (d), respectively. For all cases** $p=0.1$**. We set** $n_0=20$ **and** $m=10$ **for (a)-(c) and** $n_0=10$ **and** $m=1$ **for (d). For the HH algorithm, the parameter** β **is tuned to be** 0.5**.**

for (c) and (d), respectively. The large values of ϵ deteriorates the asymptotic efficiencies significantly; $\rho(\boldsymbol{\alpha}^{\delta,\epsilon *})/\rho^* = 0.72, 0.56$ for (c) and (d). Nevertheless, observe that the finite-time performance of AQD-D in terms of the probability of false selection is still superior to the other algorithms, indicating its performance is robust against a wide spectrum of ϵ and δ.

7. Concluding Remarks

We have shown how the problem of minimizing the probability of false selection can be analyzed using large deviations theory and certain approximations thereof. By analyzing the rate function of the probability of false selection in an asymptotic regime, we obtain a tractable objective function and structural insights that guide algorithm design. Although it is tractable, the rate function is computationally expensive to estimate from sample observations. A significant contribution of our work is that a nearly-optimal algorithm is obtained using a simple alternative to the rate function, which can be estimated using a surprisingly small number of sample observations with high accuracy.

In a wide range of applications, selecting multiple systems is likely to be a topic of interest. The theoretical results in our work can serve as a base for the case with multiple selections. On a methodological level, one can use different nonparametric estimators for quantiles, other than the sample quantile used in this paper, which can be leveraged to design even more efficient algorithms.

References

Batur D, Choobineh F (2010) A quantile-based approach to system selection. *Eur. J. Oper. Res.* 202:764–772.

Bechhofer RE (1954) A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Annals of Math. Statist.* 25:16–39.

Bekki JM, Fowler JW, Mackulak GT, Nelson BL (2007) Using quantiles in ranking and selection procedures. Henderson SG, Biller B, Hsieh MH, Tew JD, Barton RR, eds., *Proc. 2007 Winter Simulation Conf.*, 1722–1728 (Piscataway, NJ: IEEE).

Broadie M, Han M, Zeevi A (2007) Implications of heavy tails on simulation-based ordinal optimization. Henderson SG, Biller B, Hsieh MH, Tew JD, Barton RR, eds., *Proc. 2007 Winter Simulation Conf.*, 439–447 (Piscataway, NJ: IEEE).

Bubeck S, Cesa-Bianchi N (2012) Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends Machine Learn.* 5:1–122.

Chen CH, Lee LH (2011) *Stochastic simulation optimization: an optimal computing budget allocation*, volume 1 (World scientific).

Chen CH, Lin J, Yücesan E, Chick SE (2000) Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Discrete Event Dynamic Systems* 10:251–270.

Dembo A, Zeitouni O (2009) *Large deviations techniques and applications* (New York: Springer Science & Business Media).

Frazier P, Powell WB, Dayanik S (2008) A knowledge-gradient policy for sequential information collection. *SIAM J. Control Optim.* 47:2410–2439.

Glynn P, Juneja S (2004) A large deviations perspective on ordinal optimization. Ingalls RG, Rossetti MD, Smith JS, Peters BA, eds., *Proc. 2004 Winter Simulation Conf.*, 577–585 ((IEEE, Piscataway, NJ)).

Glynn P, Juneja S (2015) Ordinal optimization—empirical large deviations rate estimators, and stochastic multi-armed bandits. *arXiv:1507.04564* .

Kim SH, Nelson BL (2006) Selecting the best system. Henderson SG, Nelson BL, eds., *Handbooks in Operations Research and Management Science: Simulation*, volume 13, chapter 17, 501–534 (Boston: Elsevier).

Lee S, Nelson BL (2014) Bootstrap ranking & selection revisited. Tolk A, Diallo SY, Ryzhov IO, Yilmaz L, Buckley S, Miller JA, eds., *Proc. 2014 Winter Simulation Conf.*, 3857–3868 (Piscataway, NJ: IEEE).

Pasupathy R, Hunter SR, Pujowidianto NA, Lee LH, Chen CH (2015) Stochastically constrained ranking and selection via score. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 25(1):1.

Pasupathy R, Szechtman R, Yücesan E (2010) Selecting small quantiles. Johansson B, Jain S, Montoya-Torres J, Hugan J, Yücesan E, eds., *Proc. 2010 Winter Simulation Conf.*, 2762–2770 (Piscataway, NJ: IEEE).

Peng Y, Chen C, Fu M, Hu J, Ryzhov I (2019) Efficient sampling allocation procedures for selecting the optimal quantile, *Submitted to INFORMS J. Comput.*

Peng Y, Chong EKP, Chen CH, Fu MC (2018) Ranking and selection as stochastic control. *IEEE Transactions on Automatic Control* 63(8):2359–2373.

Rajagopalan B, Lall U (1995) A kernel estimator for discrete distributions. *Nonparametric Statist.* 4(4):409–426.

Reiss R (1981) Nonparametric estimation of smooth distribution functions. *Scandinavian J. of Statist.* 116–119.

Rinott Y (1978) On two-stage selection procedures and related probability-inequalities. *Commun. Statist.* 7:799–811.

Serfling RJ (2009) *Approximation theorems of mathematical statistics* (New York: John Wiley & Sons).

Shin D, Broadie M, Zeevi A (2016) Tractable sampling strategies for quantile-based ordinal optimization. *Proc. 2016 Winter Simulation Conf.*, 847–858 (IEEE Press).

Shin D, Broadie M, Zeevi A (2018) Tractable sampling strategies for ordinal optimization. *Oper. Res.* 66(6):1693–1712.

Silverman BW (1986) *Density estimation for statistics and data analysis* (New York: CRC press).

Szorenyi B, Busa-Fekete R, Weng P, Hüllermeier E (2015) Qualitative multi-armed bandits: A quantile-based approach. *32nd Internat. Conf. on Machine Learn.*, 1660–1668.

Yu JY, Nikolova E (2013) Sample complexity of risk-averse bandit-arm selection. *Proc. 23rd Internat. Joint Conf. on Artificial Inteligence*, 2576–2582 (AAAI Press).

Submitted to *INFORMS Journal on Computing*
manuscript (Please, provide the manuscript number!)

Online Supplement for
Practical Nonparametric Sampling Strategies for Quantile-based Ordinal Optimization

Dongwook Shin
School of Business and Management, Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon, Hong Kong
{dwshin@ust.hk}

Mark Broadie, Assaf Zeevi
Graduate School of Business, Columbia University, New York, NY 10025
{mnb2@columbia.edu, assaf@gsb.columbia.edu}

Key words: Quantile, ordinal optimization, tractable procedures, large deviations theory

---

The online supplement comprises three parts: Appendix A states additional theoretical results; Appendix B contains the proofs of main theoretical results and auxiliary lemmas; and Appendix C contains additional numerical results to provide practical guideline on the implementation of proposed algorithms.

Appendix A: Additional Theoretical Results

A.1. Properties of the Function $\rho^\delta(\cdot)$

Recall the definition of $\rho^\delta(\boldsymbol{\alpha}) = \min_{j\neq 1}\{G_j^\delta(\boldsymbol{\alpha})\}$, where

$$G_j^\delta(\boldsymbol{\alpha}) = \frac{(\xi_1 - \xi_j)^2}{2p(1-p)(1/(\alpha_1 f_1^2(\xi_1)) + 1/(\alpha_j f_j^2(\xi_j)))}.$$

It is trivial to check that $G_j^\delta(\boldsymbol{\alpha})$ is a smooth function of $\boldsymbol{\alpha} \in \Delta^0$. However, $\rho^\delta(\boldsymbol{\alpha})$ is only piecewise smooth because its derivative does not exist at $\boldsymbol{\alpha}$ such that $G_i^\delta(\boldsymbol{\alpha}) = G_j^\delta(\boldsymbol{\alpha})$ for some $i \neq j$.

Lemma A.1. *Assume (F1)-(F3). Then, $G_j^\delta(\boldsymbol{\alpha})$ is strongly concave in $\boldsymbol{\alpha}$.*

From the previous lemma, it is straightforward to see that $\rho^\delta(\boldsymbol{\alpha})$, being the minimum of $G_j^\delta(\boldsymbol{\alpha})$ over $j \neq 1$, is also strongly concave.

A.2. Strong Consistency of the Proposed Algorithms

In the main text we have shown the (weak) consistency of the naive, QD-C, and QD-D algorithms, respectively in Theorems 1, 2, and 3, respectively; that is, $N_{jt}^{\pi} \to \infty$ in probability as $t \to \infty$. In this section, we are interested in strengthening of these results to convergence with probability one. To this end, we need a stronger notion of consistency:

Definition A.1 (Strong consistency). An algorithm $\pi \in \Pi$ is strongly consistent if $N_{jt}^{\pi} \to \infty$ almost surely for each j as $t \to \infty$.

Note that the naive and QD-D algorithms can be shown to be strongly consistent without additional conditions. However, for the strong consistency of the QD-C algorithm, which is structured around the kernel-based density estimators, we need the following additional conditions:

(F4) $F_j(\cdot)$ possesses a positive, uniformly continuous density $f_j(\cdot)$ over the interval $\mathscr{H}_j$.

(K4) K is of bounded variation.

(K5) $K(x) \to 0$ as $|x| \to \infty$.

(K6) $\sum_{n=1}^{\infty} \exp(-\gamma t h(t)^2) < \infty$ for every $\gamma > 0$.

Note that (F4) is different than (F3) because it does not require twice differentiable $F_j(\cdot)$, but instead requires uniform continuity of the density function. Note also that (K5) does not follow from (K1); a sufficient condition for (K5) is that the kernel K is uniformly continuous. The last condition (K6) is stronger than (K3); for instance, the sequence of $h(t) = \log(t)/t$ satisfies (K3) but not (K6). The stronger consistency results are summarized in the following proposition.

Proposition A.1 **(Strong consistency of the proposed algorithms)**. *The naive algorithm is strongly consistent. For $\mathbf{F} \in \mathscr{D}$, QD-D algorithm is strongly consistent under (F1)-(F2). For $\mathbf{F} \in \mathscr{C}$, QD-C algorithm is strongly consistent if underlying distributions satisfy (F1),(F2), and (F4) and the kernel and bandwidth parameters satisfy (K1)-(K6).*

A.3. Strengthening of Proposition 3

An important implication of Proposition 3 is that, after taking $T \to \infty$, the probability of false selection (on a logarithmic scale) depends on the underlying distributions only through the densities at quantiles, $\{f_j(\xi_j)\}_{j=1}^k$, as the quantiles get closer to each other. To formalize this property more precisely, consider a stylized sequence of system distributions indexed by t, denoted as $\{F_{1,t}(\cdot), \dots, F_{k,t}(\cdot)\}_{t=1}^{\infty}$. Let $\xi_{j,t}$ denote the p-th quantile of the distribution function $F_{j,t}(\cdot)$ with $\xi_{1,t} \geq \cdots \geq \xi_{k,t}$ and define $\delta_t = \xi_{1,t} - \xi_{2,t}$. For $t = 1$, we fix $F_{j,1}(\cdot) = F_j(\cdot)$ so that $\xi_{j,1} = \xi_j$ for each j and $\delta_1 = \delta$. For $t \geq 2$, we fix $F_{1,t}(\cdot) = F_{1,1}(\cdot)$ for system 1 and $F_{j,t}(\cdot)$ for system $j \neq 1$ is shifted from $F_j(\cdot)$ so that $F_{j,t}(x - \delta_1 + \delta_t) = F_j(x)$. Essentially, this makes each $\xi_{j,t}$ approaches $\xi_{1,t}$ as $t \to 0$, while maintaining the ranking, $\xi_{1,t} > \xi_{2,t} \geq \cdots \geq \xi_{k,t}$, for each t. For brevity, we let the system configuration, $(F_{1,t}(\cdot), \dots, F_{k,t}(\cdot))$, be characterized by δ_t and define $\mathsf{P}(\mathrm{FS}_t; \delta_t)$ as the probability of false selection under the configuration, $(F_{1,t}(\cdot), \dots, F_{k,t}(\cdot))$.

Proposition A.2 **(Strengthening of Proposition 3)**. *Under assumptions (F1)-(F3), if $t\delta_t^2 \to \infty$ and $\delta_t \to 0$ as $t \to \infty$, then for any static algorithm $\pi(\boldsymbol{\alpha})$ that satisfies $N_{jt}/t \to \alpha_j$ as $t \to \infty$ for some $\boldsymbol{\alpha} \in \Delta^0$,*

$$\frac{1}{t\delta_t^2} \log \mathsf{P}(FS_t^{\pi(\boldsymbol{\alpha})}; \delta_t) \to -\rho^{\delta}(\boldsymbol{\alpha}) \text{ as } t \to \infty. \tag{A.1}$$

The proof for the preceding proposition is provided in §B.1. Proposition A.2 suggests a relationship between the sampling budget T and the difference in quantiles δ. In particular, the asymptotic behavior of $\mathsf{P}(\mathrm{FS}_T^\pi)$ is related to $T\delta^2$ for sufficiently small δ and large T; if the difference in quantiles halves, one may need approximately four times more sampling budget to lower $\mathsf{P}(\mathrm{FS}_T^\pi)$ to a certain level.

Appendix B: Proofs

This section is organized as follows. In §B.1 we present proofs for main results and those for auxiliary lemmas are collected in §B.2. We use the following notation for the purpose of asymptotic analysis: for real-valued functions $\ell_1(x)$ and $\ell_2(x)$, we write $\ell_1(x) = o(\ell_2(x))$ if $|\ell_1(x)|/|\ell_2(x)| \to 0$ as $\ell_2(x) \to 0$.

B.1. Proofs for Main Results

Proof of Proposition 1. Let $\mathscr{F}_t$ be the σ-field generated by the samples and sampling decisions taken up to stage t (i.e., $\{(\pi_\tau, X_{\pi_\tau,\tau})\}_{\tau=1}^t$), with the convention that $\mathscr{F}_0$ is the nominal σ-algebra associated with underlying probability space. Fix j and define a random variable P_n so that $N_{jP_n} = n$. (Note that P_n depends on the index j.) We show that $\{X_{jP_n}\}_{n=1}^\infty$ is a sequence of independent and identically distributed random variables.

First, observe that the characteristic function of X_{j,P_n} is

$$\begin{aligned}\mathsf{E}\left(e^{i\theta X_{jP_n}}\right) &= \sum_{t=1}^\infty \mathsf{E}\left(e^{i\theta X_{jt}} | P_n = t\right) \mathsf{P}\left(P_n = t\right) \\ &= \sum_{t=1}^\infty \mathsf{E}\left(e^{i\theta X_{jt}}\right) \mathsf{P}\left(P_n = t\right) \\ &= \mathsf{E}\left(e^{i\theta X_j}\right),\end{aligned} \tag{B.1}$$

where the second equality follows from the fact that the event $\{P_n = t\}$ is adapted to the filtration $\mathscr{F}_{t-1}$, and the last follows from the fact that the sequence $\{X_{jt}\}_{t=1}^\infty$ is identically distributed.

Next, we show that X_{jP_m} and X_{jP_n} are independent for $n > m$. Observe that, for any measurable sets A_1 and A_2,

$$\begin{aligned}&\mathsf{P}\left(X_{jP_m} \in A_1, X_{jP_n} \in A_2\right) \\ &= \sum_{t_1<t_2} \mathsf{P}\left(X_{jt_1} \in A_1, X_{jt_2} \in A_2\right) \mathsf{P}\left(P_m = t_1, P_n = t_2\right) \\ &= \sum_{t_1<t_2} \mathsf{E}\left(\mathbf{I}\{X_{jt_1} \in A_1\}\mathbf{I}\{X_{jt_2} \in A_2\}\right) \mathsf{P}\left(P_m = t_1, P_n = t_2\right) \\ &= \sum_{t_1<t_2} \mathsf{E}\left(\mathsf{E}\left(\mathbf{I}\{X_{jt_1} \in A_1\}\mathbf{I}\{X_{jt_2} \in A_2\}\right) | \mathcal{F}_{t_2-1}\right) \mathsf{P}\left(P_m = t_1, P_n = t_2\right) \\ &\overset{(a)}{=} \sum_{t_1<t_2} \mathsf{E}\left(\mathbf{I}\{X_{jt_1} \in A_1\}\right) \mathsf{E}\left(\mathbf{I}\{X_{jt_2} \in A_2\}\right) \mathsf{P}\left(P_m = t_1, P_n = t_2\right) \\ &= \sum_{t_1<t_2} \mathsf{P}\left(X_{jt_1} \in A_1\right) \mathsf{P}\left(X_{jt_2} \in A_2\right) \mathsf{P}\left(P_m = t_1, P_n = t_2\right) \\ &\overset{(b)}{=} \mathsf{P}\left(X_j \in A_1\right) \mathsf{P}\left(X_j \in A_2\right) \sum_{t_1<t_2} \mathsf{P}\left(P_m = t_1, P_n = t_2\right) \\ &= \mathsf{P}\left(X_j \in A_1\right) \mathsf{P}\left(X_j \in A_2\right) \\ &\overset{(c)}{=} \mathsf{P}\left(X_{jP_m} \in A_1\right) \mathsf{P}\left(X_{jP_n} \in A_2\right),\end{aligned} \tag{B.2}$$

where (a) follows from the fact that $X_{jt_1} \in \mathcal{F}_{t_2-1}$ for $t_1 < t_2$ and (b) and (c) follow from (B.1). Therefore, $\{X_{jP_n}\}_{n=1}^{\infty}$ is independent and identically distributed. The sample pth quantile from an independent and identically distributed sequence converges to the pth quantile of the distribution almost surely (see, e.g., p. 75 of Serfling 2009), and hence, also in probability. This completes the proof. □

The proof for Proposition 2 requires the following lemma. Proofs for all auxiliary lemmas are provided in §B.2.

Lemma B.1 **(Large deviations for sample quantiles)**. *Under (F1)-(F2), Equation* (9) *holds for a static algorithm* $\pi(\boldsymbol{\alpha})$ *for any* $\boldsymbol{\alpha} \in \Delta^0$.

Proof of Proposition 2. In this proof, we fix a static algorithm $\pi = \pi(\boldsymbol{\alpha})$ for some $\boldsymbol{\alpha} \in \Delta^0$ and suppress π in the superscripts to improve clarity. Observe that

$$\max_{j=2,\dots,k} \mathsf{P}(\hat{\xi}_{1t} \le \hat{\xi}_{jt}) \le \mathsf{P}(\mathrm{FS}_t) \le (k-1) \max_{j=2,\dots,k} \mathsf{P}(\hat{\xi}_{1t} \le \hat{\xi}_{jt}). \tag{B.3}$$

Hence, if, for each $j = 2, \dots, k$,

$$\lim_{t\to\infty} \frac{1}{t} \log \mathsf{P}\left(\hat{\xi}_{1t} \le \hat{\xi}_{jt}\right) = -G_j(\boldsymbol{\alpha}), \tag{B.4}$$

then

$$\lim_{t\to\infty} \frac{1}{t} \log \mathsf{P}\left(\mathrm{FS}_t\right) = -\min_{j=2,\dots,k} \{G_j(\boldsymbol{\alpha})\}. \tag{B.5}$$

Therefore, the rate function of $\mathsf{P}(\mathrm{FS}_t)$ can be immediately obtained once we prove (B.4) for some $j \ne 1$.

First, observe that

$$\mathsf{P}\left(\hat{\xi}_{1t} \le \hat{\xi}_{2t}\right) \ge \mathsf{P}\left(\hat{\xi}_{1t} \le x\right) \mathsf{P}\left(\hat{\xi}_{2t} \ge x\right) \tag{B.6}$$

for any $x \in \mathcal{R}$. Also, taking log on both sides and combining the result from Lemma B.1, it can be easily seen that

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathsf{P}\left(\hat{\xi}_{1t} \le \hat{\xi}_{2t}\right) \ge -\inf_{x \in \mathscr{H}_1 \cap \mathscr{H}_2} \left\{\alpha_1 I_1(x) + \alpha_2 I_2(x)\right\}, \tag{B.7}$$

where the inf follows from the fact that (B.6) holds for any x. Also, in the continuous case, observe that $I_j(x)$ is non-increasing for $x < \xi_j$ and non-decreasing for $x > \xi_j$, and therefore, it suffices to search for the infimum for $\xi_2 \le x \le \xi_1$. Likewise, in the discrete case, observe that $I_j(x)$ is non-increasing for $x < \xi_j - 1$ and non-decreasing for $x > \xi_j$, so that the infimum is achieved in $[\xi_2 - 1, \xi_1]$. Therefore, (B.6) reduces to

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathsf{P}(\hat{\xi}_{1t} \le \hat{\xi}_{2t}) \ge \begin{cases} -\inf_{x\in[\xi_2,\xi_1]}\{\alpha_1 I_1(x) + \alpha_2 I_2(x)\} & \text{for } \mathbf{F} \in \mathscr{C} \\ -\inf_{x\in[\xi_2-1,\xi_1]}\{\alpha_1 I_1(x) + \alpha_2 I_2(x)\} & \text{for } \mathbf{F} \in \mathscr{D}. \end{cases} \tag{B.8}$$

It remains to show the upper bound:

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathsf{P}\left(\hat{\xi}_{1t} \le \hat{\xi}_{2t}\right) \le \begin{cases} -\inf_{x\in[\xi_2,\xi_1]}\left\{\alpha_1 I_1(x) + \alpha_2 I_2(x)\right\} & \text{for } \mathbf{F} \in \mathscr{C} \\ -\inf_{x\in[\xi_2-1,\xi_1]}\left\{\alpha_1 I_1(x) + \alpha_2 I_2(x)\right\} & \text{for } \mathbf{F} \in \mathscr{D}. \end{cases} \tag{B.9}$$

(i) Continuous case. Let $\boldsymbol{x} = (x_1, x_2) \in \mathcal{R}^2$ with $x_1 \le x_2$, fix $\eta > 0$, and consider a square centered at $\boldsymbol{x}$, $S_{\boldsymbol{x}}^{\eta} = \{\boldsymbol{x}' \in \mathcal{R}^2 \mid |x_1' - x_1| \le r_\eta, |x_2' - x_2| \le r_\eta\}$, with $r_\eta > 0$ chosen small enough so that

$$\max_{j=1,2} \left\{|I_j(x_j') - I_j(x_j)|\right\} \le \eta \quad \text{for any } \boldsymbol{x}' \in S_{\boldsymbol{x}}^{\eta}. \tag{B.10}$$

Note that r_η always exists for each $\eta > 0$ because $I_j(\cdot)$ is a continuous function. Observe that

$$\begin{aligned}\mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})\in S^\eta_{\boldsymbol{x}}\right) &= \mathsf{P}\left(\hat{\xi}_{1t}\in[x_1-r_\eta,x_1+r_\eta]\right)\mathsf{P}\left(\hat{\xi}_{2t}\in[x_2-r_\eta,x_2+r_\eta]\right)\\ &\le \mathsf{P}\left(\hat{\xi}_{1t}\le x_1+r_\eta\right)\mathsf{P}\left(\hat{\xi}_{2t}\ge x_2-r_\eta\right).\end{aligned} \tag{B.11}$$

Applying Lemma B.1, it follows that

$$\begin{aligned}\limsup_{t\to\infty}\frac{1}{t}\log\mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})\in S^\eta_{\boldsymbol{x}}\right) &\le -\left(\alpha_1 I_1(x_1+r_\eta)+\alpha_2 I_2(x_2-r_\eta)\right)\\ &\le \eta-\left(\alpha_1 I_1(x_1)+\alpha_2 I_2(x_2)\right),\end{aligned} \tag{B.12}$$

where the second inequality follows from (B.10).

Now, fix $a > 0$ and consider a compact set $\Gamma_a = \{\boldsymbol{x}\in\mathcal{R}^2 \mid |x_1-\xi_1|\le a, |x_2-\xi_2|\le a\}$. Then the set, $\Gamma_a\cap\{\boldsymbol{x}\in\mathcal{R}^2 \mid x_1\le x_2\}$, can be covered by $M<\infty$ squares, each of which is centered at $\boldsymbol{x}^i=(x_1^i,x_2^i)\in\Gamma_a$ with $x_1^i\le x_2^i$. Hence, it follows that

$$\begin{aligned}\mathsf{P}\left(\hat{\xi}_{1t}\le\hat{\xi}_{2t},(\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a\right) &\le \sum_{i=1}^{M}\mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})\in S^\eta_{\boldsymbol{x}^i}\right)\\ &\le M\max_{i=1,\ldots,M}\left\{\mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})\in S^\eta_{\boldsymbol{x}^i}\right)\right\}.\end{aligned} \tag{B.13}$$

Therefore,

$$\begin{aligned}&\limsup_{t\to\infty}\frac{1}{t}\log\mathsf{P}\left(\hat{\xi}_{1t}\le\hat{\xi}_{2t},(\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a\right)\\ &\le \max_{i=1,\ldots,M}\left\{\limsup_{t\to\infty}\frac{1}{t}\log\mathsf{P}\left(\hat{\xi}_{1t}\le\hat{\xi}_{2t},(\hat{\xi}_{1t},\hat{\xi}_{2t})\in S^\eta_{\boldsymbol{x}^i}\right)\right\}\\ &\le \max_{i=1,\ldots,M}\left\{\eta-(\alpha_1 I_1(x_1^i)+\alpha_2 I_2(x_2^i))\right\}\\ &\le \eta-\inf_{x_1\le x_2,\boldsymbol{x}\in\Gamma_a}\{\alpha_1 I_1(x_1)+\alpha_2 I_2(x_2)\},\end{aligned} \tag{B.14}$$

where the second inequality follows from (B.12). Since $I_j(x)$ is non-increasing for $x<\xi_j$ and non-decreasing for $x>\xi_j$, it suffices to search for the infimum with (x_1,x_2) that satisfies $\xi_2\le x_1=x_2\le\xi_1$. By the arbitrariness of $\eta>0$, we have that

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathsf{P}\left(\hat{\xi}_{1t}\le\hat{\xi}_{2t},(\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a\right)\le-\inf_{x\in[\xi_2,\xi_1]}\{\alpha_1 I_1(x)+\alpha_2 I_2(x)\}. \tag{B.15}$$

Finally, observe that

$$\mathsf{P}\left(\hat{\xi}_{1t}<\hat{\xi}_{2t}\right)\le\mathsf{P}\left(\hat{\xi}_{1t}<\hat{\xi}_{2t}\text{ and }(\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a\right)+\mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a^c\right), \tag{B.16}$$

and hence,

$$\begin{aligned}&\limsup_{t\to\infty}\frac{1}{t}\log\mathsf{P}\left(\hat{\xi}_{1t}<\hat{\xi}_{2t}\right)\\ &\le\max\left\{\limsup_{t\to\infty}\frac{1}{t}\log\mathsf{P}\left(\hat{\xi}_{1t}<\hat{\xi}_{2t}\text{ and }(\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a\right),\limsup_{t\to\infty}\frac{1}{t}\log\mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a^c\right)\right\}\\ &\le\max\left\{-\inf_{x\in[\xi_2,\xi_1]}\{\alpha_1 I_1(x)+\alpha_2 I_2(x)\},\limsup_{t\to\infty}\frac{1}{t}\log\mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a^c\right)\right\},\end{aligned} \tag{B.17}$$

where the last inequality follows from (B.15). One can show that the first term on the right-hand side of (B.17) is bounded below since $I_1(x)$ and $I_2(x)$ are bounded for x in the compact set $[\xi_2,\xi_1]$. Also, since

$$\begin{aligned}&\mathsf{P}((\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a^c)\\ &\le 4\max\left\{\mathsf{P}\left(\hat{\xi}_{1t}<\xi_1-a\right),\mathsf{P}\left(\hat{\xi}_{1t}>\xi_1+a\right),\mathsf{P}\left(\hat{\xi}_{2t}<\xi_2-a\right),\mathsf{P}\left(\hat{\xi}_{2t}<\xi_2+a\right)\right\},\end{aligned} \tag{B.18}$$

combined with Lemma B.1, it can be seen that

$$\begin{aligned}&\lim_{a\to\infty}\limsup_{t\to\infty}\frac{1}{t}\log \mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a^c\right)\\&\le -\lim_{a\to\infty}\min\{\alpha_1 I_1(\xi_1+a),\alpha_1 I_1(\xi_1-a),\alpha_2 I_2(\xi_2+a),\alpha_2 I_2(\xi_2-a)\}\\&=-\infty,\end{aligned}\tag{B.19}$$

where the equality follows from the fact that $I_j(x)\to\infty$ as $|x|\to\infty$ for each $j=1,2$. This implies that one can choose sufficiently large a so that the second term on the right-hand side of (B.17) is smaller than the first term. This completes the proof for the upper bound (B.9) in the continuous case.

(ii) Discrete case. Let $\boldsymbol{n}=(n_1,n_2)\in\mathcal{N}^2$ with $n_1\le n_2$, where $\mathcal{N}$ is the set of non-negative integers, and observe that

$$\mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})=\boldsymbol{n}\right)=\mathsf{P}\left(\hat{\xi}_{1t}=n_1\right)\mathsf{P}\left(\hat{\xi}_{2t}=n_2\right).\tag{B.20}$$

Applying Lemma B.1, it follows that

$$\limsup_{t\to\infty}\log \mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})=\boldsymbol{n}\right)=-(\alpha_1 I_1(n_1)+\alpha_2 I_2(n_2)).\tag{B.21}$$

Now, fix $a>0$ and consider a compact set $\Gamma_a=\{\boldsymbol{n}\in\mathcal{N}^2 \mid |n_1-\xi_1|\le a, |n_2-\xi_2|\le a\}$. Denote $\{\boldsymbol{n}^i\}_{i=1}^M$ be the $M<\infty$ points such that $\boldsymbol{n}^i\in\Gamma_a$ and $n_1^i\le n_2^i$. Observe that

$$\begin{aligned}\mathsf{P}\left(\hat{\xi}_{1t}\le\hat{\xi}_{2t},(\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a\right)&\le\sum_{i=1}^M \mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})=\boldsymbol{n}^i\right)\\&\le M\max_{i=1,\dots,M}\{\mathsf{P}\left((\hat{\xi}_{1t},\hat{\xi}_{2t})=\boldsymbol{n}^i\right)\},\end{aligned}\tag{B.22}$$

from which we establish that

$$\begin{aligned}&\limsup_{t\to\infty}\frac{1}{t}\log \mathsf{P}\left(\hat{\xi}_{1t}\le\hat{\xi}_{2t},(\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a\right)\\&\le\max_{i=1,\dots,M}\left\{\limsup_{t\to\infty}\frac{1}{t}\log \mathsf{P}\left(\hat{\xi}_{1t}\le\hat{\xi}_{2t},(\hat{\xi}_{1t},\hat{\xi}_{2t})=\boldsymbol{n}^i\right)\right\}\\&\le\max_{i=1,\dots,M}\left\{-(\alpha_1 I_1(n_1^i)+\alpha_2 I_2(n_2^i)\right\}\\&\le-\inf_{n_1\le n_2,\boldsymbol{n}\in\Gamma_a}\{\alpha_1 I_1(n_1)+\alpha_2 I_2(n_2)\},\end{aligned}\tag{B.23}$$

where the second inequality follows from (B.22). It is not hard to verify that $I_j(n)$ is non-increasing for $n<\xi_j-1$ and non-decreasing for $n>\xi_j$. Hence, it suffices to search for (n_1,n_2) such that $\xi_2-1\le n_1=n_2\le\xi_1$. Therefore, we have that

$$\limsup_{t\to\infty}\frac{1}{t}\log \mathsf{P}\left(\hat{\xi}_{1t}\le\hat{\xi}_{2t},(\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a\right)\le-\inf_{n\in[\xi_2-1,\xi_2]}\{\alpha_1 I_1(n)+\alpha_2 I_2(n)\}.\tag{B.24}$$

Finally, by the same argument as in the continuous case, the event $\{(\hat{\xi}_{1t},\hat{\xi}_{2t})\in\Gamma_a\}$ is negligible with a chosen sufficiently large. This gives us the desired result for the discrete case, which concludes the proof of the theorem. □

Proof of Theorem 1. First, it is straightforward to see that the naive algorithm is consistent since $N_{jt}\ge c\log t$ for sufficiently large t. Hence, it follows that $\hat{\rho}_t(\boldsymbol{\alpha})\to\rho(\boldsymbol{\alpha})$ as $t\to\infty$ for each $\boldsymbol{\alpha}$ because $\hat{I}_{jt}(x)$ converges to $I_j(x)$ pointwise for each j and each quantile estimator $\hat{\xi}_{jt}$ converges to ξ_j by Proposition 1. Let $\boldsymbol{\alpha}^*\in\arg\max_{\boldsymbol{\alpha}\in\Delta}\{\rho(\boldsymbol{\alpha})\}$ and observe that

$$|\hat{\rho}_t(\hat{\boldsymbol{\alpha}}_t)-\rho(\boldsymbol{\alpha}^*)|\le\max(|\hat{\rho}_t(\hat{\boldsymbol{\alpha}}_t)-\rho(\hat{\boldsymbol{\alpha}}_t)|,|\hat{\rho}_t(\boldsymbol{\alpha}^*)-\rho(\boldsymbol{\alpha}^*)|),\tag{B.25}$$

and therefore, $\hat{\rho}_t(\hat{\boldsymbol{\alpha}}_t) \to \rho(\boldsymbol{\alpha}^*)$ as $t \to \infty$. Combined with the fact that $\boldsymbol{\alpha}_t - \hat{\boldsymbol{\alpha}}_t \to 0$ as $t \to \infty$ almost surely by construction of Algorithm 2, it can be easily seen that $\mathscr{R}_t^\pi = \rho(\boldsymbol{\alpha}_t)/\rho(\boldsymbol{\alpha}^*) \to 1$ as $t \to \infty$ almost surely. Since $\mathscr{R}_t^\pi$ is a bounded random variable, we establish that

$$\mathsf{E}(\mathscr{R}_t^\pi) = \mathsf{E}\left(\frac{\rho(\boldsymbol{\alpha}_t)}{\rho(\boldsymbol{\alpha}^*)}\right) \to 1 \text{ as } t \to \infty. \quad \square \tag{B.26}$$

Proof of Proposition 3. As a slight abuse of notation, we consider a sequence of system configurations where each configuration is uniquely parametrized by $\delta > 0$. We assume without loss of generality that $\xi_1 > \xi_2 = \max_{j>2}\{\xi_j\}$ for every system configuration. We prove the proposition in four steps.

Step 1. We first show that $|G_j(\boldsymbol{\alpha}) - G_j^\delta(\boldsymbol{\alpha})| = o(\delta^2)$ if $\xi_1 - \xi_j \to 0$ as $\delta \to 0$. Fix $j = 2, \ldots, k$ such that $\xi_1 - \xi_j \to 0$ as $\delta \to 0$ and observe that

$$G_j(\boldsymbol{\alpha}) = \inf_{x \in [\xi_j, \xi_1]} \{\alpha_1 I_1(x) + \alpha_j I_j(x)\} \tag{B.27}$$

and observe that

$$\begin{aligned} I_j'(x) &= f_j(x)\left(\frac{p}{F_j(x)} - \frac{1-p}{1-F_j(x)}\right) \\ I_j''(x) &= F_j''(x)\left(\frac{p}{F_j(x)} - \frac{1-p}{1-F_j(x)}\right) + f_j^2(x)\left(-\frac{p}{F_j^2(x)} + \frac{1-p}{(1-F_j(x))^2}\right), \end{aligned} \tag{B.28}$$

where $I_j'(x)$ and $I_j''(x)$ are the first and the second derivatives of $I_j(\cdot)$ at x, respectively. Hence, it can be easily seen that $I_j(\xi_j) = 0$, $I_j'(\xi_j) = 0$, and

$$I_j''(\xi_j) = \frac{f_j^2(\xi_j)}{p(1-p)}. \tag{B.29}$$

Therefore, applying a second-order Taylor expansion at $x = \xi_j$, we obtain that

$$I_j(x) = \frac{(x-\xi_j)^2}{2} I_j''(\tilde{x}) \tag{B.30}$$

for some $\tilde{x}$ on the line segment between x and ξ_j. By (F3) we know that $I_j''(x)$ is continuous. Since $f_j^2(\xi_j) \le f_{\max}$ and $(x-\xi_j)^2 \le (\xi_1 - \xi_j)^2$, we obtain that

$$I_j(x) = \frac{(x-\xi_j)^2}{2p(1-p)} f_j^2(\xi_j) + o((\xi_1-\xi_j)^2) \text{ as } \xi_1 - \xi_j \to 0. \tag{B.31}$$

for $x \in [\xi_j, \xi_1]$. Note that the observations from (B.28)-(B.31) also hold for $j = 1$. Define $I_{1j}(x) = \alpha_1 I_1(x) + \alpha_j I_j(x)$ and let

$$\begin{aligned} I_{1j}^\delta(x) &= \alpha_1 I_1^\delta(x) + \alpha_2 I_2^\delta(x) \\ &= \alpha_1 \frac{(x-\xi_1)^2}{2p(1-p)} f_1^2(\xi_1) + \alpha_2 \frac{(x-\xi_2)^2}{2p(1-p)} f_2^2(\xi_2). \end{aligned} \tag{B.32}$$

Now, let x^* and $x^{\delta*}$ be the minimizers of $I_{1j}(x)$ and $I_{1j}^\delta(x)$, respectively. Note that $I_{1j}^\delta(x)$ is a quadratic function of x. Using first order conditions, we establish that

$$\begin{aligned} x^{\delta*} &= \frac{\alpha_1 \xi_1 f_1^2(\xi_1) + \alpha_j \xi_j f_j^2(\xi_j)}{\alpha_1 f_1^2(\xi_1) + \alpha_j f_j^2(\xi_j)}, \\ I_{1j}^\delta(x^{\delta*}) &= \frac{(\xi_1-\xi_j)^2}{2p(1-p)\left(1/(\alpha_1 f_1^2(\xi_1)) + 1/(\alpha_j f_j^2(\xi_j))\right)}. \end{aligned} \tag{B.33}$$

From the definitions of $I_{1j}(x)$ and $I_{1j}^\delta(x)$ and using (B.31), it can be seen that $|I_{1j}(x) - I_{1j}^\delta(x)| = o((\xi_1-\xi_j)^2)$ for any $x \in [\xi_j, \xi_1]$. Also, since $I_j(x)$ is non-increasing for $x < \xi_j$ and non-decreasing for $x > \xi_j$ for any j, it can be easily seen that $x^* \in [\xi_j, \xi_1]$. Likewise, $I_j^\delta(x)$ is decreasing for $x < \xi_j$ and increasing for $x > \xi_j$ for any j, and hence, $x^{\delta*} \in [\xi_j, \xi_1]$. Now, observe that $|I_{1j}(x^*) - I_{1j}^\delta(x^{\delta*})| \le \max(|I(x^*) - I^\delta(x^*)|, |I^\delta(x^{\delta*}) - I(x^{\delta*})|)$, and each term of max is $o((\xi_1 - \xi_j)^2)$. Therefore, we obtain that $I_{1j}(x^*) = I_{1j}^\delta(x^{\delta*}) + o((\xi_1 - \xi_j)^2)$, and therefore, $|G_j(\boldsymbol{\alpha}) - G_j^\delta(\boldsymbol{\alpha})| = o(\delta^2)$ as $\delta \to 0$.

Step 2. Next, we show that if $\arg\min_{j\neq 1}\{G_j(\boldsymbol{\alpha})\} \to i$ as $\delta \to 0$ for some $i \neq 1$, then $\xi_1 - \xi_i \to 0$ as $\delta \to 0$. Observe that $G_i(\boldsymbol{\alpha}) \le G_2(\boldsymbol{\alpha})$ for sufficiently small δ since $\rho(\boldsymbol{\alpha}) = G_i(\boldsymbol{\alpha})$ for $\delta \in (0, \delta_0)$. From the definition of $G_j(\cdot)$ in (11) and the fact that $I_j(x)$ is strictly concave with $I_j(\mu_j) = 0$, it can be easily seen that $G_2(\boldsymbol{\alpha}) \to 0$ as $\delta \to 0$, and therefore, we establish that $G_i(\boldsymbol{\alpha}) \to 0$ as $\delta \to 0$. Now, towards a contradiction, suppose that $\liminf_{\delta\to 0}(\xi_1 - \xi_i) \ge d$ for some constant $d > 0$. Recall that $G_i(\boldsymbol{\alpha}) = \inf_x\{\alpha_1 I_1(x) + \alpha_i I_i(x)\}$ and let x_i^* be the minimizer which lies in $[\mu_i, \mu_1]$. Note that $G_i(\boldsymbol{\alpha}) \to 0$ implies both $I_1(x_i^*)$ and $I_i(x_i^*)$ converge to zero. However, due to (10) and the assumption (F3), it can be seen that $I_1(x_i^*)$ and $I_i(x_i^*)$ can converge to zero only when $x_i^* - \xi_1 \to 0$ and $x_i^* - \xi_i \to 0$ as $\delta \to 0$, respectively. This is a contradiction due to the assumption that $\liminf_{\delta\to 0}(\mu_1 - \mu_i) \ge d$. Therefore, we have that $\xi_1 - \xi_i \to 0$ as $\delta \to 0$. Exactly the same arguments can be applied to show that if $\arg\min_{j\neq 1}\{G_j^\delta(\boldsymbol{\alpha})\} \to i$ as $\delta \to 0$ for some $i \neq 1$, then it must be that $\xi_1 - \xi_i \to 0$ as $\delta \to 0$, which will be omitted.

Step 3. Next, we show that $|\rho(\boldsymbol{\alpha}) - \rho^\delta(\boldsymbol{\alpha})| = o(\delta^2)$ as $\delta \to 0$. We consider three cases.

Case (a). Suppose that the limits of $\arg\min_{j\neq 1}\{G_j(\boldsymbol{\alpha})\}$ and $\arg\min_{j\neq 1}\{G_j^\delta(\boldsymbol{\alpha})\}$ exist as $\delta \to 0$

$$\lim_{\delta\to 0}\arg\min_{j\neq 1}\{G_j(\boldsymbol{\alpha})\} = \lim_{\delta\to 0}\arg\min_{j\neq 1}\{G_j^\delta(\boldsymbol{\alpha})\}. \tag{B.34}$$

In this case, we have that $|\rho(\boldsymbol{\alpha}) - \rho^\delta(\boldsymbol{\alpha})| = |G_j(\boldsymbol{\alpha}) - G_j^\delta(\boldsymbol{\alpha})|$ for some $j \neq 1$ and sufficiently small δ. Hence we immediately establish from Step 1 that $|\rho(\boldsymbol{\alpha}) - \rho^\delta(\boldsymbol{\alpha})| = o(\delta^2)$ as $\delta \to 0$.

Case (b). Suppose that the limits of $\arg\min_{j\neq 1}\{G_j(\boldsymbol{\alpha})\}$ and $\arg\min_{j\neq 1}\{G_j^\delta(\boldsymbol{\alpha})\}$ exist as $\delta \to 0$ and

$$\lim_{\delta\to 0}\arg\min_{j\neq 1}\{G_j(\boldsymbol{\alpha})\} \neq \lim_{\delta\to 0}\arg\min_{j\neq 1}\{G_j^\delta(\boldsymbol{\alpha})\}. \tag{B.35}$$

Without loss of generality, let $\arg\min_{j\neq 1}\{G_j(\boldsymbol{\alpha})\} \to 2$ and $\arg\min_{j\neq 1}\{G_j^\delta(\boldsymbol{\alpha})\} \to 3$ as $\delta \to 0$. By *Step 2*, we know that $|\xi_1 - \xi_3| \to 0$ as $\delta \to 0$. Now, consider a sufficiently small δ and observe that if $G_2(\boldsymbol{\alpha}) > G_3^\delta(\boldsymbol{\alpha})$, then $|G_2(\boldsymbol{\alpha}) - G_3^\delta(\boldsymbol{\alpha})| \le |G_3(\boldsymbol{\alpha}) - G_3^\delta(\boldsymbol{\alpha})|$ since $G_3(\boldsymbol{\alpha}) \ge G_2(\boldsymbol{\alpha})$ for sufficiently small δ. Similarly, if $G_2(\boldsymbol{\alpha}) < G_3^\delta(\boldsymbol{\alpha})$, then $|G_2(\boldsymbol{\alpha}) - G_3^\delta(\boldsymbol{\alpha})| \le |G_2(\boldsymbol{\alpha}) - G_2^\delta(\boldsymbol{\alpha})|$ since $G_3^\delta(\boldsymbol{\alpha}) \le G_2^\delta(\boldsymbol{\alpha})$. Combining these observations, it can be seen that

$$\begin{aligned}|\rho(\boldsymbol{\alpha}) - \rho^\delta(\boldsymbol{\alpha})| &= |G_2(\boldsymbol{\alpha}) - G_3^\delta(\boldsymbol{\alpha})| \\ &\le \max(|G_2(\boldsymbol{\alpha}) - G_2^\delta(\boldsymbol{\alpha})|, |G_3(\boldsymbol{\alpha}) - G_3^\delta(\boldsymbol{\alpha})|) \\ &= o(\delta^2)\end{aligned} \tag{B.36}$$

as $\delta \to 0$, where the last equality follows from Step 1.

Case (c). Lastly, suppose that the limit of $\arg\min_{j\neq 1}\{G_j(\boldsymbol{\alpha})\}$ or $\arg\min_{j\neq 1}\{G_j^\delta(\boldsymbol{\alpha})\}$ does not exist as $\delta \to 0$. In this case, fix $i \neq 1$ and $i' \neq 1$ and consider two sets of intervals such that

$$\begin{aligned}S_i &= \{\delta \ge 0 \mid \arg\min_{j\neq 1}\{G_j(\boldsymbol{\alpha})\} = i\} \\ S_{i'} &= \{\delta \ge 0 \mid \arg\min_{j\neq 1}\{G_j^\delta(\boldsymbol{\alpha})\} = i'\}.\end{aligned} \tag{B.37}$$

Let $R_{ii'} = S_i \cup S_{i'}$ and observe that

$$\lim_{\substack{\delta \to 0 \\ \delta \in R_{ii'}}} \arg\min_{j \neq 1} \{G_j(\boldsymbol{\alpha})\} = i$$
$$\lim_{\substack{\delta \to 0 \\ \delta \in R_{ii'}}} \arg\min_{j \neq 1} \{G_j^\delta(\boldsymbol{\alpha})\} = i'. \tag{B.38}$$

Hence, we can apply the analysis of Case (a) (respectively, Case (b)) if $i = i'$ (respectively, if $i \neq i'$) to establish that $|\rho(\boldsymbol{\alpha}) - \rho^\delta(\boldsymbol{\alpha})| = o(\delta^2)$ as $\delta \to 0$. These arguments hold for arbitrary i and i', which completes the proof of this step.

Step 4. Finally, we show that, for each $\boldsymbol{\alpha} \in \Delta^0$, $\rho(\boldsymbol{\alpha}) \geq c\delta^2$ for some positive constant c. From the observation in *Step 3*, it can be easily seen that

$$|\rho(\boldsymbol{\alpha}^*) - \rho(\boldsymbol{\alpha}^\delta)| \leq \max(|\rho(\boldsymbol{\alpha}^*) - \rho^\delta(\boldsymbol{\alpha}^*)|, |\rho^\delta(\boldsymbol{\alpha}^\delta) - \rho(\boldsymbol{\alpha}^\delta)|) = o(\delta^2) \text{ as } \delta \to 0. \tag{B.39}$$

Further, from the assumption that $f_j(\xi_j) \geq f_{\min}$, we have that, for $\boldsymbol{\alpha}_{\text{eq}} = (1/k, \ldots, 1/k)$,

$$\begin{aligned} \rho^\delta(\boldsymbol{\alpha}_{\text{eq}}) &\geq \min_{j \neq 1} \frac{(\xi_1 - \xi_j)^2}{2p(1-p)\,(2k/f_{\min})} \\ &\geq \min_{j \neq 1} \frac{\delta^2}{2p(1-p)\,(2k/f_{\min})} \\ &\geq c\delta^2 \end{aligned} \tag{B.40}$$

for some positive constant c. Hence, from $\rho^\delta(\boldsymbol{\alpha}^\delta) \geq \rho^\delta(\boldsymbol{\alpha}_{\text{eq}})$ we establish that $\rho^\delta(\boldsymbol{\alpha}^\delta) \geq c\delta^2$. By Proposition 3 and (B.39), $\rho^\delta(\boldsymbol{\alpha}^\delta) \geq c\delta^2$ implies $\rho^* \geq c\delta^2 - o(\delta^2)$ as $\delta \to 0$, and therefore, we establish that

$$\frac{\rho(\boldsymbol{\alpha}^\delta)}{\rho^*} = 1 + \frac{o(\delta^2)}{\rho^*} \to 1 \text{ as } \delta \to 0, \tag{B.41}$$

which completes the proof. □

To prove Theorem 2, we define additional notation. Let $\boldsymbol{x} = (x_1, \ldots, x_k) \in \mathcal{R}^k$ and $\boldsymbol{y} = (y_1, \ldots, y_k) \in \mathcal{R}_+^k$. Define $B(\boldsymbol{x}) := \{i \in \{1, \ldots, k\} \mid x_i = \max_j \{x_j\}\}$. Note that we allow the case with $|B(\boldsymbol{x})| > 1$, where $|B(\boldsymbol{x})|$ represents the cardinality of the set $B(\boldsymbol{x})$. Let Θ be the set of all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{R}^k \times \mathcal{R}_+^k$ with $|B(\boldsymbol{x})| < k$. For each $(\boldsymbol{x}, \boldsymbol{y}) \in \Theta$ and $\boldsymbol{\alpha} \in \Delta$, define

$$J(\boldsymbol{\alpha}; \boldsymbol{x}, \boldsymbol{y}) = \min_{i \in B(\boldsymbol{x}), j \notin B(\boldsymbol{x})} \frac{(x_i - x_j)^2}{\left(1/(\alpha_i y_i^2) + 1/(\alpha_j y_j^2)\right)}. \tag{B.42}$$

Define $\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y})$ to be the maximizer of $J(\boldsymbol{\alpha}; \boldsymbol{x}, \boldsymbol{y})$, i.e.,

$$\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y}) = \arg\max_{\boldsymbol{\alpha} \in \Delta} J(\boldsymbol{\alpha}; \boldsymbol{x}, \boldsymbol{y}). \tag{B.43}$$

Note that $J(\boldsymbol{\alpha}; \boldsymbol{x}, \boldsymbol{y})$ is a strictly concave function of $\boldsymbol{\alpha}$ since it is the minimum of the strictly concave functions. Hence, $\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y})$ in (B.43) is well defined. The following two lemmas are required for the proof of Theorem 2.

Lemma B.2. *Fix $(\boldsymbol{x}, \boldsymbol{y}) \in \Theta$. Consider a sequence of parameters $(\boldsymbol{x}^t, \boldsymbol{y}^t) \in \Theta$ such that $x_j^t \to x_j$ and $y_j^t \to y_j$ as $t \to \infty$ for each j. Then, $\boldsymbol{\alpha}(\boldsymbol{x}^t, \boldsymbol{y}^t) \to \boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y})$ as $t \to \infty$ and the vector $\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y})$ is strictly positive.*

Lemma B.3. *Suppose the kernel function and the bandwidth parameter satisfy (K1)-(K3). Then, if $N_{jt} \to \infty$ in probability as $t \to \infty$, $\hat{f}_{jt}(\hat{\xi}_{jt}) \to f_j(\xi_j)$ in probability as $t \to \infty$. Suppose the kernel function and the bandwidth parameter further satisfy (K4)-(K7) and that $f_j(\cdot)$ is uniformly continuous. Then, if $N_{jt} \to \infty$ almost surely as $t \to \infty$, $\hat{f}_{jt}(\hat{\xi}_{jt}) \to f_j(\xi_j)$ almost surely as $t \to \infty$.*

Proof of Theorem 2. (i) *Consistency.* It can be easily seen that $\boldsymbol{\alpha}^\delta = \arg\max_{\alpha\in\Delta}\{\rho^\delta(\boldsymbol{\alpha})\}$ is in the interior of Δ; otherwise, $\rho^\delta(\alpha^\delta)=0$ which contradicts the definition of the maximizer $\boldsymbol{\alpha}^\delta$ since $\rho^\delta(\alpha^{\text{eq}})>0$ for $\alpha^{\text{eq}} = (1/k,\dots,1/k)$. Using Proposition A.3 of Peng et al. (2016), we conclude that $\boldsymbol{\alpha}_t \to \boldsymbol{\alpha}^\delta \in \Delta^0$ in probability as $t\to\infty$, and hence the algorithm is consistent.

(ii) *Relative efficiency.* Given the consistency of the algorithm, it follows that $\hat{\xi}_{jt}\to\xi_j$ and $\hat{f}_{jt}\to f_j(\xi_j)$ as $t\to\infty$ in probability for $j=1,\dots,k$ by Proposition 1 and Lemma B.3. Applying Lemma B.2 with $(\boldsymbol{x}^t,\boldsymbol{y}^t)$ replaced with $(\hat{\boldsymbol{\xi}}_t,\hat{\boldsymbol{f}}_t)$ and $(\boldsymbol{x},\boldsymbol{y})$ replaced with $(\boldsymbol{\xi},\boldsymbol{f})$, where $\boldsymbol{f}=(f_1(\xi_1),\dots,f_k(\xi_k))$, it follows that $\hat{\boldsymbol{\alpha}}_t\to\boldsymbol{\alpha}^\delta$ as $t\to\infty$. Moreover, by construction of Algorithm 3 it is not difficult to see that the term $\boldsymbol{\alpha}_t-\hat{\boldsymbol{\alpha}}_t^\delta\to 0$. Consequently, we have that $\boldsymbol{\alpha}_t\to\boldsymbol{\alpha}^\delta$ as $t\to\infty$ and that $\rho(\boldsymbol{\alpha}_t)/\rho(\boldsymbol{\alpha}^\delta)\to 1$ in probability as $t\to\infty$. Since $\rho(\boldsymbol{\alpha}_T)/\rho(\boldsymbol{\alpha}^\delta)$ is bounded, we also have that

$$\mathsf{E}\left(\frac{\rho(\boldsymbol{\alpha}_t)}{\rho(\boldsymbol{\alpha}^\delta)}\right)\to 1 \text{ as } t\to\infty, \tag{B.44}$$

from which (25) follows. From Proposition 3 we have that $\rho(\boldsymbol{\alpha}^\delta)/\rho(\boldsymbol{\alpha}^*)\to 1$ as $\delta\to 0$, which completes the proof of this theorem. □

To show Proposition 4 we need the following lemma.

Lemma B.4. *Consider $\boldsymbol{x}=(x_1,\dots,x_k)\in\mathbf{R}^k$ and $\boldsymbol{y}=(y_1,\dots,y_k)\in\mathbf{R}^k$ with $k\ge 1$. Suppose that there exists a constant $c>0$ such that $1-c\le x_j/y_j\le 1+c$ for each $j=1,\dots,k$, then*

$$1-c\le\frac{\min_{j=1,\dots,k}x_j}{\min_{j=1,\dots,k}y_j}\le 1+c. \tag{B.45}$$

Proof of Proposition 4. For ease of notation we define $G_j(\boldsymbol{\alpha};\boldsymbol{h})=G_j^{\delta,\epsilon}(\boldsymbol{\alpha})$ for each $j\neq 1$ and let $\rho(\boldsymbol{\alpha};\boldsymbol{h})=\rho^{\delta,\epsilon}(\boldsymbol{\alpha})$. Further, we define

$$\begin{aligned} I_{1j}(x;\boldsymbol{h}) &= p\log\left(\frac{p}{p+(x-\xi_1)h_{1j}}\right)+(1-p)\log\left(\frac{1-p}{1-p-(x-\xi_1)h_{1j}}\right)\\ I_{j1}(x;\boldsymbol{h}) &= p\log\left(\frac{p}{p+(x-\xi_j)h_{j1}}\right)+(1-p)\log\left(\frac{1-p}{1-p-(x-\xi_j)h_{j1}}\right). \end{aligned} \tag{B.46}$$

It is useful to observe that

$$\frac{\rho(\boldsymbol{\alpha})}{\rho^{\delta,\epsilon}(\boldsymbol{\alpha})}=\frac{\rho(\boldsymbol{\alpha};\boldsymbol{h})}{\rho^{\delta,\epsilon}(\boldsymbol{\alpha})}\frac{\rho(\boldsymbol{\alpha})}{\rho(\boldsymbol{\alpha};\boldsymbol{h})}=\frac{\rho(\boldsymbol{\alpha};\boldsymbol{h})}{\rho^{\delta}(\boldsymbol{\alpha};\boldsymbol{h})}\frac{\rho(\boldsymbol{\alpha})}{\rho(\boldsymbol{\alpha};\boldsymbol{h})}. \tag{B.47}$$

Further, it can be seen that if $\arg\min_{j\neq 1}\{G_j(\boldsymbol{\alpha})\}\to i$ as $\delta\to 0$ for some $i\neq 1$, then $\xi_1-\xi_i\to 0$ as $\delta\to 0$. The proof of the preceding statement follows exactly the same arguments given in *Step 2* of the proof for Proposition 3, and hence skipped.

Now, we prove the proposition in three steps. In *Step 1*, we find the lower and upper bounds of the second term on the right-hand side of (B.47). In *Step 2*, we find the lower and upper bounds of the first term on the right-hand side of (B.47). In *Step 3*, we combine these observations to conclude the proof.

Step 1. We first show that $1-\epsilon\le\rho(\boldsymbol{\alpha})/\rho(\boldsymbol{\alpha};\boldsymbol{h})\le 1+\epsilon$. To this end, it suffices to show that $1-\epsilon\le G_j(\boldsymbol{\alpha})/G_j(\boldsymbol{\alpha};\boldsymbol{h})\le 1+\epsilon$ for each j, since the desired result follows immediately from Lemma B.4. Using the relationship $\log(1\pm z)\le\pm z$ for $0<z<1$ and assumption (F3), it can be seen that

$$1-\epsilon\le\frac{I_{1j}(x)}{I_{1j}(x;\boldsymbol{h})},\frac{I_{j1}(x)}{I_{j1}(x;\boldsymbol{h})}\le 1+\epsilon \tag{B.48}$$

for each $j \neq 1$ and $x \in [\xi_j, \xi_1]$. Now, let x^* and x^h be the minimizers of $\alpha_1 I_1(x) + \alpha_j I_j(x)$ and $\alpha_1 I_{1j}(x;\boldsymbol{h}) + \alpha_j I_{j1}(x;\boldsymbol{h})$, respectively, so that $G_j(\boldsymbol{\alpha}) = \alpha_1 I_1(x^*) + \alpha_j I_j(x^*)$ and $G_j(\alpha;\boldsymbol{h}) = \alpha_1 I_{1j}(x^h;\boldsymbol{h}) + \alpha_j I_{j1}(x^h;\boldsymbol{h})$. We establish that

$$\frac{G_j(\boldsymbol{\alpha})}{G_j(\boldsymbol{\alpha};\boldsymbol{h})} \leq \frac{\alpha_1 I_1(x^h) + \alpha_j I_j(x^h)}{\alpha_1 I_{1j}(x^h;\boldsymbol{h}) + \alpha_j I_{j1}(x^h;\boldsymbol{h})} \leq (1+\epsilon), \tag{B.49}$$

where the first inequality follows by the definition of x^h and the second follows from (B.48). Likewise, we can establish the lower bound as

$$\frac{G_j(\boldsymbol{\alpha})}{G_j(\boldsymbol{\alpha};\boldsymbol{h})} \geq \frac{\alpha_1 I_1(x^*) + \alpha_j I_j(x^*)}{\alpha_1 I_{1j}(x^*;\boldsymbol{h}) + \alpha_j I_{j1}(x^*;\boldsymbol{h})} \geq (1-\epsilon). \tag{B.50}$$

This completes the proof of *Step 1*.

Step 2. It remains to bound the first term of (B.47); in particular, $1 - u(\theta) \leq \rho(\boldsymbol{\alpha};\boldsymbol{h})/\rho^\delta(\boldsymbol{\alpha};\boldsymbol{h}) \leq 1 + u(\theta)$. Using Lemma B.4 it suffices to show that $1 - u(\theta) \leq G_j(\boldsymbol{\alpha};\boldsymbol{h})/G_j^\delta(\boldsymbol{\alpha};\boldsymbol{h}) \leq 1 + u(\theta)$ for each j. To this end, recall the definitions of $\rho(\boldsymbol{\alpha};\boldsymbol{h}) = \min_{j\neq 1}\{G_j(\boldsymbol{\alpha};\boldsymbol{h})\}$ and $\rho^\delta(\boldsymbol{\alpha};\boldsymbol{h}) = \min_{j\neq 1}\{G_j^\delta(\boldsymbol{\alpha};\boldsymbol{h})\}$, where

$$\begin{aligned} G_j(\boldsymbol{\alpha};\boldsymbol{h}) &= \inf_x \{\alpha_1 I_{1j}(x;\boldsymbol{h}) + \alpha_j I_{j1}(x;\boldsymbol{h})\} \\ G_j^\delta(\boldsymbol{\alpha};\boldsymbol{h}) &= \inf_x \{\alpha_1 I_{1j}^\delta(x;\boldsymbol{h}) + \alpha_j I_{j1}(x;\boldsymbol{h})\}, \end{aligned} \tag{B.51}$$

and

$$\begin{aligned} I_{1j}^\delta(x;\boldsymbol{h}) &= \frac{(x-\xi_j)^2}{2p(1-p)/h_{1j}^2} \\ I_{j1}^\delta(x;\boldsymbol{h}) &= \frac{(x-\xi_j)^2}{2p(1-p)/h_{j1}^2}. \end{aligned} \tag{B.52}$$

Applying a second-order Taylor expansion for $I_{j1}(x;\boldsymbol{h})$ at $x = \xi_j$, we obtain that

$$\begin{aligned} I_{j1}(x;\boldsymbol{h}) &= I_{j1}(\xi_j;\boldsymbol{h}) + (x-\xi_j) I'_{j1}(\xi_j;\boldsymbol{h}) + \frac{(x-\xi_j)^2}{2} I''_{j1}(\xi_j;\boldsymbol{h}) + \frac{(x-\xi_j)^3}{6} I'''_{j1}(\tilde{x};\boldsymbol{h}) \\ &= \frac{(x-\xi_j)^2}{2p(1-p)} h_{j1}^2 + \frac{(x-\xi_j)^3}{6} I'''_{j1}(\tilde{x};\boldsymbol{h}) \\ &= I_{j1}^\delta(x;\boldsymbol{h}) + \frac{(x-\xi_j)^3}{6} I'''_{j1}(\tilde{x};\boldsymbol{h}) \end{aligned} \tag{B.53}$$

for $x \in [\xi_j, \xi_1]$, where $\tilde{x}$ lies between x and ξ_j and

$$I'''_{j1}(x;\boldsymbol{h}) = 2h_{j1}^3 \left(\frac{-p}{(p + (x-\xi_j)h_{j1})^3} + \frac{1-p}{(1-p-(x-\xi_j)h_{j1})^3} \right). \tag{B.54}$$

Note that the first equality of (B.53) follows from the fact that $I_{j1}(\xi_j;\boldsymbol{h}) = 0$ and $I'_{j1}(\xi_j;\boldsymbol{h}) = 0$, and the last equality follows from the definition of $I_{j1}^\delta(\cdot)$ in (B.52). From (B.53), we have that

$$\frac{I_{j1}(x;\boldsymbol{h})}{I_{j1}^\delta(x;\boldsymbol{h})} \leq 1 + \left| \frac{(x-\xi_j)^3 I'''_{j1}(\tilde{x};\boldsymbol{h})/6}{(x-\xi_j)^2 h_{j1}^2/(2p(1-p))} \right| \leq 1 + u(\theta), \tag{B.55}$$

which follows from the definition of $u(\theta)$ and straightforward algebra. Likewise, we can establish the lower bound as

$$\frac{I_{j1}(x;\boldsymbol{h})}{I_{j1}^\delta(x;\boldsymbol{h})} \geq 1 - u(\theta). \tag{B.56}$$

To summarize our observations in Step 4 so far, we have that

$$1 - u(\theta) \leq \frac{I_{j1}(x;\boldsymbol{h})}{I_{j1}^\delta(x;\boldsymbol{h})} \leq 1 + u(\theta). \tag{B.57}$$

Further, exactly the same steps from (B.52) to (B.56) can be applied to show that

$$1-u(\theta) \le \frac{I_{1j}(x;\boldsymbol{h})}{I^{\delta}_{1j}(x;\boldsymbol{h})} \le 1+u(\theta) \tag{B.58}$$

(the detailed derivation is omitted).

Next, define x^h and $x^{h\delta}$ as the minimizers of $\alpha_1 I_{1j}(x;\boldsymbol{h})+\alpha_j I_{j1}(x;\boldsymbol{h})$ and $\alpha_1 I^{\delta}_{1j}(x;\boldsymbol{h})+\alpha_j I^{\delta}_{j1}(x;\boldsymbol{h})$, respectively. Then we can write

$$\frac{G_j(\boldsymbol{\alpha};\boldsymbol{h})}{G^{\delta}_j(\boldsymbol{\alpha};\boldsymbol{h})} \le \frac{\alpha_1 I_{1j}(x^{h\delta};\boldsymbol{h})+\alpha_j I_{j1}(x^{h\delta};\boldsymbol{h})}{\alpha_1 I^{\delta}_{1j}(x^{h\delta};\boldsymbol{h})+\alpha_j I^{\delta}_{j1}(x^{h\delta};\boldsymbol{h})} \le 1+u(\theta), \tag{B.59}$$

where the first inequality follows by the definition of $x^{h\delta}$ and the second follows from (B.57)-(B.58). Likewise, we can write

$$\frac{G_j(\boldsymbol{\alpha};\boldsymbol{h})}{G^{\delta}_j(\boldsymbol{\alpha};\boldsymbol{h})} \ge \frac{\alpha_1 I_{1j}(x^{h};\boldsymbol{h})+\alpha_j I_{j1}(x^{h};\boldsymbol{h})}{\alpha_1 I^{\delta}_{1j}(x^{h};\boldsymbol{h})+\alpha_j I^{\delta}_{j1}(x^{h};\boldsymbol{h})} \le 1-u(\theta). \tag{B.60}$$

The proof of this step is complete from the last two inequalities.

Step 3. Finally, observe that

$$\begin{aligned}\frac{\rho(\boldsymbol{\alpha}^{\delta,\epsilon*})}{\rho(\boldsymbol{\alpha}^{*})} &= \frac{\rho(\boldsymbol{\alpha}^{\delta,\epsilon*})}{\rho(\boldsymbol{\alpha}^{*};\boldsymbol{h})}\frac{\rho(\boldsymbol{\alpha}^{*};\boldsymbol{h})}{\rho(\boldsymbol{\alpha}^{*})}\\ &\ge \frac{\rho(\boldsymbol{\alpha}^{\delta,\epsilon*};\boldsymbol{h})}{\rho(\boldsymbol{\alpha}^{*};\boldsymbol{h})}\frac{1-\epsilon}{1+\epsilon}\\ &= \frac{\rho^{\delta}(\boldsymbol{\alpha}^{\delta,\epsilon*};\boldsymbol{h})}{\rho(\boldsymbol{\alpha}^{*};\boldsymbol{h})}\frac{\rho(\boldsymbol{\alpha}^{\delta,\epsilon*};\boldsymbol{h})}{\rho^{\delta}(\boldsymbol{\alpha}^{\delta,\epsilon*};\boldsymbol{h})}\frac{1-\epsilon}{1+\epsilon}\\ &\ge \frac{\rho^{\delta}(\boldsymbol{\alpha}^{*};\boldsymbol{h})}{\rho(\boldsymbol{\alpha}^{*};\boldsymbol{h})}\frac{\rho(\boldsymbol{\alpha}^{\delta,\epsilon*};\boldsymbol{h})}{\rho^{\delta}(\boldsymbol{\alpha}^{\delta,\epsilon*};\boldsymbol{h})}\frac{1-\epsilon}{1+\epsilon},\end{aligned} \tag{B.61}$$

where the first inequality follows from the fact that $1-\epsilon \le \rho(\boldsymbol{\alpha})/\rho(\boldsymbol{\alpha};\boldsymbol{h}) \le 1+\epsilon$ for any $\boldsymbol{\alpha}\in\Delta^0$ and the second inequality follows from the fact that $\boldsymbol{\alpha}^{\delta,\epsilon*}$ is the maximizer of $\rho^{\delta}(\cdot;\boldsymbol{h})=\rho^{\delta,\epsilon}(\cdot)$. Using the observations from *Step 2*, it can be seen that

$$\begin{aligned}\frac{\rho^{\delta}(\boldsymbol{\alpha}^{*};\boldsymbol{h})}{\rho(\boldsymbol{\alpha}^{*};\boldsymbol{h})} &\ge \frac{1}{1+u(\theta)}\\ \frac{\rho(\boldsymbol{\alpha}^{\delta,\epsilon*};\boldsymbol{h})}{\rho^{\delta}(\boldsymbol{\alpha}^{\delta,\epsilon*};\boldsymbol{h})} &\ge 1-u(\theta),\end{aligned} \tag{B.62}$$

which establishes the desired result of the proposition. □

The proof of Theorem 3 requires the following lemma.

LEMMA B.5. *If $N_{jt}\to\infty$ almost surely (respectively, in probability) as $t\to\infty$ for each j, then $\hat{h}_{jbt}\to h_{j1}$ and $\hat{h}_{bjt}\to h_{1j}$ almost surely (respectively, in probability) as $t\to\infty$.*

Proof of Theorem 3. We omit the proof for consistency because it follows from the identical steps as those in the proof of Theorem 2, with $\hat{f}_{jt}$ being replaced by $\hat{h}_{jbt}$ or $\hat{h}_{bjt}$ and $f_j(\xi_j)$ by h_{j1} or h_{1j}, along with Lemma B.5.

Given the consistency of the algorithm, it follows that $\hat{\xi}_{jt}\to\xi_j$ and $\hat{h}_{jt}\to h_j(u)$ as $t\to\infty$ in probability for all j by Proposition 1 and Lemma B.5. Applying Lemma B.2 with $(\boldsymbol{x}^t,\boldsymbol{y}^t)$ replaced with $(\hat{\boldsymbol{\xi}}_t,\hat{\boldsymbol{h}}_t)$ and $(\boldsymbol{x},\boldsymbol{y})$ replaced with $(\boldsymbol{\xi},\boldsymbol{h})$, where $\hat{\boldsymbol{h}}_t=(\hat{h}_{1t}(u),\ldots,\hat{h}_{kt}(u))$ and $\boldsymbol{h}=(h_1(u),\ldots,h_k(u))$, it follows that $\hat{\boldsymbol{\alpha}}^{\delta,\epsilon}_t\to\boldsymbol{\alpha}^{\delta,\epsilon*}$ as $t\to\infty$. Moreover, by construction of Algorithm 4 it is not difficult to see that the term $\boldsymbol{\alpha}_t-\hat{\boldsymbol{\alpha}}^{\delta,\epsilon}_t\to 0$.

Consequently, we have that $\boldsymbol{\alpha}_t \to \boldsymbol{\alpha}^{\delta,\epsilon*}$ as $t \to \infty$ and that $\rho(\boldsymbol{\alpha}_t)/\rho(\boldsymbol{\alpha}^{\delta,\epsilon*}) \to 1$ in probability as $t \to \infty$. Since $\rho(\boldsymbol{\alpha}_t)/\rho(\boldsymbol{\alpha}^{\delta,\epsilon*})$ is bounded, we establish that

$$\mathsf{E}\left(\frac{\rho(\boldsymbol{\alpha}_t)}{\rho(\boldsymbol{\alpha}^{\delta,\epsilon*})}\right) \to 1 \text{ as } t \to \infty. \quad \square \tag{B.63}$$

Proof of Lemma A.1. In the definition of $G_j^\delta(\boldsymbol{\alpha})$, p, ξ_j, and $f_j(\xi_j)$ are constants that are independent of $\boldsymbol{\alpha}$. Hence, without loss of generality, we can show that the function $\ell_j(\boldsymbol{\alpha})$, defined as

$$\ell_j(\boldsymbol{\alpha}) = \frac{1}{1/\alpha_1 + 1/\alpha_j} \tag{B.64}$$

is strongly concave. Observe that the Hessian matrix for $\ell_j(\boldsymbol{\alpha})$ with respect to α_1 and α_j is

$$\nabla^2 \ell_j(\boldsymbol{\alpha}) = \frac{1}{(1/\alpha_1 + 1/\alpha_j)^2}\begin{bmatrix} \frac{2}{\alpha_1^3}\left(\frac{1/\alpha_1}{1/\alpha_1+1/\alpha_j} - 1\right) & \frac{2/\alpha_1^2\alpha_j^2}{1/\alpha_1+1/\alpha_j} \\ \frac{2/\alpha_1^2\alpha_j^2}{1/\alpha_1+1/\alpha_j} & \frac{2}{\alpha_j^3}\left(\frac{1/\alpha_j}{1/\alpha_1+1/\alpha_j} - 1\right) \end{bmatrix}. \tag{B.65}$$

After some straightforward algebra, the determinant of the Hessian can be written as

$$|\nabla^2 \ell_j(\boldsymbol{\alpha})| = \frac{4((\alpha_1+\alpha_j) - (\alpha_1^4 + \alpha_j^4))}{\alpha_1\alpha_j(\alpha_1+\alpha_j)^3}. \tag{B.66}$$

In the remainder of the proof, we show that there exists a positive constant d such that

$$\min_{\boldsymbol{\alpha}\in\Delta} \; |\nabla^2 \ell_j(\boldsymbol{\alpha})| \ge 4d. \tag{B.67}$$

First, it is trivial to check that $|\nabla^2 \ell_j(\boldsymbol{\alpha})| \to \infty$ as $\alpha_1 \to 0$ and/or $\alpha_j \to 0$. Hence, there exists a small $d \in (0,1)$ such that

$$\min_{\boldsymbol{\alpha}\in\Delta} \; |\nabla^2 \ell_j(\boldsymbol{\alpha})| = \min_{\boldsymbol{\alpha}\in\Delta^d} \; |\nabla^2 \ell_j(\boldsymbol{\alpha})|, \tag{B.68}$$

where $\Delta^d = \{\boldsymbol{\alpha} \;:\; d \le \sum_{j=1}^k \alpha_j \le 1 - kd\}$. Further, since $(x+y)/2 \ge \sqrt{xy}$, observe that for $\boldsymbol{\alpha} \in \Delta^d$,

$$\begin{aligned} |\nabla^2 \ell_j(\boldsymbol{\alpha})| &\ge \frac{(\alpha_1+\alpha_j) - (\alpha_1^4+\alpha_j^4)}{(\alpha_1+\alpha_j)^5} \\ &\ge \frac{(\alpha_1+\alpha_j) - (\alpha_1+\alpha_j)^4}{(\alpha_1+\alpha_j)^5} \\ &\ge (\alpha_1+\alpha_j) - (\alpha_1+\alpha_j)^4 \\ &\ge d, \end{aligned} \tag{B.69}$$

which completes the proof of the lemma. $\square$

Proof for Proposition A.1. First, the strong consistency of the naive algorithm immediately follows, because $N_{jt}^\pi \ge c\log t \to \infty$ as $t \to \infty$ almost surely.

Next, we show that QD-C is strongly consistent under the additional assumptions in the statement of Proposition A.1. Fix a sequence of samples, $\{(X_{j1}, X_{j2}, \ldots)\}_{j=1}^k$ and let $A \subset \{1,2,\ldots,k\}$ be the set of systems that are sampled only finitely many times and let $I := \{1,2,\ldots,k\} \setminus A$. Suppose, towards a contradiction, that F is non-empty and let $\tau < \infty$ be the last time that the systems in A are sampled. It is straightforward to see, using the proof of Proposition 1, that if QD-C is strongly consistent, then $\hat{\xi}_{jt} \to \xi_j$ almost surely as $t \to \infty$. Further, from Lemma B.3 we observe that $\hat{f}_{jt}$ converges to $f_j(\xi_j)$ almost surely as $t \to \infty$ for each system $j \in I$. Also, for any system $j \in A$, $\hat{\xi}_{jt}$ and $\hat{f}_{jt}$ are constant subsequent to the stage τ. Therefore, as $t \to \infty$,

$$\hat{\xi}_{jt} \to \hat{\xi}_{j\infty} = \begin{cases} \xi_j \text{ for } j \in I \\ \hat{\xi}_{j\tau} \text{ for } j \in A, \end{cases} \tag{B.70}$$

and

$$\tilde{f}_{jt}(\hat{\xi}_{jt}) \to \tilde{f}_{j\infty} = \begin{cases} f_j(\xi_j) \text{ for } j \in I \\ \tilde{f}_{jt}(\hat{\xi}_{jt}) \text{ for } j \in A. \end{cases} \tag{B.71}$$

Let $\hat{\boldsymbol{\xi}}_t = (\hat{\xi}_{1t}, \ldots, \hat{\xi}_{kt})$ and $\hat{\boldsymbol{f}}_t = (\hat{f}_{1t}(\hat{\xi}_{jt}), \ldots, f_{kt}(\hat{\xi}_{jt}))$. We may assume that $|B(\hat{\boldsymbol{\xi}}_\infty)| < k$, i.e., $\max_j\{\hat{\xi}_{j\infty}\} > \min_j\{\hat{\xi}_{j\infty}\}$, since the event, $\hat{\xi}_{1\infty} = \cdots = \hat{\xi}_{k\infty}$, occurs with zero probability under the assumption (F3). Define $\hat{\boldsymbol{\alpha}}_t = \boldsymbol{\alpha}(\hat{\boldsymbol{\xi}}_t, \hat{\boldsymbol{f}}_t)$ for each $t \geq k$ and $\hat{\boldsymbol{\alpha}}_\infty = \boldsymbol{\alpha}(\hat{\boldsymbol{\xi}}_\infty, \hat{\boldsymbol{f}}_\infty)$. Applying Lemma B.2 with $(\boldsymbol{x}^t, \boldsymbol{y}^t)$ replaced with $(\hat{\boldsymbol{\xi}}_t, \hat{\boldsymbol{f}}_t)$ and $(\boldsymbol{x}, \boldsymbol{y})$ replaced with $(\hat{\boldsymbol{\xi}}_\infty, \hat{\boldsymbol{f}}_\infty)$, it follows that $\hat{\boldsymbol{\alpha}}_t \to \hat{\boldsymbol{\alpha}}_\infty$ as $t \to \infty$ with $\hat{\boldsymbol{\alpha}}_\infty > 0$. Further, by construction of the algorithm, $\boldsymbol{\alpha}_t - \hat{\boldsymbol{\alpha}}_t \to 0$, and therefore, $\boldsymbol{\alpha}_t \to \hat{\boldsymbol{\alpha}}_\infty$ as $t \to \infty$. However, this contradicts our assumption because $\alpha_{jt} \to 0$ for each $j \in A$ since such a system is sampled only finitely many times. Consequently, F is empty with probability 1 and QD-C is strongly consistent.

The proof for the strong consistency of QD-D follows using exactly the same logical steps as that for QD-C, with $\hat{f}_{jt}$ being replaced by $\hat{h}_{jbt}$ or $\hat{h}_{bjt}$, $f_j(\xi_j)$ by h_{j1} or h_{1j}, and Lemma B.3 by Lemma B.5. Hence, this will be omitted. □

The proof for Proposition A.2 requires the following lemma.

LEMMA B.6 **(Moderate deviations for sample quantiles)**. *Under assumptions (F1)-(F3), for any positive sequence $\{\delta_t\}$ such that $t\delta_t^2 \to \infty$ and $\delta_t \to 0$ as $t \to \infty$, a static algorithm $\pi(\boldsymbol{\alpha})$ for some $\boldsymbol{\alpha} \in \Delta^0$ satisfies*

$$\frac{1}{t\delta_t^2} \log \mathsf{P}(|\hat{\xi}_{jt} - \xi_j| > \delta_t) \to -\frac{\alpha_j f_j^2(\xi_j)}{2p(1-p)} \quad \textit{as } t \to \infty \quad \textit{for } j = 1, \ldots, k \tag{B.72}$$

Proof of Proposition A.2. In this proof, we fix a static algorithm $\pi = \pi(\boldsymbol{\alpha})$ for some $\boldsymbol{\alpha} \in \Delta^0$ and suppress π in the superscripts for clarity. Observe that

$$\max_{j=2,\ldots,k} \mathsf{P}(\hat{\xi}_{1t} \leq \hat{\xi}_{jt}; \delta_t) \leq \mathsf{P}(\mathrm{FS}_t; \delta_t) \leq (k-1) \max_{j=2,\ldots,k} \mathsf{P}(\hat{\xi}_{1t} \leq \hat{\xi}_{jt}; \delta_t). \tag{B.73}$$

Hence, if, for each $j = 2, \ldots, k$,

$$\lim_{t\to\infty} \frac{1}{t\delta_t^2} \log \mathsf{P}\left(\hat{\xi}_{1t} \leq \hat{\xi}_{jt}; \delta_t\right) = -G_j(\boldsymbol{\alpha}), \tag{B.74}$$

then

$$\lim_{t\to\infty} \frac{1}{t\delta_t^2} \log \mathsf{P}\left(\mathrm{FS}_t; \delta_t\right) = -\min_{j=2,\ldots,k} \{G_j(\boldsymbol{\alpha})\}. \tag{B.75}$$

Therefore, the rate function of $\mathsf{P}(\mathrm{FS}_t; \delta_t)$ can be immediately obtained once we prove (B.74) for some $j \neq 1$.

For simplicity, we consider a two-system case, but the proof can be extended in a straightforward manner to $k \geq 3$. Without loss of generality, assume that $\delta_1 = \xi_1 - \xi_2 = 1$. Observe that

$$\begin{aligned} \mathsf{P}(\mathrm{FS}_t; \delta_t) &= \mathsf{P}(\hat{\xi}_{1t} < \hat{\xi}_{2t}; \delta_t) \\ &= \mathsf{P}(\hat{\xi}_{1t} - \xi_{1t} < \hat{\xi}_{2t} - \xi_{2t} - \delta_t; \delta_t) \\ &= \mathsf{P}(\hat{\xi}_{1t} - \xi_1 < \hat{\xi}_{2t} - \xi_2 - \delta_t) \\ &= \mathsf{P}(\hat{\xi}_{1t} - \xi_1 + \delta_t/2 < \hat{\xi}_{2t} - \xi_2 - \delta_t/2). \end{aligned} \tag{B.76}$$

Define $\eta_{1t} = (\hat{\xi}_{1t} - \xi_1)/\delta_t + 1/2$ and $\eta_{2t} = (\hat{\xi}_{2t} - \xi_2)/\delta_t - 1/2$. Then, $\mathsf{P}(\mathrm{FS}_t; \delta_t)$ can be written as

$$\mathsf{P}(\mathrm{FS}_t; \delta_t) = \mathsf{P}(\eta_{1t} < \eta_{2t}). \tag{B.77}$$

Note that for $x \le 1/2$, $\mathsf{P}(\eta_{1t} < x) = \mathsf{P}(\hat{\xi}_{1t} < \delta_t(x - 1/2) + \xi_1)$ and therefore,

$$\lim_{t\to\infty} \frac{1}{t\delta_t^2} \log \mathsf{P}(\eta_{1t} < x) = -\left(x - \frac{1}{2}\right)^2 \frac{\alpha_1 f_1^2(\xi_1)}{2p(1-p)}. \tag{B.78}$$

Likewise, we have that, for $x \ge -1/2$,

$$\lim_{t\to\infty} \frac{1}{t\delta_t^2} \log \mathsf{P}(\eta_{2t} > x) = -\left(x + \frac{1}{2}\right)^2 \frac{\alpha_2 f_2^2(\xi_2)}{2p(1-p)}. \tag{B.79}$$

Using the same steps in the proof of Proposition 2, it can be easily seen that

$$\lim_{t\to\infty} \frac{1}{t\delta_t^2} \log \mathsf{P}(\eta_{1t} < \eta_{2t}) = -\inf_{x\in[-1/2,1/2]} \{\alpha_1 \mathscr{I}_1(x) + \alpha_2 \mathscr{I}_2(x)\}, \tag{B.80}$$

where

$$\begin{aligned} \mathscr{F}_1(x) &:= \left(x - \frac{1}{2}\right) \frac{f_1^2(\xi_1)}{2p(1-p)} \\ \mathscr{F}_2(x) &:= \left(x + \frac{1}{2}\right) \frac{f_2^2(\xi_2)}{2p(1-p)}. \end{aligned} \tag{B.81}$$

Since $\alpha_1 \mathscr{I}_1(x) + \alpha_2 \mathscr{I}_2(x)$ is a quadratic function of x, using first order conditions, one can show that the infimum is

$$\frac{1}{2p(1-p)(1/\alpha_1 f_1^2(\xi_1) + 1/\alpha_2 f_2^2(\xi_2))}, \tag{B.82}$$

which is attained at

$$x = \frac{\alpha_1 f_1^2(\xi_1) - \alpha_2 f_2^2(\xi_2)}{2(\alpha_1 f_1^2(\xi_2) + \alpha_2 f_2^2(\xi_2))}. \tag{B.83}$$

This completes the proof. □

B.2. Proofs for Auxiliary Lemmas

Proof of Lemma B.1. First, fix $x < \xi_j$ and observe that

$$\mathsf{P}\left(\hat{\xi}_{jt}^p \le x\right) = \mathsf{P}\left(\sum_{\tau=1}^{t} \mathbf{I}\{X_{j\tau} \le x\}\mathbf{I}\{\pi_\tau = j\} > [pN_{jt}]\right). \tag{B.84}$$

Note that $\mathbf{I}\{X_j \le x\}$ is a Bernoulli random variable with success probability $F_j(x)$. Hence, applying Cramer's theorem, it can be seen that

$$\frac{1}{t} \log \mathsf{P}\left(\sum_{\tau=1}^{t} \mathbf{I}\{X_{j\tau} \le x\}\mathbf{I}\{\pi_\tau = j\} > [pN_{jt}]\right) \to -\alpha_j \sup_\lambda \{\lambda x - \Lambda_j(\lambda)\} \tag{B.85}$$

as $t \to \infty$, where $\Lambda_j(\lambda) = \log \mathsf{E}[\exp(\lambda \mathbf{I}\{X_j \le x\})]$. After some straightforward algebra one can show that

$$\sup_\lambda \{\lambda x - \Lambda_j(\lambda)\} = p \log\left(\frac{p}{F_j(x)}\right) + (1-p)\log\left(\frac{1-p}{1-F_j(x)}\right). \tag{B.86}$$

From identical arguments, one can prove the case with $x' > \xi_j^p$ and this concludes the proof of the lemma. □

Proof of Lemma B.2. First, we show that $\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y})$ is strictly positive. Note that the objective function of the optimization problem (B.43) is strictly positive at its optimal solution; to see this, notice that $\boldsymbol{\alpha}_e = (1/k, \ldots, 1/k)$ is a feasible solution and the objective function is positive at $\boldsymbol{\alpha}_e$. Suppose that $\alpha_j(\boldsymbol{x}, \boldsymbol{y}) = 0$ for some j. Then the objective function of the optimization problem (B.43) is 0, which contradicts the optimality of $\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y})$.

Without loss of generality, assume that $x_1 \geq \cdots \geq x_k$ and let $b = |B(\boldsymbol{x})| < k$. Define $\epsilon \in (0, (x_1 - x_{b+1})/2)$ and let $t_0 < \infty$ such that $\max_j\{|x_j^t - x_j|\} \leq \epsilon$ and $\max_j\{|y_j^t - y_j|\} \leq \epsilon$ for all $t \geq t_0$. That is, for each $t \geq t_0$, $(\boldsymbol{x}^t, \boldsymbol{y}^t)$ is in a compact set defined by $(\boldsymbol{x}, \boldsymbol{y})$ and ϵ. Also, x_j^t is sufficiently close to x_j for each $j = 1, \ldots, k$ so that $B(\boldsymbol{x}^t) = B(\boldsymbol{x})$ for $t \geq t_0$. In the rest of the proof, it suffices to consider $t \geq t_0$.

Suppose, towards a contradiction, that $\boldsymbol{\alpha}(\boldsymbol{x}^t, \boldsymbol{y}^t)$ does not converge to $\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y})$. Then there exists a convergent subsequence, $\{t_1, t_2, \ldots\}$, such that

$$\boldsymbol{\alpha}(\boldsymbol{x}^{t_n}, \boldsymbol{y}^{t_n}) \to \tilde{\boldsymbol{\alpha}} \neq \boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y}), \tag{B.87}$$

as $n \to \infty$. Since $\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y})$ is the unique maximizer of $H(\boldsymbol{\alpha}; \boldsymbol{x}, \boldsymbol{y})$ and $\tilde{\boldsymbol{\alpha}} \neq \boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y})$, it can be seen that

$$H(\tilde{\boldsymbol{\alpha}}; \boldsymbol{x}, \boldsymbol{y}) < H(\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y}); \boldsymbol{x}, \boldsymbol{y}). \tag{B.88}$$

On the other hand, since $\boldsymbol{\alpha}(\boldsymbol{x}^{t_n}, \boldsymbol{y}^{t_n})$ is the unique maximizer of $H(\boldsymbol{\alpha}; \boldsymbol{x}^{t_n}, \boldsymbol{y}^{t_n})$,

$$H(\boldsymbol{\alpha}(\boldsymbol{x}^{t_n}, \boldsymbol{y}^{t_n}); \boldsymbol{x}^{t_n}, \boldsymbol{y}^{t_n}) \geq H(\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y}); \boldsymbol{x}^{t_n}, \boldsymbol{y}^{t_n}). \tag{B.89}$$

Note that $H(\boldsymbol{\alpha}; \boldsymbol{x}, \boldsymbol{y})$ is continuous in $\boldsymbol{\alpha}$ and $(\boldsymbol{x}, \boldsymbol{y})$ because it is minimum of the continuous functions, $\hat{d}_{ij}(\boldsymbol{\alpha}; \boldsymbol{\mu}, \boldsymbol{\sigma})$. Since $\boldsymbol{\alpha}(\boldsymbol{x}^{t_n}, \boldsymbol{y}^{t_n}) \to \tilde{\boldsymbol{\alpha}}$, $\boldsymbol{x}^{t_n} \to \boldsymbol{x}$, and $\boldsymbol{y}^{t_n} \to \boldsymbol{y}$, taking $n \to \infty$ on both sides of (B.89), we obtain that

$$H(\tilde{\boldsymbol{\alpha}}; \boldsymbol{x}, \boldsymbol{y}) \geq H(\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y}); \boldsymbol{x}, \boldsymbol{y}), \tag{B.90}$$

which contradicts (B.88). Therefore, $\boldsymbol{\alpha}(\boldsymbol{x}^t, \boldsymbol{y}^t) \to \boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{y})$ as $t \to \infty$ and the proof is complete. □

Proof of Lemma B.3. We first show weak convergence. From Proposition 1 we have that $\hat{\xi}_{jt} \to \xi_j$ in probability as $t \to \infty$. Further, observe that $\{X_{jP_n}\}$ is a sequence of independent random variables, where a random variable P_n is define as $N_{jP_n} = n$; see (B.1) in the proof of Proposition 1. Since $\hat{f}_{jt}$ is the density estimator of the independent sequence of random variables $\{X_{jP_n}\}$, we may apply Parzen (1962) and obtain that $\hat{f}_{jt}(x) \to f(x)$ almost surely as $t \to \infty$ for any $x \in \mathscr{H}_j$. Using the continuous mapping theorem, we establish the weak convergence of $\hat{f}_{jt}(\hat{\xi}_{jt})$ to $f_j(\xi_j)$.

To show strong convergence, it is easy to extend Proposition 1 to see that $\hat{\xi}_{jt} \to \xi_j$ almost surely if $N_{jt} \to \infty$ almost surely as $t \to \infty$, which follows from the independence argument above and p. 75 of Silverman (1986). Further, from Theorem 1 of Nadaraya (1965) we have that $\hat{f}_{jt}(x) \to f_j(x)$ almost surely as $t \to \infty$ under the additional conditions on the bandwidth parameter and the kernel. Combined with the continuous mapping theorem, the proof of the lemma is complete. □

Proof of Lemma B.4. Without loss of generality, let $1 = \arg\min_j\{x_j\}$ and $2 = \arg\min_j\{y_j\}$. Then,

$$\frac{\min_j\{x_j\}}{\min_j\{y_j\}} = \frac{x_1}{y_2}. \tag{B.91}$$

By definition, observe that

$$\frac{x_1}{y_2} \leq \frac{x_1}{y_1} \leq 1 + e \text{ and } \frac{x_1}{y_2} \geq \frac{x_2}{y_2} \geq 1 - e, \tag{B.92}$$

which completes the proof of the lemma. □

Proof of Lemma B.5. First, suppose $N_{jt} \to \infty$ in probability as $t \to \infty$. Then, we have that $\hat{\xi}_{jt} \to \xi_j$ in probability from Proposition 1. Also, from the same argument as in the proof for Proposition 1, it can be seen that $\hat{F}_{jt}(x)$ is an average of indicator random variables that are independent and identically distributed. Therefore, by Glivenko-Cantelli theorem, it follows that $\sup_{x\in\mathscr{X}_j} |\hat{F}_{jt}(x) - F_j(x)| \to 0$ in probability as $t \to \infty$. Consequently, from continuous mapping theorem, we have that $\hat{F}_{jt}(\hat{\xi}_{jt}) \to F_j(\xi_j)$ in probability as $t \to \infty$ for each j, from which the desired conclusion follows. The other case where $N_{jt} \to \infty$ almost surely as $t \to \infty$ follows immediately from the same steps as above, which will be omitted. □

Proof of Lemma B.6. Note that N_{jt} under the *static* policy $\pi(\boldsymbol{\alpha})$ is a deterministic value for each j and t, satisfying $N_{jt}/t \to \alpha_j$ as $t \to \infty$. We first show the upper bound:

$$\limsup_{t\to\infty} \frac{1}{t\delta_t^2} \log \mathsf{P}(|\hat{\xi}_{jt} - \xi_j| > \delta_t) \le -\frac{-\alpha_j f_j^2(\xi_j)}{2p(1-p)}. \tag{B.93}$$

Observe that, for $\theta < 0$,

$$\begin{aligned} \mathsf{P}(\hat{\xi}_{jt} > \xi_j + \delta_t) &= \mathsf{P}\left(\sum_{\tau=1}^{t} \mathbf{I}\{X_{j\tau} \le \xi_j + \delta_t\}\mathbf{I}\{\pi_\tau = j\} < pN_{jt}\right) \\ &\le \exp\left(-\theta p N_{jt}\right) \mathsf{E}\left(\exp(\theta \sum_{\tau=1}^{t} \mathbf{I}\{X_{j\tau} \le \xi_j + \delta_t\}\mathbf{I}\{\pi_\tau = j\})\right) \\ &= \left(\mathsf{E}\left(\exp(\theta(\mathbf{I}\{X_j \le \xi_j + \delta_t\} - p))\right)\right)^{N_{jt}}, \end{aligned} \tag{B.94}$$

where the inequality follows from Chernoff's inequality. Further observe that

$$\begin{aligned} \frac{1}{t\delta_t^2} \log \mathsf{P}(\hat{\xi}_{jt} > \xi_j + \delta_t) &\le -\frac{\alpha_j}{\delta_t^2} \sup_{\theta<0} \{\theta p - \log \mathsf{E}(\exp(\theta \mathbf{I}\{X_j \le \xi_j + \delta_t\}))\} \\ &= -\frac{\alpha_j}{\delta_t^2}\left(p \log\left(\frac{p}{F_j(\xi_j + \delta_t)}\right) + (1-p)\log\left(\frac{1-p}{1-F_j(\xi_j+\delta_t)}\right)\right), \end{aligned} \tag{B.95}$$

where the inequality follows from (B.94) and the fact that the inequality holds for every $\theta < 0$; the equality follows from the fact that that $\mathbf{I}\{X_j \le \xi_j + \delta_t\}$ is a Bernoulli random variable with success probability $F_j(\xi_j + \delta_t)$. Now, from the second order Taylor expansion, we have that

$$\begin{aligned} \log(F_j(\xi_j + \delta_t)) &= \log(p) - \delta_t \frac{f_j(\xi_j)}{p} + \frac{\delta_t^2}{2}\left(\frac{f_j^2(\xi_j)}{p^2} - \frac{f_j'(\xi_j)}{p}\right) + o(\delta_t^2) \\ \log(1 - F_j(\xi_j + \delta_t)) &= \log(1-p) + \delta_t \frac{f_j(\xi_j)}{1-p} + \frac{\delta_t^2}{2}\left(\frac{f_j^2(\xi_j)}{(1-p)^2} - \frac{f_j'(\xi_j)}{(1-p)}\right) + o(\delta_t^2). \end{aligned} \tag{B.96}$$

Substituting (B.96) into (B.95) and taking lim sup on both sides of the inequality, we establish that

$$\limsup_{t\to\infty} \frac{1}{t\delta_t^2} \log \mathsf{P}(\hat{\xi}_{jt} > \xi_j + \delta_t) \le -\frac{\alpha_j f_j^2(\xi_j)}{2p(1-p)}. \tag{B.97}$$

Similarly, one can easily show that

$$\limsup_{t\to\infty} \frac{1}{t\delta_t^2} \log \mathsf{P}(\hat{\xi}_{jt} < \xi_j - \delta_t) \le -\frac{\alpha_j f_j^2(\xi_j)}{2p(1-p)}, \tag{B.98}$$

and we complete the proof for the upper bound, (B.93).

Next, we show the lower bound:

$$\liminf_{t\to\infty} \frac{1}{t\delta_t^2} \log \mathsf{P}(|\hat{\xi}_{jt} - \xi_j| > \delta_t) \ge -\frac{\alpha_j f_j^2(\xi_j)}{2p(1-p)}. \tag{B.99}$$

To this end, define $Y_{j\tau}^t = \mathbf{I}\{X_{j\tau} \le \xi_j + \delta_t\}$ for $\tau \le t$ and let $Y_j^t = \mathbf{I}\{X_j \le \xi_j + \delta_t\}$ be identically distributed with $Y_{j\tau}^t$, $\tau \le t$. Also define $M_{jt}(\theta) = \mathsf{E}(\exp(\theta Y_j^t))$, $q_{jt} = F_j(\xi_j + \delta_t)$, and $q_{jt}(\theta) = F_j(\xi_j + \delta_t)\exp(\theta)/M_{jt}(\theta)$. If we let $\theta_t = \arg\max_{\theta \in \mathcal{R}}\{\theta p - \log M_{jt}(\theta)\}$, then it can be easily seen that $\theta_t > 0$ is well defined and $q_{jt}(\theta_t) = p$. For any function $g : \mathcal{R}^t \to \mathcal{R}$ with $\mathsf{E}|g(Y_{j1}^t, \ldots, Y_{jt}^t)| < \infty$, $\mathsf{E}_\theta(g(Y_{j1}^t, \ldots, Y_{jt}^t))$ denotes the expectation with respect to the measure $q_{jt}(\theta)$ for which $Y_{jt}^t = 1$ with probability $q_{jt}(\theta)$ and 0 with probability $1 - q_{jt}(\theta)$. Observe that

$$\begin{aligned}
\mathsf{P}(\hat{\xi}_{jt} > \xi_j + \delta_t) &= \mathsf{P}\left(\sum_{\tau=1}^t Y_{j\tau}\mathbf{I}\{\pi_\tau = j\} > pN_{jt}\right) \\
&= \mathsf{E}\left(\mathbf{I}\left\{\sum_{\tau=1}^t Y_{j\tau}\mathbf{I}\{\pi_\tau = j\} > pN_{jt}\right\}\right) \\
&= \mathsf{E}_{\theta_t}\left(\mathbf{I}\left\{\sum_{\tau=1}^t Y_{j\tau}\mathbf{I}\{\pi_\tau = j\} > pN_{jt}\right\}\prod_{\tau=1}^t \left(\frac{q_{jt}}{p}\right)^{Y_{j\tau}^t\mathbf{I}\{\pi_\tau=j\}}\left(\frac{1-q_{jt}}{1-p}\right)^{(1-Y_{j\tau}^t)\mathbf{I}\{\pi_\tau=j\}}\right)
\end{aligned} \tag{B.100}$$

by the change of measure and the fact that $q_{jt}(\theta_t) = p$ by definition of θ_t. After some straightforward algebra, we have that

$$\begin{aligned}
&\prod_{\tau=1}^t \left(\frac{q_{jt}}{p}\right)^{Y_{j\tau}^t\mathbf{I}\{\pi_\tau=j\}}\left(\frac{1-q_{jt}}{1-p}\right)^{(1-Y_{j\tau}^t)\mathbf{I}\{\pi_\tau=j\}} \\
&= \exp\left(-\log\left(\frac{p}{q_{jt}}\sum_{\tau=1}^t Y_{j\tau}^t\mathbf{I}\{\pi_\tau = j\}\right) - \log\left(\frac{1-p}{1-q_{jt}}\sum_{\tau=1}^t (1 - Y_{j\tau}^t)\mathbf{I}\{\pi_\tau = j\}\right)\right) \\
&= \exp\left(-N_{jt}\left(p\log\left(\frac{p}{q_{jt}}\right) + (1-p)\log\left(\frac{1-p}{1-q_{jt}}\right)\right)\right)\exp\left(-\beta_1\sum_{\tau=1}^t (Y_{j\tau}^t - p)\mathbf{I}\{\pi_\tau = j\}\right),
\end{aligned} \tag{B.101}$$

where $\beta_1 = \log(p/q_{jt}) + \log((1-p)/(1-q_{jt}))$. Therefore, rewriting (B.100),

$$\begin{aligned}
&\mathsf{P}(\hat{\xi}_{jt} > \xi_j + \delta_t) \\
&= \exp\left(-N_{jt}\left(p\log\left(\frac{p}{q_{jt}}\right) + (1-p)\log\left(\frac{1-p}{1-q_{jt}}\right)\right)\right) \times \\
&\quad \mathsf{E}_{\theta_t}\left(\mathbf{I}\left\{\sum_{\tau=1}^t Y_{j\tau}\mathbf{I}\{\pi_\tau = j\} > pN_{jt}\right\}\exp\left(-\beta_1\sum_{\tau=1}^t (Y_{j\tau}^t - p)\mathbf{I}\{\pi_\tau = j\}\right)\right),
\end{aligned} \tag{B.102}$$

Further, for some $\beta_2 > 0$, the expectation on the right-hand side of (B.102) can be bounded as

$$\begin{aligned}
&\mathsf{E}_{\theta_t}\left(\mathbf{I}\left\{\sum_{\tau=1}^t Y_{j\tau}\mathbf{I}\{\pi_\tau = j\} > pN_{jt}\right\}\exp\left(-\beta_1\sum_{\tau=1}^t (Y_{j\tau}^t - p)\mathbf{I}\{\pi_\tau = j\}\right)\right) \\
&\ge \mathsf{E}_{\theta_t}\left(\mathbf{I}\left\{pN_{jt} < \sum_{\tau=1}^t Y_{j\tau}\mathbf{I}\{\pi_\tau = j\} < pN_{jt} + \beta_2 N_{jt}^{0.5}\right\}\exp\left(-\beta_1\beta_2 N_{jt}^{0.5}\right)\right) \\
&= \exp\left(-\beta_1\beta_2 N_{jt}^{0.5}\right)\mathsf{P}_{\theta_t}\left(pN_{jt} < \sum_{\tau=1}^t Y_{j\tau}\mathbf{I}\{\pi_\tau = j\} < pN_{jt} + \beta_1\beta_2 N_{jt}^{0.5}\right),
\end{aligned} \tag{B.103}$$

where the inequality follows from bounding the summation, $\sum_{\tau=1}^t Y_{j\tau}\mathbf{I}\{\pi_\tau = j\}$, in the indicator function. Note that

$$\begin{aligned}
&\mathsf{P}_{\theta_t}\left(pN_{jt} < \sum_{\tau=1}^t Y_{j\tau}\mathbf{I}\{\pi_\tau = j\} < pN_{jt} + \beta_1\beta_2 N_{jt}^{0.5}\right) \\
&= \mathsf{P}_{\theta_t}\left(0 < \frac{1}{N_{jt}^{0.5}}\sum_{\tau=1}^t (Y_{j\tau} - p)\mathbf{I}\{\pi_\tau = j\} < \beta_1\beta_2\right) \\
&\to c \in (0, 1),
\end{aligned} \tag{B.104}$$

since $Y_{j\tau}$, $\tau \leq t$, is a Bernoulli random variable with success probability p under $\mathsf{P}_{\theta_t}(\cdot)$. Combining (B.102), (B.103), (B.104), and (B.96), we establish that

$$\liminf_{t\to\infty} \frac{1}{t\delta_t^2} \log \mathsf{P}(\hat{\xi}_{jt} > \xi_j + \delta_t) \geq -\frac{\alpha_j f_j^2(\xi_j)}{2p(1-p)}. \tag{B.105}$$

Likewise, one can easily obtain that

$$\liminf_{t\to\infty} \frac{1}{t\delta_t^2} \log \mathsf{P}(\hat{\xi}_{jt} < \xi_j - \delta_t) \geq -\frac{\alpha_j f_j^2(\xi_j)}{2p(1-p)}, \tag{B.106}$$

which completes the proof for the lower bound (B.99). Combining (B.93) and (B.99), the proof for the lemma is complete. □

Appendix C: Parameter Sensitivity of the Algorithms

The algorithms proposed in this paper commonly take the number of initial samples n and the batch size m as parameters. These slightly affect the performance of the proposed algorithms, but we remark that the qualitative conclusions from §6.2 are robust relative to the choice of these parameters. In this section, we examine the sensitivity of performance with respect to parameters that are specific to each algorithm.

First, the naive algorithm takes c as an additional parameter, which ensures that each system is sampled at least at a logarithmic rate. To see how the performance of the naive algorithm is affected by the choice of c, we consider a two-system configuration with binomial distributions; the number of trials is 10 for both systems and the success probabilities are $\boldsymbol{\mu} = (0.62, 0.55)$, respectively. We set $p = 0.1$. In this case, the vector $\boldsymbol{\alpha}^* = (0.76, 0.24)$ maximizes the rate function. We vary $c = 0, 20$.

Figure 1(a) shows a sample path of α_{1t} under the naive algorithm with $c = 0$, in which system 1 is not sampled after $n_0 = 10$ initial samples. The relative frequency of α_{1t} for $t = 4000$ over many paths is illustrated in Figure 1(b) using 10^4 simulation replications. As noted in Remark 2, the naive algorithm may not sample a particular system for a long period without forced sampling, potentially increasing the likelihood of selecting a non-best system. Figure 1(c) presents a typical sample path α_{1t} under the naive algorithm with forced sampling and Figure 1(d) gives the relative frequency chart of α_{1t} for $t = 4000$.

Next, the QD-C and AQD-C algorithms take the kernel $K(\cdot)$ and $h(\cdot)$ as parameters. The optimal choice of the bandwidth parameter is $h(n) = dn^{-1/5}$, where n is the number of observations and the constant d depends on the true density function that is a priori unknown (Silverman 1986). In order to show the sensitivity of these algorithms with respect to the kernel and the bandwidth parameter, we use normal and triangular kernels and four bandwidth functions; $h(n) = 1.06\sigma n^{-1/5}$, where σ is replaced with sample standard deviation, and $h(n) = h = 0.1, 1, 10$. We use the configuration with $k = 4$ normally distributed systems with $\boldsymbol{\mu} = (0, 0, 0, 0)$ and $\boldsymbol{\sigma} = (1, 1.1, 1.2, 1.3)$. We let $p = 0.1$.

Figures 2(a)-(b) illustrate the performance of the QD-C and AQD-C algorithms in terms of the probability of false selection when we use the normal kernel. It can be seen that the performance of the two algorithms is degraded when the bandwidth parameter is $h = 0.1$ (too little smoothing) or $h = 10$ (too much smoothing). The performance is robust when we use the adaptive bandwidth $h(n) = 1.06\sigma n^{-1/5}$. Figures 2(c)-(d) show the same experiments with the triangular kernel, in which similar conclusions hold.

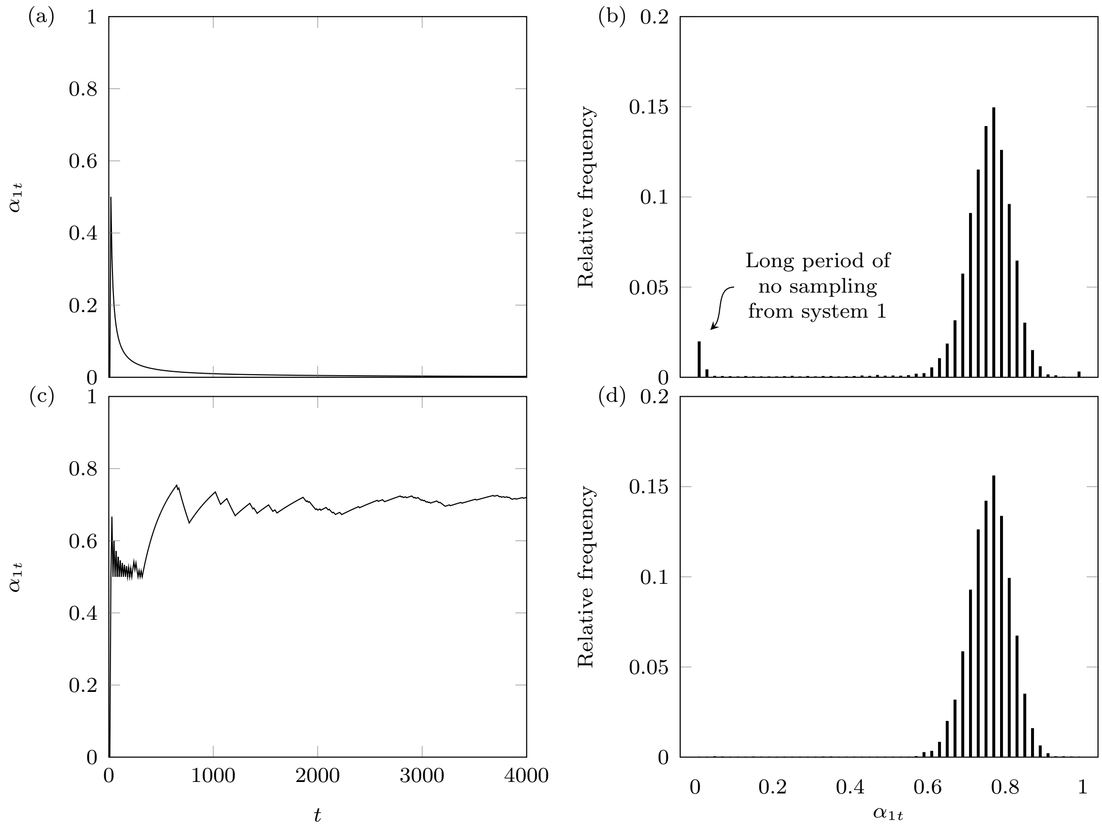

Figure 1 **The frequencies of α_{1t} under the naive algorithms for different values of parameter c. In all cases, we consider two systems with binomial distributions parameterized by the number of trials $n = (10, 10)$ and the success probability $\mu = (0.62, 0.55)$. We set $p = 0.1$, in which case $\alpha_1^* = 0.76$. We use $n_0 = 10$ and $m = 10$. Panel (a) shows a sample path of the naive algorithm with $c = 0$, where the system 1 is not sampled over the long horizon. Panel (c) shows a sample path of the naive algorithm with $c = 20$ using the same sequence of random samples. Panels (b) and (d) give the relative frequence of α_{1t} at $t = 4000$ using 10^4 trials of the naive algorithm.**

References

Nadaraya EA (1965) On non-parametric estimates of density functions and regression curves. *Theory of Prob. & Its Applications* 10(1):186–190.

Parzen E (1962) On estimation of a probability density function and mode. *Annals of Math. Statist.* 33:1065–1076.

Peng Y, Chen CH, Fu MC, Hu JQ (2016) Dynamic sampling allocation and design selection. *INFORMS J. on Computing* 28(2):195–208.

Serfling RJ (2009) *Approximation theorems of mathematical statistics* (New York: John Wiley & Sons).

Silverman BW (1986) *Density estimation for statistics and data analysis* (New York: CRC press).

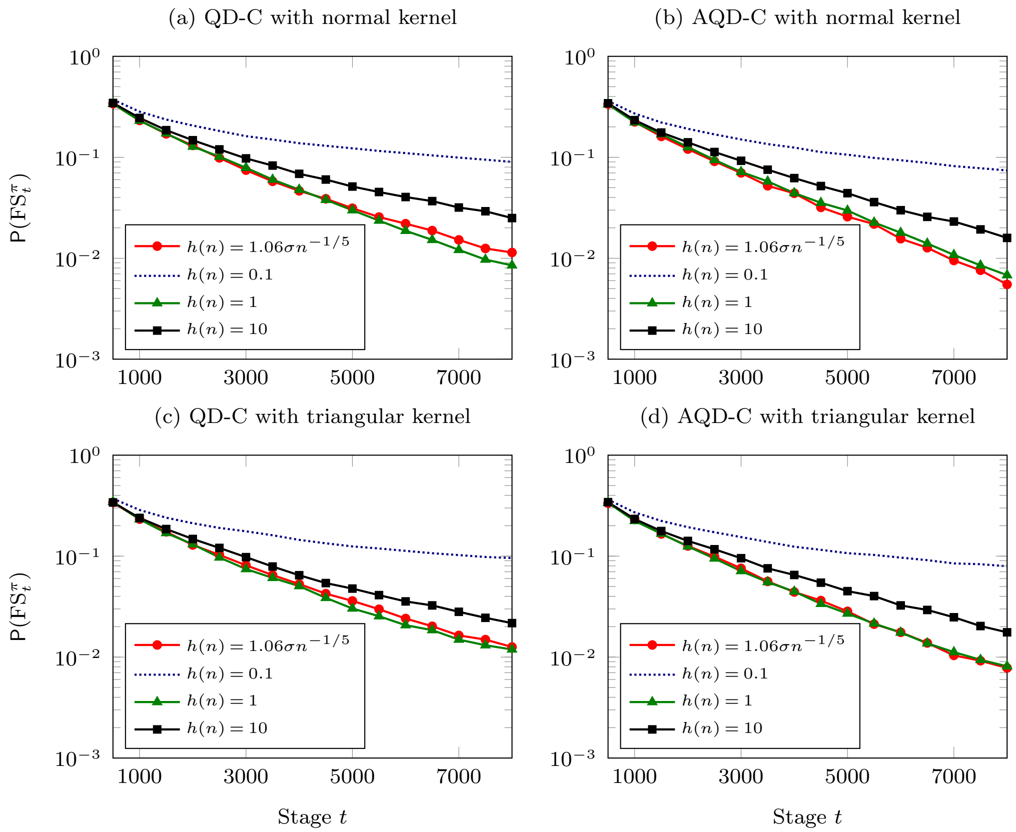

Figure 2 $\mathsf{P}(\mathbf{FS}_t^\pi)$ **as a function of stage** t **on a log-linear scale for different kernels and bandwidth parameters. We test the QD-C and AQD-C algorithms in the case with four normally distributed systems;** $\boldsymbol{\mu} = (0,0,0,0)$ **and** $\boldsymbol{\sigma} = (1, 1.1, 1.2, 1.3)$**. We let** $p = 0.1$ **and** $(n_0, m) = (20, 10)$ **for both algorithms. The performance of the two algorithms is not sensitive with respect to the choice of the kernel, but it degrades when the bandwidth parameter** $h(\cdot)$ **is too small or too large.**