Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

TRACTABLE SAMPLING STRATEGIES FOR QUANTILE-BASED ORDINAL OPTIMIZATION

Dongwook Shin
Mark Broadie
Assaf Zeevi

Graduate School of Business
3022 Broadway
Columbia University
New York, NY 10027, U.S.A.

ABSTRACT

This paper describes and analyzes the problem of selecting the best of several alternatives ("systems"), where they are compared based on quantiles of their performances. The quantiles cannot be evaluated analytically but it is possible to sequentially sample from each system. The objective is to dynamically allocate a finite sampling budget to minimize the probability of falsely selecting non-best systems. To formulate this problem in a tractable form, we introduce an objective associated with the probability of false selection using large deviations theory and leverage it to design well-performing dynamic sampling policies. We first propose a naive policy that optimizes the aforementioned objective when the sampling budget is sufficiently large. We introduce two variants of the naive policy with the aim of improving finite-time performance; these policies retain the asymptotic performance of the naive one in some cases, while dramatically improving its finite-time performance.

1 INTRODUCTION

Given a finite number of alternatives, henceforth referred to as *systems*, we are concerned with the problem of selecting the best system in the situation where performances of the systems are initially unknown but it is possible to sequentially sample from each system. A critical assumption made in most academic studies is that a decision maker is primarily interested in the average performances of the systems. However, these mean-based procedures are not flexible to accommodate various risk preferences of the decision maker.

Alternatively, the systems can be compared based on pth quantiles of their performances, where p can be chosen based on the decision maker's risk preference. The quantile-based procedure has a practical importance when downside or upside risk is more critical than mean performance. As a prototypical example, consider a call center system in which response time to customer is important. When evaluating possible designs of the systems, analysts may be interested in exceptionally high response times, which may lead to a significant decrease in potential profit. In this case the desired performance measure is a quantile.

The main objective of this paper is to design a dynamic sampling policy that minimizes the probability of falsely selecting non-best systems subject to a given sampling budget. Unfortunately, as is well documented in the literature, this objective is not analytically tractable. To arrive at a tractable objective, our departure point will be an asymptotic benchmark characterized by large sampling budget. In this regime, the objective can be written in an analytically tractable form and we leverage it to design a well-performing policy.

Building on structural insights from the asymptotic benchmark, we introduce three dynamic sampling policies. The first policy is a naive approach; it repeatedly estimates the aforementioned objective function from history of sample observations, and then allocates a sample in each stage as if the estimated objective

function is the true objective. We show that this policy is asymptotically optimal with respect to the aforementioned objective, but exhibits poor finite-time performance because such an objective function is difficult to estimate accurately. The other two policies are designed with the aim of improving the finite-time performance of the former for continuous and discrete cases, respectively. The key idea is to approximate the objective function, which is quite susceptible to sampling errors, by a simple function that is stable against such errors. We show that the alternative algorithms retain the asymptotic performance of the former in some cases, while dramatically improving the finite-time performance.

An area closely related to the ordinal optimization is the Ranking and Selection (R&S) problem, where the goal is to take as few samples as possible to satisfy a desired guarantee on the probability of correct selection. See a survey paper by Kim and Nelson (2006). Despite wide usage of quantile as a performance metric, the topic of quantile-based R&S procedures has not received much attention in this literature. Bekki et al. (2007) modify a traditional two-stage IZ procedure by Rinott (1978) to suggest a grouped quantile approach where a quantile estimate from micro-replications is taken as a single estimate of macro-replications. Although normality assumption is violated, this procedure takes the average of macro-replications to make quantile estimates nearly normal. Batur and Choobineh (2010) also suggest a two-stage procedure based on Rinott (1978), where a set of quantile values is compared between two systems. Recently, 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 the ordinal optimization framework, Pasupathy et al. (2010) characterize the rate function associated with the probability of false selection using large deviations theory when full information on the underlying distribution functions is given a priori. The performance criterion introduced in this paper has a close connection to this rate function, however, our work is fundamentally different than Pasupathy et al. (2010) in that one needs to learn the underlying probability distributions and simultaneously allocate budget to optimize the objective.

The remainder of the paper is organized as follows. In Section 2 we present large deviation preliminaries and formulate the problem. In Section 3 we propose dynamic policies and provide main theoretical results. In Section 4 we give numerical experimentation using the proposed policies and discuss the results. All proofs can be found in Shin et al. (2016).

2 FORMULATION

2.1 Model Primitives

Consider k stochastic systems, whose performance is governed by a random variable X_j with a distribution function $F_j(\cdot)$, $j = 1,\ldots,k$. Fix $p \in (0,1)$ that represents the quantile of interest and define the pth quantile of $F_j(\cdot)$ as

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

Denote $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_k)$ the k-dimensional vector of the pth quantiles. We assume that $\xi_1^p > \xi_2^p \geq \cdots \geq \xi_k^p$. A decision maker is given a sampling budget T, which means T independent samples can be drawn from the k systems, and the goal is to correctly identify the system with the largest pth quantile. Further, to avoid trivial cases where the probability of false selection is zero, we assume that $[\xi_k^p, \xi_1^p] \subset \cap_{j=1}^k \mathscr{H}_j^0$, where $\mathscr{H}_j = \{x \in \mathcal{R} \mid F_j(x) \in (0,1)\}$ and A^0 denotes an interior of any set A. Essentially, this ensures the sample pth quantile (to be made precise below) from each system can take any value in the interval $[\xi_k^p, \xi_1^p]$.

Let $\boldsymbol{\pi}$ denote a policy, 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 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 based on all the sampling decisions and samples observed in previous stages.

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

j is defined as

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

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

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

For brevity, the superscript $\boldsymbol{\pi}$ may be dropped when it is clear from the context. With a single subscript t, $\boldsymbol{N}_t = (N_{1t}, \ldots, N_{kt})$ and $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{kt})$ denote vectors of the cumulative number of samples and the sampling rates in stage t, respectively. Likewise, we let $\hat{\boldsymbol{\xi}}_t^{\,p} = (\hat{\xi}_{1t}^{p}, \ldots, \hat{\xi}_{kt}^{p})$ be the vector of sample quantile estimates.

In what follows, we also consider static policies, in which the sampling decisions up to stage T are fixed in the beginning of stage 1. A static policy $\boldsymbol{\pi}^{\boldsymbol{\alpha}}$ is characterized by a vector $\boldsymbol{\alpha} \in \Delta$ with

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

so that $N_{jt} = \alpha_j t$ for each j, ignoring integrality constraint on N_{jt}.

2.2 Large Deviations Preliminaries

The probability of false selection, denoted $\mathsf{P}(\mathrm{FS}_t)$ with $\mathrm{FS}_t = \{\hat{\xi}_{1t}^{p} < \max_{j\neq 1} \hat{\xi}_{jt}^{p}\}$, is a widely used criterion for the efficiency of a sampling policy (see, e.g., a survey paper by Kim and Nelson 2006). However, the exact evaluation of $\mathsf{P}(\mathrm{FS}_t)$ is not analytically tractable. Alternatively, we view this measure in light of large deviations theories. In this regime, one can characterize how fast $\mathsf{P}(\mathrm{FS}_t)$ converges to 0.

To this end, fix a static policy $\boldsymbol{\pi}^{\boldsymbol{\alpha}}$ for some $\boldsymbol{\alpha} \in \Delta$ and observe that $\mathsf{P}(\hat{\xi}_{jt}^{p} > x) = \mathsf{P}(\sum_{s=1}^{N_{jt}} \mathbf{I}\{X_{j\tau_j(s)} < x\} < [pN_{jt}])$, where $\tau_j(s) = \inf\{t : N_{jt} \ge s\}$ and $[y]$ is the greatest integer less than y. Similarly, $\mathsf{P}(\hat{\xi}_{jt}^{p} < x) = \mathsf{P}(\sum_{s=1}^{N_{jt}} \mathbf{I}\{X_{j\tau_j(s)} < x\} > [pN_{jt}])$. Hence, applying the Cramer's theorem for the sum of Bernoulli random variables (Dembo and Zeitouni 2009), the large deviation probability for $\hat{\xi}_{jt}^{p}$ can be characterized as follows:

$$\begin{aligned} \lim_{t\to\infty} \frac{1}{t} \log \mathsf{P}(\hat{\xi}_{jt}^{p} > x) &= -\alpha_j I_j(x) \quad \text{for } x > \xi_j^p \\ \lim_{t\to\infty} \frac{1}{t} \log \mathsf{P}(\hat{\xi}_{jt}^{p} < x') &= -\alpha_j I_j(x') \quad \text{for } x' < \xi_j^p, \end{aligned} \tag{5}$$

where

$$I_j(x) = p \log\left(\frac{p}{F_j(x)}\right) + (1-p)\log\left(\frac{1-p}{1-F_j(x)}\right). \tag{6}$$

Next, we impose the following assumptions on the distribution functions to avoid technical difficulties in the development of theoretical results.

(F1) ξ_j^p is the unique solution x of $F_j(x-) \le p \le F_j(x)$

(F2) For each $j \neq 1$, $I_1(x_1) < I_j(x_1)$ and $I_1(x_j) > I_j(x_j)$, where x_j is the smallest minimizer of $I_j(x)$

Assumption (F1) is a mild assumption that ensures $F_j(\cdot)$ is not flat around the pth quantile. Assumption (F2) is trivially satisfied for continuous distributions, but it rules out certain families of discrete distribution functions with large jumps.

The following proposition characterizes the convergence rate of the probability of false selection.

Proposition 1 Suppose assumption (F1) holds. For a static policy $\boldsymbol{\pi}^{\boldsymbol{\alpha}}$ for some $\boldsymbol{\alpha} \in \Delta$,

$$\lim_{t\to\infty} \frac{1}{t} \log \mathsf{P}(\mathrm{FS}_t) = -\rho(\boldsymbol{\alpha}), \tag{7}$$

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

$$G_j(\boldsymbol{\alpha}) = \inf_{x\in[\xi_j^p, \xi_1^p]} \left\{ \alpha_1 I_1(x) + \alpha_j I_j(x) \right\}. \tag{8}$$

To rephrase Proposition 1, $\mathsf{P}(\mathrm{FS}_t)$ behaves roughly like $\exp(-\rho(\boldsymbol{\alpha})t)$ for large values of t. It follows that the best possible asymptotic performance can be characterized by the allocation that maximizes $\rho(\cdot)$. The properties of the function $\rho(\boldsymbol{\alpha})$ is summarized in the next proposition.

Proposition 2 Under assumption (F1), $\rho(\boldsymbol{\alpha})$ is a continuous, concave function of $\boldsymbol{\alpha} \in \Delta$. If assumption (F2) is further satisfied, each element of $\boldsymbol{\alpha}^* \in \mathrm{argmax}_{\boldsymbol{\alpha}\in\Delta}\{\rho(\boldsymbol{\alpha})\}$ is strictly positive.

2.3 Problem Formulation

The function $\rho(\cdot)$ measures the asymptotic efficiency of an allocation in relation to the rate function associated with $\mathsf{P}(\mathrm{FS}_t)$. We define the *relative efficiency* $\mathscr{R}_t^{\boldsymbol{\pi}}$ for any given policy $\boldsymbol{\pi} \in \Pi$ in stage t to be

$$\mathscr{R}_t^{\boldsymbol{\pi}} = \frac{\rho(\boldsymbol{\alpha}_t)}{\rho^*}, \tag{9}$$

where $\rho^* = \max_{\boldsymbol{\alpha}\in\Delta}\{\rho(\boldsymbol{\alpha})\}$. By definition, the value of $\mathscr{R}_t^{\boldsymbol{\pi}}$ lies in the interval $[0,1]$; an allocation is considered efficient when $\mathscr{R}_t^{\boldsymbol{\pi}}$ is close to 1 for sufficiently large t.

Definition 1 (Consistency) A policy $\boldsymbol{\pi} \in \Pi$ is consistent if $N_{jt} \to \infty$ almost surely as $t \to \infty$ for each j.

Recall that Π is the set of all non-anticipating policies. 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 3 (Consistency of quantile estimators) Under assumption (F1),

$$\hat{\xi}_{jt}^{\,p} \to \xi_j^p \tag{10}$$

almost surely as $t \to \infty$ for any consistent policy $\boldsymbol{\pi} \in \bar{\Pi}$.

Note that consistent policies ensure $\mathsf{P}(\mathrm{FS}_t) \to 0$ as $t \to \infty$ since each sample quantile converges to its population counterpart. We remark that the consistency is not a trivial result for a dynamic policy, although any static policy $\boldsymbol{\pi}^{\boldsymbol{\alpha}}$ with $\boldsymbol{\alpha} \in \Delta^0$ is consistent since $N_{jt} = \alpha_j t \to \infty$ as $t \to \infty$.

We are interested in the policy that maximizes the expected relative efficiency with the budget T:

$$\sup_{\boldsymbol{\pi}\in\bar{\Pi}} \mathsf{E}(\mathscr{R}_T^{\boldsymbol{\pi}}). \tag{11}$$

We introduce a notion of asymptotic optimality, which will be used in the analysis of dynamic sampling policies in the following section.

Definition 2 (Asymptotic optimality) A policy $\boldsymbol{\pi} \in \bar{\Pi}$ is asymptotically optimal if

$$\mathsf{E}(\mathscr{R}_T^{\boldsymbol{\pi}}) \to 1 \text{ as } T \to \infty. \tag{12}$$

3 PROPOSED POLICIES AND MAIN THEORETICAL RESULTS

3.1 A Naive Approach

The preliminary observations in Section 2 imply that a policy may achieve optimal relative efficiency asymptotically if its allocation $\boldsymbol{\alpha}_t$ converges to $\boldsymbol{\alpha}^* \in \text{argmax}_{\boldsymbol{\alpha}\in\Delta}\{\rho(\boldsymbol{\alpha})\}$ as $t \to \infty$. We first suggest a naive policy that iteratively estimate such an optimal allocation from the history of sample observations. Define $\hat{\rho}_t(\boldsymbol{\alpha}) = \min_{j\neq b}\{\hat{G}_{jt}(\boldsymbol{\alpha})\}$ with $b = \text{argmax}_{j\neq 1}\{\hat{\xi}_{jt}^{\,p}\}$,

$$\hat{G}_{jt}(\boldsymbol{\alpha}) = \inf_{x\in[\hat{\xi}_{jt}^{\,p},\hat{\xi}_{bt}^{\,p}]}\{\alpha_b \hat{I}_{bt}(x) + \alpha_j \hat{I}_{jt}(x)\} \tag{13}$$

for $j \neq b$, and

$$\hat{I}_{jt}(x) = p\log\left(\frac{p}{\hat{F}_{jt}(x)}\right) + (1-p)\log\left(\frac{1-p}{1-\hat{F}_{jt}(x)}\right) \tag{14}$$

for $j = 1,\ldots,k$. Note that $\hat{\rho}_t(\boldsymbol{\alpha})$ is the estimation of $\rho(\boldsymbol{\alpha})$ in stage t with $\{F_j(\cdot)\}_{j=1}^k$ replaced with its empirical counterpart, $\{\hat{F}_{jt}(\cdot)\}_{j=1}^k$. Define $\hat{\boldsymbol{\alpha}}_t \in \text{argmax}_{\boldsymbol{\alpha}\in\Delta}\{\hat{\rho}_t(\boldsymbol{\alpha})\}$. The naive policy is designed to make $\boldsymbol{\alpha}_t$ approach $\boldsymbol{\alpha}^*$ as $t \to \infty$.

Algorithm 1 $\boldsymbol{\pi}(n_0)$

For each j, take n_0 samples and let $t = kn_0$
while $t \leq T$ **do**
Solve for $\hat{\boldsymbol{\alpha}}_t \in \text{argmax}_{\boldsymbol{\alpha}\in\Delta}\{\hat{\rho}_t(\boldsymbol{\alpha})\}$
Take a sample from system π_{t+1} such that

$$\pi_{t+1} = \begin{cases} \text{argmin}_j\{\alpha_{jt}\} & \text{if } \hat{\alpha}_{jt} = 0 \text{ for some } j \\ \text{argmax}_j\{\hat{\alpha}_{jt} - \alpha_{jt}\} & \text{otherwise} \end{cases} \tag{15}$$

Let $t = t+1$
end while

Theorem 1 (Asymptotic performance of Algorithm 1) Under assumptions (F1)-(F2), a policy by Algorithm 1 is consistent and asymptotically optimal.

Remark 1 (Poor finite-time performance of Algorithm 1) Despite the theoretical guarantee on its asymptotic performance, the Algorithm 1 exhibits poor finite-time performance as will be illustrated in Section 4, because the estimations of $\rho(\cdot)$ and its maximizer $\boldsymbol{\alpha}^*$ require full information on the distribution functions for certain ranges of their domains. This causes significant portion of the budget wasted in estimating them, leaving less budget to optimize the objective function. This intuitive argument will be more precise in the next subsection by comparing with an alternative scheme that judiciously balances the two competing goals.

3.2 Alternative Approach for Continuous Distributions

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

(F3) $F_j(x)$ possesses a positive continuous density $f_j(x)$ for $x \in \mathscr{H}_j$ and a bounded second derivative at $x = \xi_j^{\,p}$.

It is trivial to check that the smoothness assumption (F3) implies (F1)-(F2). Note that the bounded second derivative at $x = \xi_j^{\,p}$ ensures that $I_j(x)$ can be closely approximated, using a Taylor expansion, by a quadratic function of x in a neighborhood of $\xi_j^{\,p}$, which will be a key to the theoretical results in this subsection.

Let $\delta = \xi_1^p - \xi_2^p$ 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

$$G_j^\delta(\boldsymbol{\alpha}) = \frac{(\xi_1^p - \xi_j^p)^2}{2p(1-p)\left(1/(\alpha_1 f_1^2(\xi_1^p)) + 1/(\alpha_j f_j^2(\xi_j^p))\right)}. \tag{16}$$

To provide some intuition behind the definition of $\rho^\delta(\cdot)$, note that, under assumption (F3), the sample pth quantile of system j is asymptotically normal with mean ξ_j^p and variance $p(1-p)/(\alpha_j f_j^2(\xi_j^p)t)$ (see, e.g., Serfling 2009). After scaling by $1/t$, each term on the right-hand side of (16) can be viewed as the difference between the two quantiles, in terms of the number of standard errors of the difference. Therefore, if $\rho^\delta(\boldsymbol{\alpha})$ is large, one can conclude with more confidence whether ξ_1^p is greater than ξ_j^p, $j \neq 1$, because of the smaller (asymptotic) variance of the difference. Hence, one might expect that $\rho^\delta(\boldsymbol{\alpha})$ is closely aligned with $\rho(\boldsymbol{\alpha})$. This intuitive observation is summarized in the following proposition.

Proposition 4 (Validity of approximation $\rho^\delta(\cdot)$ for $\rho(\cdot)$) Under assumption (F3), for any $\boldsymbol{\alpha} \in \Delta$

$$|\rho(\boldsymbol{\alpha}) - \rho^\delta(\boldsymbol{\alpha})| = o(\delta^2) \text{ as } \delta \to 0 \tag{17}$$

Also, $\rho^\delta(\boldsymbol{\alpha})$ has a unique maximum $\boldsymbol{\alpha}^{\delta *}$ which is strictly positive.

Proposition 4 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.

Remark 2 (Structural properties of $\boldsymbol{\alpha}^{\delta *}$) We note that, as opposed to the nested structure in the optimization problem of Algorithm 1, it is more computationally efficient to solve for $\boldsymbol{\alpha}^{\delta *}$. In particular, the objective function of (19), being a minimum of concave functions, is concave for $\boldsymbol{\alpha} \in \Delta$. Further, from the first order conditions (Avriel 2003), the following equations can be used to determine $\boldsymbol{\alpha}^{\delta *} = \text{argmax}_{\boldsymbol{\alpha}\in\Delta}\{\rho^\delta(\boldsymbol{\alpha})\}$:

$$\begin{aligned} \frac{(\xi_1^p - \xi_i^p)^2}{1/(\alpha_b^{\delta *} f_1^2(\xi_1^p)) + 1/(\alpha_i^{\delta *} f_i^2(\xi_i^p))} &= \frac{(\xi_1^p - \xi_i^p)^2}{1/(\alpha_1^{\delta *} f_1^2(\xi_1^p)) + 1/(\alpha_j^{\delta *} f_j^2(\xi_j^p))}, \text{ for } i, j \neq 1 \\ (\alpha_1^{\delta *})^2 f_1^2(\xi_1^p) &= \sum_{j\neq 1} (\alpha_j^{\delta *})^2 \hat{f}_j^2(\xi_j^p). \end{aligned} \tag{18}$$

Based on Proposition 4, we now propose an alternative dynamic policy that iteratively estimates $\boldsymbol{\alpha}^{\delta *} = \text{argmax}_{\boldsymbol{\alpha}\in\Delta}\{\rho^\delta(\boldsymbol{\alpha})\}$ from the history of sample observations. Specifically, denote $\hat{\boldsymbol{\alpha}}_t^\delta$ as the estimator of $\boldsymbol{\alpha}^{\delta *}$ in stage t; formally,

$$\hat{\boldsymbol{\alpha}}_t^\delta = \underset{\boldsymbol{\alpha}\in\Delta}{\text{argmax}} \left\{ \min_{j\neq b} \frac{(\hat{\xi}_{bt}^p - \hat{\xi}_{jt}^p)^2}{2p(1-p)\left(1/(\alpha_b \hat{f}_{bt}^2) + 1/(\alpha_j \hat{f}_{jt}^2)\right)} \right\}, \tag{19}$$

where $b = \text{argmax}_j\{\hat{\xi}_{jt}^p\}$ and for each j,

$$\hat{f}_{jt} = \frac{1}{N_{jt}} \sum_{\tau=1}^{t} K_{h(\tau)}(\hat{\xi}_{j\tau}^p - X_{j\tau}) \mathbf{I}\{\pi_\tau = j\} \tag{20}$$

is the kernel-based estimator of density at $\hat{\xi}_{jt}^p$ with a kernel function $K(\cdot)$ and a bandwidth parameter $h(t) \geq 0$ for each t. A kernel with subscript h is called a scaled kernel and defined as $K_h(x) = 1/hK(x/h)$. The optimal choices of the kernel function and the bandwidth parameter are not in the scope of this paper (see, e.g., Silverman 1986), but we impose 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 a policy that matches $\boldsymbol{\alpha}_t$ with $\hat{\boldsymbol{\alpha}}_t^{\delta}$ in each stage, simultaneously making $\hat{\boldsymbol{\alpha}}_t^{\delta}$ approach $\boldsymbol{\alpha}^{\delta *}$ as $t \to \infty$. The algorithm is summarized below, with n_0 being a parameter of the algorithm.

Algorithm 2 $\boldsymbol{\pi}(n_0)$

For each j, take n_0 samples and let $t = kn_0$
while $t \leq T$ **do**
Obtain $\hat{\boldsymbol{\alpha}}_t^{\delta}$ from (19)
Take a sample from system $\pi_{t+1} = \text{argmax}_j\{\hat{\alpha}_{jt}^{\delta} - \alpha_{jt}\}$ and let $t = t+1$
end while

Theorem 2 (Asymptotic performance of Algorithm 2) Under assumption (F3), a policy by Algorithm 2 is consistent and its asymptotic performance is characterized as

$$\lim_{T\to\infty} \mathsf{E}(\mathscr{R}_T^{\boldsymbol{\pi}}) = \frac{\rho(\boldsymbol{\alpha}^{\delta *})}{\rho^*}. \tag{21}$$

Also, if $\lim_{\delta\to 0} f_j(\xi_j^p) \geq c$ for some exogenous $c > 0$, then $\rho(\boldsymbol{\alpha}^{\delta *})/\rho^* \to 1$ as $\delta \to 0$.

To rephrase (21), the Algorithm 2 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 $\delta \to 0$. In Theorem 2, the condition that $f_j(\xi_j^p)$ is bounded below is a mild restriction that is introduced to exclude some improbable situations. We provide a simple example to better understand this condition.

Example 1 (Exponential systems) Consider two exponential systems with means $(\mu, \mu - \delta)$, for which $(f_1(\xi_1^p), f_2(\xi_2^p)) \to ((1-p)/\mu, (1-p)/\mu)$ as $\delta \to 0$, satisfying the condition in Theorem 2. As an extreme counterexample, consider two exponential systems with means $(\delta + 1/\delta, 1/\delta)$, for which $(\xi_1^p, \xi_2^p) = (-(\delta + 1/\delta)\log(1-p), -(1/\delta)\log(1-p))$. In this case, $(f_1(\xi_1^p), f_2(\xi_2^p)) \to (0,0)$ as $\delta \to 0$, violating the condition in Theorem 2.

Remark 3 (Bias-variance tradeoff) The major advantage of Algorithm 2 over Algorithm 1 is that the former only requires *local* information on densities at particular points, while the latter requires the full knowledge of the distribution functions on certain regions. As a result, $\boldsymbol{\alpha}_t$ under Algorithm 1 is quite volatile around $\boldsymbol{\alpha}^*$. However, Algorithm 2 exhibit smaller variance of $\boldsymbol{\alpha}_t$ around $\boldsymbol{\alpha}^{\delta *}$, while introducing a small bias $\boldsymbol{\alpha}^{\delta *} - \boldsymbol{\alpha}^*$ which is in an order of $o(\delta^2)$. To be more precise, 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

$$\rho(\alpha_1^*) - \rho(\alpha_{1t}) = -\frac{\rho''(\alpha_1^*)}{2}(\alpha_{1t} - \alpha_1^*)^2 + o((\alpha_{1t} - \alpha_1^*)^2), \tag{22}$$

where $\rho''(\alpha_1^*)$ is the second derivative of $\rho(\cdot)$ at α_1^*. Hence, the expected optimality gap, $\mathsf{E}(\rho(\alpha_1^*) - \rho(\alpha_{1t}))$, can be largely explained by the mean squared error (MSE), $\mathsf{E}(\alpha_{1t} - \alpha_1^*)^2$, which can be further decomposed as

$$\mathsf{E}(\alpha_{1t} - \alpha_1^*)^2 = \mathsf{E}(\alpha_{1t} - \mathsf{E}(\alpha_{1t}))^2 + (\mathsf{E}(\alpha_{1t}) - \alpha_1^*)^2. \tag{23}$$

The first term represents the variance component and the second represents the bias component of the MSE. As illustrated in Figure 1, most part of the MSE is due to the variance under the policy by Algorithm 1. On the other hand, Algorithm 2 significantly reduces the variance, while introducing a small bias that decreases with δ. Further, if let T_δ denote the sampling budget where the MSE's under the two algorithms are equal, then it can be seen that Algorithm 2 performs better for $T \ll T_\delta$. The figure also suggests that T_δ increases as $\delta \to 0$.

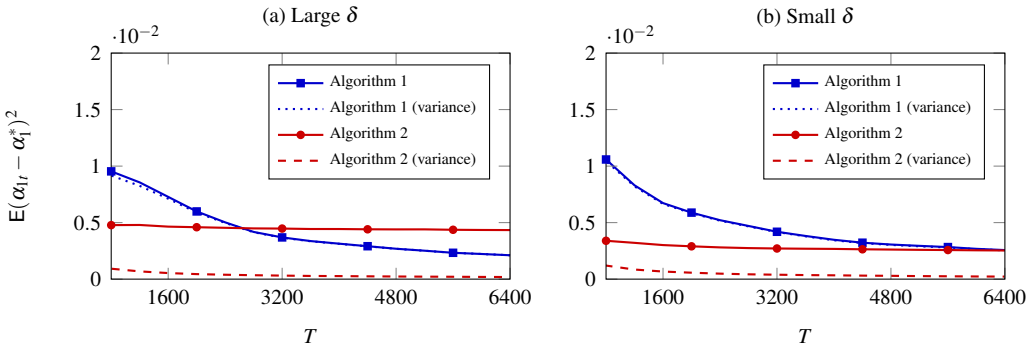

Figure 1: Bias-variance tradeoff. The mean squared error (MSE), $\mathsf{E}(\alpha_{1t} - \alpha_1^*)^2$, is estimated via simulation for the two policies by Algorithms 1-2. The system configurations are (a) two normal systems with means $(0,0)$ and standard deviations $(1,3)$ and (b) two normal systems with the same means and standard deviations $(1,2)$. The solid lines are the MSE's and the dotted and the dashed lines are the variance components for Algorithms 1 and 2, respectively.

3.3 Alternative Approach for Discrete Distributions

We now assume that the distributions are discrete and, without loss of generality, that they have supports in the set of nonnegative integers. Define

$$\begin{aligned} h_{j1} &= \frac{F_j(\xi_1^p) - F_j(\xi_j^p)}{\xi_1^p - \xi_j^p} \\ h_{1j} &= \frac{F_1(\xi_1^p) - F_1(\xi_j^p)}{\xi_1^p - \xi_j^p} \end{aligned} \tag{24}$$

for each $j \neq 1$ and let $\boldsymbol{h} = \{(h_{j1}, h_{1j}) \mid j \neq 1\}$. Also define $\varepsilon \in (0,1)$ as a constant such that, for each $j \neq 1$ and $x \in [\xi_j^p, \xi_1^p]$,

$$\begin{aligned} 1-\varepsilon \leq \frac{F_j(x)}{p + (x - \xi_j^p)h_{j1}}, \frac{1 - F_j(x)}{1 - p - (x - \xi_j^p)h_{j1}} \leq 1+\varepsilon \\ 1-\varepsilon \leq \frac{F_1(x)}{p + (x - \xi_1^p)h_{1j}}, \frac{1 - F_1(x)}{1 - p - (x - \xi_1^p)h_{1j}} \leq 1+\varepsilon. \end{aligned} \tag{25}$$

The constant ε represents (multiplicative) errors when $F_j(x)$ is approximated by a linear function with its slope characterized by h_{j1} or h_{1j}. We consider a set of discrete distributions that satisfies the following condition, which ensures that h_{j1} and h_{1j} defined in (24) 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 $x \in \mathscr{H}_j$

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

$$G_j^{\delta,\varepsilon}(\boldsymbol{\alpha}) = \frac{(\xi_1^p - \xi_j^p)^2}{2p(1-p)\left(1/(\alpha_1 h_{1j}^2) + 1/(\alpha_j h_{j1}^2(u))\right)}, \tag{26}$$

which has a similar structure as that of $\rho^{\delta}(\cdot)$ in (16) in the continuous case, except that $f_j(\xi_j^p)$ in $\rho^{\delta}(\cdot)$ is replaced with h_{j1} or h_{1j}. To provide some intuition behind the definition of $\rho^{\delta,\varepsilon}(\boldsymbol{\alpha})$, recall that the asymptotic variance of $\hat{\xi}_{jt}^p$ is inversely proportional to $f_j(\xi_j^p)$ in the continuous case. In other words, the

Algorithm 3 $\boldsymbol{\pi}(n_0)$

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

while $t \leq T$ **do**

If $\hat{\xi}^p_{bt} = \hat{\xi}^p_{jt}$, $\hat{h}_{jbt} = 0$, or $\hat{h}_{bjt} = 0$ for some $j \neq b$, then take a sample from system $\pi_{t+1} = \text{argmin}_{i=j,b}\{\alpha_{it}\}$. Otherwise, solve for $\hat{\boldsymbol{\alpha}}^{\delta,\varepsilon}_t = \text{argmax}_{\boldsymbol{\alpha}\in\Delta}\{\hat{\rho}^{\delta,\varepsilon}_t(\boldsymbol{\alpha})\}$ and let

$$\pi_{t+1} = \underset{j}{\text{argmax}}\{\hat{\alpha}^{\delta,\varepsilon}_{jt} - \alpha_{jt}\} \tag{31}$$

Let $t = t+1$

end while

asymptotic variance is low (high) when samples from system j are more (less) likely lie around ξ^p_j. Note that h_{j1} or h_{1j} captures the likelihood of samples lying around ξ^p_j, and hence, plays a role of $f_j(\xi^p_j)$ in the continuous case. The following proposition validates this intuition in a precise mathematical manner. We use the following notation: $\bar{h} = \max_{j\neq 1}\{\max(h_{j1}, h_{1j})\}$, $r = \max_{j\neq 1}\{\xi^p_1 - \xi^p_j\}/\min_{j\neq 1}\{\xi^p_1 - \xi^p_j\}$, $\theta = r\bar{h}\delta$, and $u : \mathcal{R}_+ \to \mathcal{R}_+$ is a non-decreasing function of θ defined as

$$u(\theta) = \frac{2\theta p(1-p)(\theta^3 + 3\theta p(1-p) + p(1-p))}{3(p-\theta)^3(1-p-\theta)^3} \tag{27}$$

Proposition 5 (Characteristic of $\rho^{\delta,\varepsilon}$) Under assumptions (F1)-(F2) and (F3'), for any $\boldsymbol{\alpha} \in \Delta$

$$(1-\varepsilon)(1-u(\theta)) \leq \frac{\rho(\boldsymbol{\alpha})}{\rho^{\delta,\varepsilon}(\boldsymbol{\alpha})} \leq (1+\varepsilon)(1+u(\theta)), \tag{28}$$

for θ sufficiently small so that $u(\theta) < 1$.

Proposition 5 validates approximation of $\rho(\boldsymbol{\alpha})$ by $\rho^{\delta,\varepsilon}(\boldsymbol{\alpha})$ for small δ (equivalently, small θ) and ε. We now present an alternative algorithm that iteratively maximizes $\rho^{\delta,\varepsilon}(\cdot)$ from history of sample observations. Let $\hat{\boldsymbol{\alpha}}^{\delta,\varepsilon}_t$ be the estimator of $\boldsymbol{\alpha}^{\delta,\varepsilon *}$ in stage t; formally,

$$\hat{\boldsymbol{\alpha}}^{\delta,\varepsilon}_t \in \underset{\boldsymbol{\alpha}\in\Delta}{\text{argmax}} \left\{ \min_{j\neq b} \frac{(\hat{\xi}^p_{bt} - \hat{\xi}^p_{jt})^2}{2p(1-p)\left(1/(\alpha_b \hat{h}^2_{bjt}) + 1/(\alpha_j \hat{h}^2_{jbt})\right)} \right\}, \tag{29}$$

where $b = \text{argmax}_j\{\hat{\xi}^p_{jt}\}$ and for each j,

$$\begin{aligned} \hat{h}_{jbt} &= \frac{\hat{F}_{jt}(\hat{\xi}^p_{bt}) - \hat{F}_{jt}(\hat{\xi}^p_{jt})}{\hat{\xi}^p_{bt} - \hat{\xi}^p_{jt}} \\ \hat{h}_{bjt} &= \frac{\hat{F}_{bt}(\hat{\xi}^p_{bt}) - \hat{F}_{bt}(\hat{\xi}^p_{jt})}{\hat{\xi}^p_{bt} - \hat{\xi}^p_{jt}} \end{aligned} \tag{30}$$

The algorithm for the discrete case is summarized below, with n_0 being a parameter.

We note that the event $\{\hat{\xi}^p_{bt} = \hat{\xi}^p_{jt}\}$ or the event $\{\hat{h}_{jbt} = 0 \text{ or } \hat{h}_{bjt} = 0\}$ for some $j \neq b$ can occur with positive probability, in which case $\hat{\boldsymbol{\alpha}}^{\delta,\varepsilon}_t$ in (29) may not be well defined. When these cases occur, Algorithm 3 takes a sample from system j or b, whichever is sampled less.

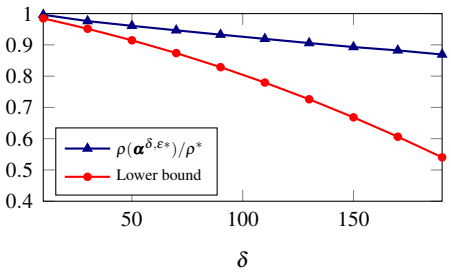

Figure 2: Asymptotic performance of Algorithm 3 and its lower bound defined in (32) for three discrete uniform systems with quantiles $(0,-\delta,-2\delta)$ for $p=0.1$. The lengths of supports are 10^4, or equivalently, $h_{j1}=h_{1j}=10^{-4}$ for each $j\neq 1$.

Theorem 3 (Asymptotic performance of Algorithm 3) Under assumptions (F1)-(F2) and (F3'), a policy by Algorithm 3 is consistent and its asymptotic performance is characterized as

$$\lim_{T\to\infty}\mathsf{E}(\mathscr{R}_T^{\boldsymbol{\pi}})=\frac{\rho(\boldsymbol{\alpha}^{\delta,\varepsilon*})}{\rho^*}\geq\frac{(1-u(\theta))(1-\varepsilon)}{(1+u(\theta))(1+\varepsilon)} \tag{32}$$

for sufficiently small θ such that $u(\theta)<1$.

Remark 4 (Tightness of the lower bound in Theorem 3) We note that $u(\theta)$ is a continuous function of $\theta\geq 0$ with $u(0)=0$. Hence, when θ and ε are close to 0, the bound in (32) is tight, uniformly across *all* discrete distributions satisfying mild assumptions (F1), (F2), and (F3'). In Figure 2 we provide a numerical example to illustrate the tightness of the bound for the case with discrete uniform distributions. One can observe that the bound becomes loose as δ increases, or equivalently, as $u(\theta)$ approaches 1.

4 COMPARISON WITH BENCHMARK POLICIES

4.1 Benchmark Sampling Policies

We compare the proposed policies with a couple of other policies: the equal allocation and a simple heuristic policy based on Hoeffding's inequality; see details below.

- *Equal allocation (EA).* This is a simple static policy, where all systems are equally sampled, i.e., $\pi_t=\text{argmin}_j\{N_{jt}\}$, breaking ties by selecting the system with the smallest index.
- *Heuristic based on Hoeffding's inequality (HH).* This policy takes two parameters: the number of initial samples n_0 and $\beta\in(0,1)$. The policy is summarized in Algorithm 4.

To provide some intuition behind the HH policy, denote ξ^p as the pth quantile for distribution $F(\cdot)$ and $\hat{\xi}_n^p$ as the sample quantile from n independent samples. From Hoeffding's inequality, observe that

$$\mathsf{P}(\hat{\xi}_n^{p+\zeta}<\xi^p)=\mathsf{P}\left(\sum_{i=1}^n(\mathbf{I}\{X_i\leq\xi^p\}-p)\geq\zeta n\right)\leq e^{-2n\zeta^2}. \tag{34}$$

In other words, the population quantile ξ^p is less than $\hat{\xi}_n^{p+\zeta}$ with probability greater than or equal to $1-\exp(-2n\zeta^2)$. Note that, in Algorithm 4, a confidence interval is characterized by the parameter β; the upper bound of the confidence interval for system $j\neq b$ is v_t, while the lower bound for system b is v_t. The significance level of each confidence interval can be bounded by a function of $N_{jt}\zeta_{jt}^2$ as seen from the Hoeffding's inequality (34). This implies that the algorithm is designed to take a sample from the system with the least significance level in each stage.

Algorithm 4 $\boldsymbol{\pi}(n_0, \beta)$

For each j, take n_0 samples and let $t = kn_0$
while $t \leq T$ **do**
Let $b = \text{argmax}_j\{\hat{\xi}^p_{jt}\}$ and $b' = \text{argmax}_{j\neq b}\{\hat{\xi}^p_{jt}\}$
Fix $v_t = \beta\hat{\xi}^p_{bt} + (1-\beta)\hat{\xi}^p_{b't}$
Define $\{\zeta_{jt}\}_{j=1}^k$ as follows

$$\zeta_{jt} = \begin{cases} p - \hat{F}_{bt}(v_t) & \text{for } j = b \\ \hat{F}_{jt}(v_t) - p & \text{for } j \neq b \end{cases} \tag{33}$$

Take a sample from system $\pi_{t+1} = \text{argmin}_j\{N_{jt}\zeta^2_{jt}\}$ and let $t = t+1$
end while

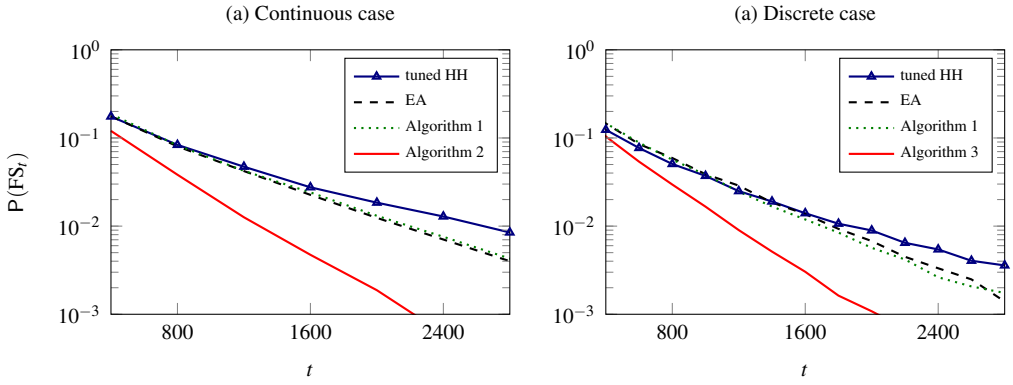

Figure 3: Probability of false selection in continuous and discrete cases. The system configurations are (a) normal distributions with means $(0,0,0,0)$ and standard deviations $(1,1.2,1.4,1.6)$ with $p = 0.05$ and (b) Poisson distributions with parameters $(1000,990,980,970)$ with $p = 0.5$. For all algorithms, $n_0 = 0.02T$. For HH algorithm, the parameter β is tuned to be 0.5.

4.2 Numerical Experiments

We perform numerical tests for continuous and discrete cases; we take normal and Poisson distributions as representative examples of the two. $\mathsf{P}(\text{FS}_t)$ is estimated by counting the number of false selections out of m simulation trials, which is chosen so that:

$$\sqrt{\frac{P_t(1-P_t)}{m}} \leq \frac{P_t}{10}, \tag{35}$$

where P_t is the order of magnitude of $\mathsf{P}(\text{FS}_t)$. This implies standard errors for the estimates of $\mathsf{P}(\text{FS}_t)$ are at least ten times smaller than the value of $\mathsf{P}(\text{FS}_t)$ so that we have sufficiently high confidence that the results are not driven by simulation error. For all policies we let $n_0 = 0.02T$; this is by no means an optimal choice for any specific policy but the same qualitative results hold for different values of n_0.

On the left panel of Figure 3, it can be seen that Algorithm 2 significantly improves the performance of Algorithm 1 in terms of $\mathsf{P}(\text{FS}_t)$. Note that Algorithm 1 performs no better than the equal allocation policy. This provide a crucial insight for practitioners who need to choose an algorithm most suitable for problem instances they face: When only a small budget is given and the difference between best and non-best systems are small enough, as is the case of Figure 3a, we recommend to use Algorithm 2. Algorithm 1 would be appropriate only when the sampling budget is extraordinarily large, which makes it less attractive in most applications. These conclusions are consistent with our theoretical results and Remark 3.

Further, although the parameter β is tuned, we observe that the HH policy does not perform well compared to the equal allocation. This result is expected, taking into account the fact that it is based on the Hoeffding's inequality, (34), which does not precisely capture the behavior of $\mathsf{P}(\mathrm{FS}_t)$. This in turn suggests that any policy without close association with the rate function $\rho(\cdot)$ can perform arbitrarily poorly. Essentially the same arguments hold for the discrete case in the right panel of Figure 3, but we highlight that Algorithm 3 significantly outperforms Algorithm 1.

REFERENCES

Avriel, M. 2003. *Nonlinear Programming: Analysis and Methods*. Courier Corporation.

Batur, D., and F. Choobineh. 2010. "A Quantile-Based Approach to System Selection". *European Journal of Operations Research* 202:764–772.

Bekki, J. M., J. W. Fowler, G. T. Mackulak, and B. L. Nelson. 2007. "Using Quantiles in Ranking and Selection Procedures". In *Proceedings of the 2007 Winter Simulation Conference*, edited by S. G. Henderson, B. Biller, M.-H. Hsieh, J. D. Tew, and R. R. Barton, 1722–1728. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Dembo, A., and O. Zeitouni. 2009. *Large Deviations Techniques and Applications*. New York: Springer Science & Business Media.

Kim, S. H., and B. L. Nelson. 2006. "Selecting the Best System". In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson, 501–534. Boston: Elsevier.

Lee, S., and B. L. Nelson. 2014. "Bootstrap Ranking & Selection Revisited". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 3857–3868. Piscataway, New Jersey.

Pasupathy, R., R. Szechtman, and E. Yücesan. 2010. "Selecting Small Quantiles". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 2762–2770. Piscataway, New Jersey.

Rinott, Y. 1978. "On Two-Stage Selection Procedures and Related Probability Inequalities". *Communications in Statistics* 7:799–811.

Serfling, R. J. 2009. *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons.

Shin, D., M. Broadie, and A. Zeevi. 2016. "Tractable Dynamic Sampling Strategies for Quantile-based Ordinal Optimization". *Columbia Business School Working Paper*.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. New York: CRC press.

AUTHOR BIOGRAPHIES

DONGWOOK SHIN is a PhD candidate at the Graduate School of Business, Columbia University. His research centers on optimal learning, at the interface between machine learning and sequential decision making under uncertainty. His email address is dshin17@gsb.columbia.edu.

MARK BROADIE is the Carson Family Professor of Business at the Graduate School of Business, Columbia University. His research interests include the pricing of derivative securities, risk management and, more generally, quantitative methods for decision-making under uncertainty. His email address is mnb2@columbia.edu

ASSAF ZEEVI is the Vice Dean for Research and Henry Kravis Professor of Business at the Graduate School of Business at Columbia University. He is particularly interested in the areas of stochastic modeling and statistics, and their synergistic application to problems arising in service operations, revenue management, and financial services. His email address is assaf@gsb.columbia.edu