Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

ORDINAL OPTIMIZATION WITH GENERALIZED LINEAR MODEL

Dohyun Ahn

Department of SEEM
The Chinese University of Hong Kong
Shatin, New Territories, HONG KONG

Dongwook Shin

Department of ISOM
HKUST Business School
Clear Water Bay, Kowloon, HONG KONG

ABSTRACT

Given a number of stochastic systems, we consider an ordinal optimization problem to find an optimal allocation of a finite sampling budget, which maximizes the likelihood of selecting the "best" system, where the "best" is defined as the one with the highest mean. The statistical characteristics of each system are described by the generalized linear model, where unknown parameters are estimated using maximum likelihood estimation. To formulate the problem in a tractable form, we use the large deviations theory to characterize the structural properties of the optimal allocation. Further, motivated by Euclidean information theory, we obtain an approximate solution for the optimal allocation, which is leveraged to design a sampling strategy that is near-optimal in a suitable asymptotic sense. The proposed sampling strategy is computationally tractable, and we show via numerical testing that it performs competitively even in the presence of model misspecification.

1 INTRODUCTION

Given a finite number of systems, we are concerned with the problem of dynamically learning the unknown statistical characteristics of the systems to ultimately select the one with the highest mean. This is an instance of ordinal optimization where a given sampling budget can be efficiently allocated to minimize the probability of mistakenly selecting a suboptimal system; see, e.g., Ho et al. (2008). An important extension of the ordinal optimization problem is when the underlying probability distribution for each system is characterized by known features, as well as an unknown parameter that is common across systems. When a system is sampled, it is possible to learn about the unknown parameter, which enables us to understand the statistical characteristics of all systems.

The aforementioned problem arises in many practical settings. A representative example is the adaptive design of clinical trials (Martin 2010). There are three salient aspects of this problem: firstly, the primary goal is to select the best drug after the trial, and the performance of drugs during the trial is of secondary importance; secondly, drugs can be described by a vector of features, e.g., the presence of particular molecules and the structural properties of chemical compounds; and lastly, the functional relationship between the drug's features and the outcome can be discrete (possibly non-numeric) and nonlinear with respect to the features. This paper focuses on the settings of ordinal optimization described above: the optimal allocation of sampling budgets that consider *multi-dimensional features* of systems, where the performance of each system is characterized by the *generalized linear model* (McCullagh and Nelder 1989), to minimize the *probability of falsely selecting the system with sub-optimal performance*.

Unfortunately, as is well documented in the literature of ordinal optimization, the probability of false selection is not an analytically tractable objective. This is especially so in our problem with the generalized linear model, where the selection is based on the maximum likelihood estimator of the unknown parameter, because its exact probability distribution does not admit a tractable form. This situation can be drastically simplified if one *assumes* that the estimated parameter is normally distributed. Although this approach

might help develop practically implementable allocation strategies, the actual performance with respect to the probability of false selection would be subject to the normality assumption, which often is not exactly true in general applications; although true, it is difficult to verify in practice. The main contribution of our work is to formulate the probability of false selection in a tractable form without making such an assumption on the maximum likelihood estimation.

To that end, our departure point is the large deviations framework: if the underlying probability distributions belong to an exponential family, and if the sampling budget is sufficiently large, then the *rate function* of the maximum likelihood estimator can be expressed using the Kullback-Leibler (KL) divergence, and the the optimal allocation of the sampling budget can be obtained by maximizing the KL divergence. This maximization problem, however, rarely admits an analytical solution nor provides structural insights into the optimal allocation of the sampling budget. Using ideas from Euclidean information theory, we show that, under mild conditions, the probability of false selection (on a logarithmic scale) can be closely approximated in a closed form, structured around the statistical characteristics of the best and the second-best systems.

In spite of the ever-growing importance of the generalized linear model in a wide range of applications, most of the existing studies focus on the cases with independent systems, where sampling from one system does not reveal any information about other systems. In the context with independent systems, the existing literature can be roughly categorized into fixed budget and fixed confidence settings; the goal of the former setting is to minimize the probability of false selection given a sampling budget, whereas in the latter setting, the objective is to devise a sampling procedure that satisfies the desired guarantee on the probability of false selection by taking as few samples as possible; we only mention Audibert et al. (2010) and Even-Dar et al. (2006) for fixed budget and fixed confidence settings, respectively.

As alluded to earlier, this paper focuses on parametric systems, rather than on independent systems. In fixed confidence settings, there has been considerable work in the situation where the mean system performance is described by a linear function of features. Soare et al. (2014) propose a static algorithm in the situation where the system performance is described by a linear function of features, and Xu et al. (2018) develop an adaptive algorithm that can make substantial improvement over static ones. Recently, Kazerouni and Wein (2019) extend the results of Xu et al. (2018) to the case with the generalized linear model. To our knowledge, our work is the first paper to consider the generalized linear model in a fixed budget setting.

The ordinal optimization is related to the multi-armed bandit (MAB) problem, insofar as the underlying distributions of independent systems (arms) are initially unknown but can be learned via sequential sampling; see, e.g., Lai and Robbins (1985) and the reference therein. An important extension of the MAB problem is the linear-bandit problem, where the mean of the underlying distribution is characterized by a linear function of features; see Abbasi-Yadkori et al. (2011) and Li et al. (2010). Although our paper shares an important common theme with the linear-bandit problem in that it, too, highlights a trade-off between learning and optimizing, it is worthy noting that the goal of the MAB problems is to maximize the cumulative rewards, and the reward-maximizing algorithms are not well-suited for the ordinal optimization problem.

The remainder of the paper is organized as follows. In Section 2, we formulate the problem using large deviations theory. In Section 3, we propose an approximation based on the Euclidean information theory and provide main theoretical results. In Section 4, we conduct numerical experiments and discuss the results. Section 5 concludes the paper.

2 PROBLEM FORMULATION

This paper focuses on the problem of minimizing the probability of false selection in a large deviations framework (Dembo and Zeitouni 2009). In particular, we first introduce model preliminaries and elementary principles of large deviations in Sections 2.1 and 2.2, respectively, and then formulate the problem in Section 2.3.

2.1 Model Preliminaries

We consider k stochastic systems indexed by $i = 1, \ldots, k$, where each system is characterized by a feature vector $\mathbf{x}_i \in \mathbb{R}^d$ which is known to the decision maker and summarizes the available information of the system. Denote by $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_k) \in \mathbb{R}^{d \times k}$ the feature matrix. The relationship between the feature vector and the performance of each system is described by a generalized linear model. In particular, the performance of system i, denoted by a random variable $Y_i \in \mathbb{R}$, is governed by a distribution that belongs to an exponential family with a canonical parametrization; see, e.g., McCullagh and Nelder (1989). Specifically, the density function associated with system i is parametrized by $\theta \in \mathbb{R}^d$ and can be represented as

$$f_i(y;\theta) = h_i(y) \exp\left\{\eta_i(\theta' \mathbf{x}_i) y - A_i(\eta_i(\theta' \mathbf{x}_i))\right\}, \tag{1}$$

where $h_i(\cdot)$, $\eta_i(\cdot)$, and $A_i(\cdot)$ are known functions, whereas the parameter θ is unknown. The mean of the distribution is denoted by $\mu_i(\theta)$, which depends on θ through a strictly increasing link function $g \in \mathscr{C}^2$ as $\mu_i(\theta) = g(\theta' \mathbf{x}_i)$, and the standard deviation of the distribution is denoted by $\sigma_i(\theta)$. Without loss of generality, we assume that system 1 has the largest mean; that is, $\mu_1(\theta) > \max_{i \neq 1} \mu_i(\theta)$, or equivalently, $\theta' \mathbf{x}_1 > \max_{i \neq 1} \theta' \mathbf{x}_i$. For simple exposition, we often write $\mu_i = \mu_i(\theta)$ and $\sigma_i = \sigma_i(\theta)$ for each i.

A decision maker is given a *fixed* sampling budget T, which means T independent samples can be drawn from the k systems. A sampling policy π is defined as a sequence of random variables, π_1, π_2,..., taking values in the index set $\{1, \ldots, k\}$; the event $\{\pi_t = i\}$ means a sample from system i is taken at stage t. Define Y_{it}, $t = 1, \ldots, T$, as a random sample from system i in stage t, and let $\mathscr{F}_t$ be the σ-field generated by the samples and sampling decisions taken up to stage t (i.e., $\{(\pi_s, Y_{i\pi_s})\}_{s=1}^t$, with the convention that $\mathscr{F}_0$ is the nominal sigma-algebra associated with underlying probability space. 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 stage (i.e., $\{\pi_t = i\} \in \mathscr{F}_{t-1}$ for $i = 1, \ldots, k$ and $t = 1, \ldots, T$).

We denote by $n_{it}(\pi)$ the number of samples from system i up to stage t and let $\alpha_{it}(\pi)$ be the sampling rate at stage t; that is, $\alpha_{it}(\pi) = n_{it}(\pi)/t$. The maximum likelihood estimator (MLE) of θ using sampling observations up to stage t is denoted by $\theta_t(\pi)$. For brevity, the argument π may be dropped when it is clear from the context.

2.2 Large Deviations Preliminaries

Define $\mathrm{FS}_i = \{\tilde{\theta} \in \mathbb{R}^d \mid \tilde{\theta}' \mathbf{x}_1 < \tilde{\theta}' \mathbf{x}_i\}$ for each $i \neq 1$ and let $\mathrm{FS} = \bigcup_{i \neq 1} \mathrm{FS}_i$. The probability of false selection, denoted by $\mathsf{P}(\theta_T(\pi) \in \mathrm{FS})$, is a widely used criterion for the efficiency of a sampling policy, but it is not analytically tractable to precisely evaluate $\mathsf{P}(\theta_T(\pi) \in \mathrm{FS})$. Hence, we focus on the asymptotic regime where the sampling budget goes to infinity, in which case $\mathsf{P}(\theta_T(\pi) \in \mathrm{FS})$ can be expressed in a closed form using the large deviations principle.

Consider a fixed vector $\alpha = (\alpha_1, \ldots, \alpha_k) \in \Delta^{k-1}$, where $\Delta^{k-1} = \{(\alpha_1, \ldots, \alpha_k) | \sum_{i=1}^k \alpha_i = 1,\ \alpha_i \geq 0\}$, and we let $\pi(\alpha) = (\pi_1(\alpha), \pi_2(\alpha), \ldots) \in \mathscr{F}_0$ be a deterministic allocation rule such that $\alpha_i T$ samples out of the total T samples are allocated to system i, ignoring minor technicalities associated with $\alpha_i T$ not being an integer. In this section, we fix $\pi = \pi(\alpha)$ and omit π in function arguments for simple exposition. The probability of false selection can be bounded as

$$\max_{i \neq 1} \mathsf{P}\left(\theta_T \in \mathrm{FS}_i\right) \leq \mathsf{P}\left(\theta_T \in \mathrm{FS}\right) \leq (k-1) \max_{i \neq 1} \mathsf{P}\left(\theta_T \in \mathrm{FS}_i\right). \tag{2}$$

Joutard (2004) shows that the maximum-likelihood estimator θ_T satisfies the large deviations principle in the following form:

$$\frac{1}{T} \log \mathsf{P}\left(\theta_T \in \mathrm{FS}_i\right) \to -\rho_i(\alpha) \text{ as } T \to \infty.$$

Further, the function $\rho_i(\alpha)$ satisfies

$$\rho_i(\alpha) = \inf\left\{I_\theta(\tilde{\theta}; \alpha) \;\middle|\; \tilde{\theta} \in \mathrm{FS}_i\right\}, \tag{3}$$

where

$$I_\theta(\tilde{\theta};\alpha) = -\inf_{\lambda\in\mathbb{R}^d}\left\{\lim_{T\to\infty}\frac{1}{T}\log \mathsf{E}\left[\exp\left(\sum_{t=1}^{T}\lambda'\nabla_\theta \log f_{\pi_t}(Y_{\pi_t};\theta)\right)\right]\right\}.$$

Hence, from (2) it can be seen that

$$\frac{1}{T}\log \mathsf{P}\left(\theta_T \in \mathrm{FS}\right) \to -\rho(\alpha) \text{ as } T\to\infty$$

where $\rho(\alpha) = \min_{i\neq 1}\{\rho_i(\alpha)\}$ which is referred to as the rate function for the probability of false selection. We let $\alpha^* = \mathrm{argmax}_{\alpha\in\Delta^{k-1}}\{\rho(\alpha\}$ and $\rho^* = \rho(\alpha^*)$.

2.3 Problem Formulation

The observations in Section 2.2 suggest that, if the sampling budget T is allocated in proportion to a vector $\alpha\in\Delta^{k-1}$, then $\mathsf{P}(\theta_T\in\mathrm{FS})$ behaves roughly similar to $\exp(-\rho(\alpha)T)$ for large values of T. Hence, $\rho(\cdot)$ is an appropriate measure of efficiency associated with the probability of false selection. Define the relative efficiency $R_T(\pi)$ for a policy $\pi\in\Pi$ to be

$$R_T(\pi) = \frac{\rho(\alpha_T(\pi))}{\rho^*}.$$

A policy π is called efficient if $R_T(\pi)$ is close to one. We aim at finding an allocation policy π that maximizes the relative efficiency given a sampling budget T:

$$\sup_{\pi\in\Pi}\mathsf{E}\left[R_T(\pi)\right]. \tag{4}$$

It is difficult, if not impossible, to find an exact solution to the stochastic dynamic programming problem (4) due to the curse of dimensionality. In addition, the objective function $R_T(\pi)$ is not known a priori because the parametric structure of the underlying distributions, as well as the parameter θ, is unknown. In this situation, one would need to explore the objective function while simultaneously maximizing it within a sampling budget given. We alternatively focus on *asymptotically optimal* sampling policies such that

$$\mathsf{E}[R_T(\pi)]\to 1 \text{ as } T\to\infty. \tag{5}$$

Since $R_T(\pi)\in[0,1]$ by definition, it can be seen from the bounded convergence theorem that (5) is achieved if $\alpha_T(\pi)\to\alpha^*$ in probability as $T\to\infty$. Hence, in the remaining part of the paper, we aim to design a sampling strategy that sequentially learns the unknown parameter θ using sample observations, while simultaneously making the allocation $\alpha_T(\pi)$ close to α^*.

Notational convention. It will be convenient to define $\kappa = (\kappa_1,\ldots,\kappa_k)$, where $\kappa_i = \theta'\mathbf{x}_i$, so that $\mu_i(\theta) = g(\kappa_i)$. Also, $u(x) = o(v(x))$ implies $|u(x)/v(x)|\to 0$ as $v(x)\to 0$. We may suppress α in function arguments if it is clear from the context. We use Newton's notation for differentiation, i.e., if u is a function of x, then

$$\dot{u} = \frac{\partial u}{\partial x} \quad \text{and} \quad \ddot{u} = \frac{\partial^2 u}{\partial x^2}.$$

All vectors are column vectors and we use $\mathbf{z}'$ to denote the transpose of a vector $\mathbf{z}$. The Euclidean norm of a vector $\mathbf{z}$ is denoted by $||\mathbf{z}||$.

3 EUCLIDEAN APPROXIMATION FOR THE RATE FUNCTION

In a general setting, the rate function $\rho(\cdot)$ for the probability of false selection does not admit a closed form, so one would need to solve a bi-level optimization problem to obtain $\alpha^* = \mathrm{argmax}_{\alpha\in\Delta^{k-1}}\{\rho(\alpha)\}$; the

inner layer to evaluate $\rho_i(\alpha)$ in (3) for each i and α and the outer layer to maximize $\rho(\alpha) = \min_{i \neq 1}\{\rho_i(\alpha)\}$ over $\alpha \in \Delta^{k-1}$. This may pose significant computational challenges, especially when the problem instance is large in terms of the parameter dimension (d) and the number of systems (k). This is particularly problematic in practical applications where the estimate of θ is sequentially updated according to sample observations, so that the bi-level problem needs to be solved repeatedly over the sampling horizon.

In order to address the aforementioned issues, we derive a proxy $\rho^{\mathrm{E}}(\alpha)$ for the rate function $\rho(\alpha)$ for each α, hereafter referred to as the *Euclidean approximation of the rate function*. As it turns out, this proxy is sufficiently tractable and serves as a close approximation under certain conditions. To this end, we first show that the function $I_\theta(\tilde{\theta};\alpha)$ is closely related to the Kullback-Leibler divergence (Kullback 1997).

Lemma 1 (An upper bound and its tightness) For any $\alpha \in \Delta^{k-1}$ and $\tilde{\theta} \in \mathbb{R}^d$,

$$I_\theta(\tilde{\theta};\alpha) \leq \bar{I}_\theta(\tilde{\theta};\alpha) := \sum_{i=1}^{k} \alpha_i D_{\mathrm{KL}}(f_i(\cdot;\tilde{\theta})||f_i(\cdot;\theta)),$$

where $D_{\mathrm{KL}}(f_i(\cdot;\tilde{\theta})||f_i(\cdot;\theta))$ is the Kullback-Leibler divergence between the two density functions parametrized by $\tilde{\theta}$ and θ, respectively. Further, the upper bound $\bar{I}_\theta(\tilde{\theta};\alpha)$ is tight when $\tilde{\theta}$ is close to θ; formally,

$$\bar{I}_\theta(\tilde{\theta};\alpha) - I_\theta(\tilde{\theta};\alpha) = o\left(||(\tilde{\theta}-\theta)'\mathbf{X}||^2\right) \quad \text{as} \quad (\tilde{\theta}-\theta)'\mathbf{X} \to 0.$$

The proof of the preceding lemma follows from the same logical steps in the proof of Theorem 3.2 of (Arcones 2006), which will be omitted. According to the preceding lemma, the function $I_\theta(\tilde{\theta};\alpha)$ can be closely approximated by its upper bound $\bar{I}_\theta(\tilde{\theta};\alpha)$ as long as $\tilde{\theta}$ differs by only a small amount from θ. If $\tilde{\theta}$ lies in the vicinity of θ, the following lemma establishes that the upper bound $\bar{I}_\theta(\tilde{\theta};\alpha)$, and hence $I_\theta(\tilde{\theta};\alpha)$, can be approximated by a quadratic function.

Lemma 2 (Quadratic approximation) For any $\alpha \in \Delta^{k-1}$ and $\tilde{\theta} \in \mathbb{R}^d$,

$$I_\theta(\tilde{\theta};\alpha) = I^{\mathrm{E}}_\theta(\tilde{\theta};\alpha) + o\left(||(\tilde{\theta}-\theta)'\mathbf{X}||^2\right),$$

where

$$I^{\mathrm{E}}_\theta(\tilde{\theta};\alpha) = \sum_{i=1}^{k} \frac{\alpha_i \left(\mu_i(\tilde{\theta}) - \mu_i(\theta)\right)^2}{2\sigma_i^2}. \tag{6}$$

We omit the proof of Lemma 2 since it is straightforward by the standard argument using a second order Taylor expansion: we refer interested readers to Chapter 2.6 of Kullback (1997). Lemmas 1 and 2 suggest that $I_\theta(\tilde{\theta};\alpha)$ can be approximated by $I^{\mathrm{E}}_\theta(\tilde{\theta};\alpha)$ if $\tilde{\theta}$ is sufficiently close to θ. Recall that $\rho_i(\alpha) = \inf\{I_\theta(\tilde{\theta};\alpha) \mid \tilde{\theta} \in \mathrm{FS}_i\}$. If the infimum is achieved near θ, then one may intuit that $\rho_i(\alpha)$ can be approximated by $\rho^{\mathrm{E}}_i(\alpha) := \inf\{I^{\mathrm{E}}_\theta(\tilde{\theta};\alpha) \mid \tilde{\theta} \in \mathrm{FS}_i\}$, which is formalized in the next proposition.

Proposition 1 (Euclidean approximation) For each $i \neq 1$,

$$\rho_i(\alpha) = \rho^{\mathrm{E}}_i(\alpha) + o(\delta_i^2),$$

where $\delta_i := \theta'(\mathbf{x}_1 - \mathbf{x}_i)$,

$$\rho^{\mathrm{E}}_i(\alpha) = \frac{(\theta'(\mathbf{x}_1 - \mathbf{x}_i))^2}{2(\mathbf{x}_1 - \mathbf{x}_i)'\Sigma(\theta;\alpha)(\mathbf{x}_1 - \mathbf{x}_i)}, \tag{7}$$

and

$$\Sigma(\theta;\alpha) := \left(\sum_{i=1}^{k} \frac{\alpha_i}{\sigma_i^2}(\dot{g}(\theta'\mathbf{x}_i))^2 \mathbf{x}_i\mathbf{x}_i'\right)^{-1}.$$

Proof of Proposition 1. Fix $j \neq 1$ and let θ^* be the minimizer of (3) for system j. Define

$$\bar{\theta} = \operatorname{argmin} \left\{ I_\theta^{\mathrm{E}}(\tilde{\theta}; \alpha) \;\middle|\; \tilde{\theta} \in \mathrm{FS}_j \right\}. \tag{8}$$

In this proof we interchangeably use $I_\theta(\tilde{\theta};\alpha)$ and $I_\kappa(\tilde{\kappa};\alpha)$ since the former depends on θ and $\tilde{\theta}$ only through $\kappa = \theta'\mathbf{X}$ and $\tilde{\kappa} = \tilde{\theta}'\mathbf{X}$. Also, define $I_\kappa^{\mathrm{E}}(\tilde{\kappa};\alpha) = \sum_{i=1}^k \alpha_i (g(\tilde{\kappa}_i) - g(\kappa_i))^2/(2\sigma_i^2)$. Then, it is equal to $I_\theta^{\mathrm{E}}(\tilde{\theta};\alpha)$ if $\kappa = \theta'\mathbf{X}$ and $\tilde{\kappa} = \tilde{\theta}'\mathbf{X}$. For simplicity of notation, we let $\kappa_i^* = (\theta^*)'\mathbf{x}_i$, $\bar{\kappa}_i = \bar{\theta}'\mathbf{x}_i$, $\mu_i^* = g(\kappa_i^*)$, and $\bar{\mu}_i = g(\bar{\kappa}_i)$. Also, we fix $\alpha \in \Delta^{k-1}$ and omit it in function arguments for clarity of exposition.

Step 1. We show that κ^* and $\bar{\kappa}$ are not too far from true κ. The Lagrangian function for (3) is $L(\tilde{\theta}, \lambda) = I_\theta(\tilde{\theta}) + \lambda \tilde{\theta}'(\mathbf{x}_1 - \mathbf{x}_j)$ for $\tilde{\theta} \in \mathbb{R}^d$. Then, from the first-order condition for θ^* and the chain rule,

$$\sum_{i=1}^k \alpha_i \mathbf{x}_i' \frac{\partial D_{\mathrm{KL}}(f_i(\cdot;\tilde{\kappa}_i) || f_i(\cdot;\kappa_i))}{\partial \tilde{\kappa}_i}\Big|_{\tilde{\kappa}_i = \kappa_i^*} = -\lambda(\mathbf{x}_1 - \mathbf{x}_j)'. \tag{9}$$

Using (1) and the fact that $A_i'(\eta_i(\kappa_i)) = g(\kappa_i)$, we have that

$$\frac{\partial D_{\mathrm{KL}}(f_i(\cdot;\tilde{\kappa}_i) || f_i(\cdot;\kappa_i))}{\partial \tilde{\kappa}_i}\Big|_{\tilde{\kappa}_i = \kappa_i^*} = \dot{g}(\kappa_i^*)\big(\eta_i(\kappa_i^*) - \eta_i(\kappa_i)\big) = \dot{g}(\kappa_i^*)\dot{\eta}_i(\hat{\kappa}_i^*)(\kappa_i^* - \kappa_i) \tag{10}$$

for some $\hat{\kappa}_i^*$ between κ_i and κ_i^*. Define the matrix $\hat{\Sigma}(\kappa^*) := \hat{\Sigma}(\kappa^*;\alpha) = (\sum_{i=1}^k \alpha_i \dot{g}(\kappa_i^*)\dot{\eta}_i(\hat{\kappa}_i^*)\mathbf{x}_i\mathbf{x}_i')^{-1}$, which is well-defined since $\dot{g}(\cdot), \dot{\eta}_i(\cdot) > 0$. From (9) and (10), and using the fact that $(\kappa_i^* - \kappa_i) = (\theta^* - \theta)'\mathbf{x}_i$, we deduce that

$$(\theta^* - \theta)' = -\lambda(\mathbf{x}_1 - \mathbf{x}_j)'\hat{\Sigma}(\kappa^*). \tag{11}$$

For each i, by multiplying $\mathbf{x}_i$ on both sides of (11), we obtain

$$\kappa_i^* - \kappa_i = -\lambda(\mathbf{x}_1 - \mathbf{x}_j)'\hat{\Sigma}(\kappa^*)\mathbf{x}_i. \tag{12}$$

If $\lambda = 0$, then the preceding equation implies that $\kappa_i^* = \kappa_i$ for each i, which is a contradiction. Hence, $\lambda > 0$ and $(\theta^*)'(\mathbf{x}_1 - \mathbf{x}_j) = \kappa_1^* - \kappa_j^* = 0$ by complementary slackness. Further, multiplying $(\mathbf{x}_1 - \mathbf{x}_j)$ on both sides of (11) gives

$$\lambda = \frac{\kappa_1 - \kappa_j}{(\mathbf{x}_1 - \mathbf{x}_j)'\hat{\Sigma}(\kappa^*)(\mathbf{x}_1 - \mathbf{x}_j)}. \tag{13}$$

Combining (13) into (12), the following holds for each i:

$$\kappa_i^* - \kappa_i = \frac{(\kappa_1 - \kappa_j)(\mathbf{x}_1 - \mathbf{x}_j)'\hat{\Sigma}(\kappa^*)\mathbf{x}_i}{(\mathbf{x}_1 - \mathbf{x}_j)'\hat{\Sigma}(\kappa^*)(\mathbf{x}_1 - \mathbf{x}_j)},$$

from which we establish that there exists a constant $B^* < \infty$ such that $|\kappa_i^* - \kappa_i| \leq B^*|\kappa_1 - \kappa_j|$ for each $i = 1, \ldots, k$. Using the same logical steps for (8), it can be easily seen that

$$\bar{\kappa}_i - \kappa_i = \frac{(\kappa_1 - \kappa_j)(\mathbf{x}_1 - \mathbf{x}_j)'\check{\Sigma}(\bar{\kappa})\mathbf{x}_i}{(\mathbf{x}_1 - \mathbf{x}_j)'\check{\Sigma}(\bar{\kappa})(\mathbf{x}_1 - \mathbf{x}_j)}, \tag{14}$$

where $\check{\Sigma}(\bar{\kappa}) := \check{\Sigma}(\bar{\kappa};\alpha) = \left(\sum_{i=1}^k (\alpha_i/\sigma_i^2)\dot{g}(\bar{\kappa}_i)\dot{g}(\check{\kappa}_i)\mathbf{x}_i\mathbf{x}_i'\right)^{-1}$ and $\check{\kappa}_i$ is a constant between $\bar{\kappa}_i$ and κ_i satisfying $g(\bar{\kappa}_i) - g(\kappa_i) = \dot{g}(\check{\kappa}_i)(\bar{\kappa}_i - \kappa_i)$. Thus, $|\bar{\kappa}_i - \kappa_i| \leq \bar{B}|\kappa_1 - \kappa_j|$ for some constant $\bar{B} < \infty$.

Step 2. We claim that $\rho_j(\alpha) = I_\theta^{\mathrm{E}}(\bar{\theta}) + o(\delta_j^2)$. To this end, observe that $|I_\theta(\theta^*) - I_\theta^{\mathrm{E}}(\bar{\theta})| \leq |I_\theta(\theta^*) - I_\theta^{\mathrm{E}}(\theta^*)| + |I_\theta^{\mathrm{E}}(\theta^*) - I_\theta^{\mathrm{E}}(\bar{\theta})|$. From Lemma 2, we know that $|I_\theta(\theta^*) - I_\theta^{\mathrm{E}}(\theta^*)| = o(||\kappa^* - \kappa||^2)$. Further, using the observations from *Step 1* that $|\kappa_i^* - \kappa_i| \leq B^*|\kappa_1 - \kappa_j|$ for each $i = 1, \ldots, k$, we establish that

$$|I_\theta(\theta^*) - I_\theta^{\mathrm{E}}(\theta^*)| = o(\delta_j^2) \tag{15}$$

due to the Lipschitz continuity of $g(\cdot)$. Since $|I_\kappa(\tilde{\kappa}) - I^{\mathrm{E}}_\kappa(\tilde{\kappa})| = o(\delta_j^2)$ for any $\tilde{\kappa}$ and $I^{\mathrm{E}}_\kappa(\cdot)$ satisfies the conditions of Proposition 6.1 of Bonnans and Shapiro (1998), we deduce that the minimizers, κ^* and $\bar{\kappa}$, satisfy $||\kappa^* - \bar{\kappa}|| = o(\delta_j)$. From the definition of $I^{\mathrm{E}}_\theta(\cdot)$ in (6), one can write that

$$|I^{\mathrm{E}}_\theta(\theta^*) - I^{\mathrm{E}}_\theta(\bar{\theta})| = |I^{\mathrm{E}}_\kappa(\kappa^*) - I^{\mathrm{E}}_\kappa(\bar{\kappa})| \leq \sum_{i=1}^{k} \left| \frac{\alpha_i (g(\kappa_i^*) - g(\kappa_i))^2 - \alpha_i (g(\bar{\kappa}_i) - g(\kappa_i))^2}{2\sigma_j^2} \right| = o(\delta_j^2). \tag{16}$$

The last equality follows since $g(\cdot)$ is Lipschitz continuous and $\max(|\kappa_i^* - \kappa_i|, |\bar{\kappa}_i - \kappa_i|) \leq \max(B^*, \bar{B})|\kappa_1 - \kappa_j|$ from *Step 1*. Combining (15) and (16), we deduce that $\rho_j(\alpha) = I_\theta(\theta^*) = I^{\mathrm{E}}_\theta(\bar{\theta}) + o(\delta_j^2)$.

Step 3. We finally show that $I^{\mathrm{E}}_\theta(\bar{\theta}) = \rho^{\mathrm{E}}_j(\alpha) + o(\delta_j^2)$. Recall that $\mu_i(\bar{\theta}) - \mu_i(\theta) = g(\bar{\kappa}_i) - g(\kappa_i) = \dot{g}(\check{\kappa}_i)(\bar{\kappa}_i - \kappa_i)$. Hence, we can write that

$$\begin{aligned}
I^{\mathrm{E}}_\theta(\bar{\theta}) &= \sum_{i=1}^{k} \frac{\alpha_i (\mu_i(\bar{\theta}) - \mu_i(\theta))^2}{2\sigma_i^2} \\
&= \sum_{i=1}^{k} \frac{\alpha_i (\dot{g}(\check{\kappa}_i))^2 (\bar{\kappa}_i - \kappa_i)^2}{2\sigma_i^2} \\
&= \sum_{i=1}^{k} \frac{\alpha_i (\dot{g}(\check{\kappa}_i))^2 (\kappa_1 - \kappa_j)^2 ((\mathbf{x}_1 - \mathbf{x}_j)' \check{\Sigma}(\bar{\kappa}) \mathbf{x}_i)^2}{2\sigma_i^2 ((\mathbf{x}_1 - \mathbf{x}_j)' \check{\Sigma}(\bar{\kappa})(\mathbf{x}_1 - \mathbf{x}_j))^2} \\
&= \frac{(\kappa_1 - \kappa_j)^2}{2(\mathbf{x}_1 - \mathbf{x}_j)' \Sigma(\kappa)(\mathbf{x}_1 - \mathbf{x}_j)} + o(\delta_j^2),
\end{aligned}$$

where the third equality follows from (14) and the last equality holds since $\check{\Sigma}(\bar{\kappa}) = \Sigma(\kappa) + o(1)$ and $\dot{g}(\check{\kappa}_i) = \dot{g}(\kappa_i) + o(1)$ as $\delta_j \to 0$. Consequently, the desired result follows. □

Proposition 1 indicates that the Euclidean approximation $\rho^{\mathrm{E}}(\alpha) = \min_{j \neq 1}\{\rho^{\mathrm{E}}_j(\alpha)\}$ can be used as a proxy for the rate function for the probability of false selection. An important property of the approximation is that $\rho^{\mathrm{E}}(\cdot)$ is concave, and therefore, the first order conditions would be sufficient to characterize the maximizer $\alpha^{\mathrm{E}} := \mathrm{argmax}_{\alpha \in \Delta^{k-1}}\{\rho^{\mathrm{E}}(\alpha)\}$.

It is worthy noting that both functions, $\rho(\alpha)$ and $\rho^{\mathrm{E}}(\alpha)$, tend to zero for each $\alpha \in \Delta^{k-1}$ as δ approaches zero. Hence, Proposition 1 is not sufficient to guarantee the convergence of the maximizers, α^* and α^{E}, respectively, in the limit as $T \to \infty$. To strengthen Proposition 1, we consider a class $\mathbf{C}$ of system configurations, each characterized by $\theta \in \mathbb{R}^d$ such that $\sigma_i(\theta)/\dot{g}(\theta' \mathbf{x}_i) \in [\omega_{\min}, \omega_{\max}]$ for some positive constants $\omega_{\min} \leq \omega_{\max}$ for each i. This regularity condition rules out trivial cases where the probability of false selection approaches one (as $\sigma_i(\theta) \to 0$) or zero (as $\sigma_i(\theta) \to \infty$ or $\dot{g}(\theta' \mathbf{x}_i) \to 0$). For any system configuration in $\mathbf{C}$, it can be seen that $\rho^{\mathrm{E}}_i(\alpha)$ is quadratic with respect to δ_i for each $i \neq 1$, so there exists a constant $c > 0$ such that $\rho^{\mathrm{E}}(\alpha^{\mathrm{E}}) \geq \rho^{\mathrm{E}}(\alpha^{\mathrm{eq}}) \geq c\delta^2$, where $\alpha^{\mathrm{eq}} = (1/k, \ldots, 1/k)$. From Proposition 1 we deduce that $\rho(\alpha^*) \geq \tilde{c}\delta^2$ for some constant $\tilde{c} > 0$. Therefore, both $\rho(\alpha^*)$ and $\rho^{\mathrm{E}}(\alpha^{\mathrm{E}})$ are not dominated by $o(\delta^2)$, which enables us to establish the following theorem which suggests that the maximizers α^* and α^{E} coincide in the limit as the gap in means between the best and second-best systems approaches zero.

Theorem 1 (Validity of the Euclidean Approximation) Consider a class $\mathbf{C}$ of system configurations and define $\delta = \min_{i \neq 1}\{\theta'(\mathbf{x}_1 - \mathbf{x}_i)\}$ for each configuration in $\mathbf{C}$. If $g(\cdot)$ is continuously differentiable, then

$$\frac{\rho(\alpha^{\mathrm{E}})}{\rho(\alpha^*)} \to 1 \text{ as } \delta \to 0.$$

Several remarks on Theorem 1 are in order. First, the proximity between $\rho(\alpha^{\mathrm{E}})$ and $\rho(\alpha^*)$ depends essentially on the gap in means between the best and the second-best systems; specifically, in the (possibly

Table 1: Average CPU time (sec) of estimating α^* and $\alpha^{\mathtt{E}}$ 100 times with different distributions and gaps (δ). The negative inverse link function and logistic link function are applied to the cases of exponential distribution and Bernoulli distribution, respectively.

	Exponential dist.		Bernoulli dist.	
Allocation	$\delta = 0.1$	$\delta = 0.3$	$\delta = 0.1$	$\delta = 0.3$
α^*	11.40	35.91	7.05	22.55
$\alpha^{\mathtt{E}}$	0.13	0.09	0.09	0.07

unlikely) event that the true best system is not selected, the second-best system is most likely to be the best one. Therefore, if a sufficiently large sampling budget is given, the optimal allocation must be designed to inform more about the statistical difference between these two systems than about other pairs. The validity of focusing on top-two systems is discussed in the context of ordinal optimization with independent systems (i.e., each system is not parametrized by features); see, e.g., Shin et al. (2018), Russo (2020).

Second, in the context of ordinal optimization, the effect of model misspefication regarding the GLM can be significantly alleviated. Specifically, the Euclidean approximation $\rho^{\mathtt{E}}(\cdot)$ is structured around the local behavior of the functions $\eta_i(\cdot)$ and $A_i(\cdot)$, rather than their global behavior. That is, although these two functions can be significantly misspecified, the performance in terms of the probability of false selection can be competitive as long as they are close to the true model in the vicinity of the parameter θ. In Section 4, we report the performance of the sampling policies based on the Euclidean approximation in the presence of model misspecification.

Last but not least, the maximization of $\rho^{\mathtt{E}}(\cdot)$ is computationally tractable. Specifically, it is much more efficient to find the maximizer $\alpha^{\mathtt{E}}$ for the Euclidean approximation $\rho^{\mathtt{E}}(\cdot)$ than to find the maximizer α^* for the rate function $\rho(\cdot)$, because the latter entails a bi-level optimization. Table 1 illustrates the computation time of maximizing the rate function $\rho(\cdot)$ and its approximation $\rho^{\mathtt{E}}(\cdot)$. Furthermore, according to Theorem 1, the Euclidean approximation is particularly merited in situations when the means of the top two systems are sufficiently close to each other, in which case the benefit in terms of computational efficiency comes without much sacrifice in performance with respect to the probability of false selection.

4 NUMERICAL EXPERIMENTS

In this section, we numerically examine the performance of our approximate solution based on the Euclidean approximation developed in the previous section. The numerical experiments include a series of tests using normal and Bernoulli distributions. We estimate the probability of false selection $\mathsf{P}(\theta_T \in \mathrm{FS})$ using the Monte Carlo simulation, and our criterion for the number of simulation trials, denoted by N, is as follows:

$$\sqrt{\frac{P_T(1-P_T)}{N}} \leq \frac{P_T}{10}$$

where P_T is the order of magnitude of $\mathsf{P}(\theta_T \in \mathrm{FS})$. This leads to the standard error for each estimate of the probability at least 10 times smaller than the estimate, and makes the confidence level high enough to guarantee the results. Accordingly, we use 10^4 trials in our experiments.

4.1 Normal Distribution

We consider two system configurations, Ξ_1 and Ξ_2, each of which consists of six systems whose performances follow normal distributions equipped with a linear link function $g(z) = z$ and known variances $\sigma_i^2 = 1$ for $i = 1, \ldots, k$. The system configurations Ξ_1 and Ξ_2 are governed by parameters $\theta = (0.2, 0.1, 0.05, 0.1)'$ and

$\theta = (-0.3, 0.2, 0.1, 0.2)'$, respectively. The common feature matrix associated with Ξ_1 and Ξ_2 is given by

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 1 & 3 & 1 \\ 2 & 2 & 1 & 1 & 1 & 0 \\ 2 & 1 & 1 & 2 & 0 & 1 \end{bmatrix}.$$

Thus, the mean vectors for configurations Ξ_1 and Ξ_2 are $\mu = (0.70, 0.60, 0.55, 0.55, 0.55, 0.40)'$ and $\mu = (0.70, 0.50, 0.40, 0.40, 0.40, 0.10)'$, respectively. The gap between the best and the second-best systems is represented by $\delta = \min_{i \neq 1}\{\theta'(\mathbf{x}_1 - \mathbf{x}_i)\}$, where $\delta = 0.1$ in configuration Ξ_1 and $\delta = 0.2$ in configuration Ξ_2.

In order to substantiate our theoretical outcomes, we temporarily assume that the allocation decisions can be made with the knowledge of true θ and consider static allocation rules characterize by a vector α, where system i is sampled $\alpha_i T$ times. For each configuration, we analyze and compare four allocation vectors $\alpha \in \{\alpha^{\mathrm{E}}, \alpha^{\mathrm{E}}_{\text{Bernoulli}}, \alpha^{\mathrm{E}}_{\text{logistic}}, \alpha^{\text{eq}}\}$ defined as follows:

1. the Euclidean approximation α^{E} when the link function g and the density functions f_i are known;
2. the Euclidean approximation $\alpha^{\mathrm{E}}_{\text{Bernoulli}}$ when the link function g is known but the density functions f_i are misspecified as Bernoulli distributions;
3. the Euclidean approximation $\alpha^{\mathrm{E}}_{\text{logistic}}$ when the density functions f_i are known but the link function g is misspecified as a logistic function $\hat{g}(z) = e^z/(1+e^z)$; and
4. the equal allocation α^{eq} when the link function g and the density functions f_i are known.

We note that $\alpha^{\mathrm{E}} = \alpha^*$ in the linear-normal case, so α^{E} is the allocation that maximizes the rate function for the probability of false selection.

Figure 1 depicts the performance of the static allocation rules. In all cases, the static allocation rule based on α^{E} outperforms the other rules. It is important to observe that the allocation rules based on misspecified distribution ($\alpha^{\mathrm{E}}_{\text{Bernoulli}}$) and misspecified link function ($\alpha^{\mathrm{E}}_{\text{logistic}}$) perform competitively with α^{E}, especially in configuration Ξ_1 that features small gap between the best and the second-best systems. These confirm our theoretical findings that, as long as the gap δ is small, the misspecification of distributions and link functions does not have any significant effect on the performance in terms of the probability of false selection, largely owing to the fact that the rate function $\rho(\cdot) = \rho^{\mathrm{E}}(\cdot)$ depends only on the mean $\mu_i(\theta)$ and variance $\sigma_i^2(\theta)$ of each system.

Motivated by the observations under the static allocation rules, we develop a dynamic sampling policy using our Euclidean allocation α^{E}, which is computationally tractable and practically applicable. This policy is denoted by $\pi^{\mathrm{E}}(n_0, m)$ which is described in Algorithm 1, where the parameter n_0 indicates the number of initial samples for each system and m is the batch size. When the distribution functions and link functions are misspecified, one may similarly define dynamic sampling policies, denoted by $\pi^{\mathrm{E}}_{\text{normal}}(n_0, m)$ and $\pi^{\mathrm{E}}_{\text{linear}}(n_0, m)$, respectively. For ease of exposition, we assume that m is a submultiple of $T - kn_0$. Although a mathematical proof is not included in the paper, we numerically observe that under our policy, $\hat{\alpha}^{\mathrm{E}}_t$ approaches α^{E} as T increases if $n_0/T < \min_i\{\alpha^{\mathrm{E}}_i\}$, which implies that the policy is asymptotically optimal as $T \to \infty$ and $\delta \to 0$ according to Theorem 1.

In Figure 2, we present the probability of false selection under our dynamic sampling policy in Algorithm 1. For all cases, we use 20% of the total sampling budget for the initial samples (i.e., $n_0 = 0.2T/k$) and each batch size is set equal to $m = 0.1T$. As in the previous experiments, we use 10^4 simulation trials. Compared to the static allocation cases, except for the fact that the probability of false selection has decreased, we obtain a similar result. Specifically, in configuration Ξ_2 with a relatively large gap (δ) between the best and non-best systems, the dynamic sampling policies with misspecification, $\pi^{\mathrm{E}}_{\text{Bernoulli}}$ and $\pi^{\mathrm{E}}_{\text{logistic}}$, perform significantly poorly compared to π^{E}. However, the performance gap becomes negligible in configuration Ξ_1 with small δ.

Algorithm 1 Dynamic sampling policy $\pi^{\mathrm{E}}(n_0, m)$

Take n_0 samples of Y_i for each system i
Set $t = kn_0$
while $t < T$ **do**
    Find the maximum likelihood estimator θ_t and sample variance s_{it}^2 for each i
    Compute $\hat{\alpha}_t^{\mathrm{E}} = \mathrm{argmax}_{\alpha \in \Delta^{k-1}} \{\rho_t^{\mathrm{E}}(\alpha)\}$, where $\rho_t^{\mathrm{E}}(\alpha)$ is defined in (7) with $\theta = \theta_t$ and $\sigma_i^2 = s_{it}^2$
    Generate a sample vector $(n_{1t}, \ldots, n_{kt})$ from the multinomial distribution with parameters $(m, \hat{\alpha}_t^{\mathrm{E}})$
    Take n_{it} samples of Y_i for each i and set $t = t + m$
end while

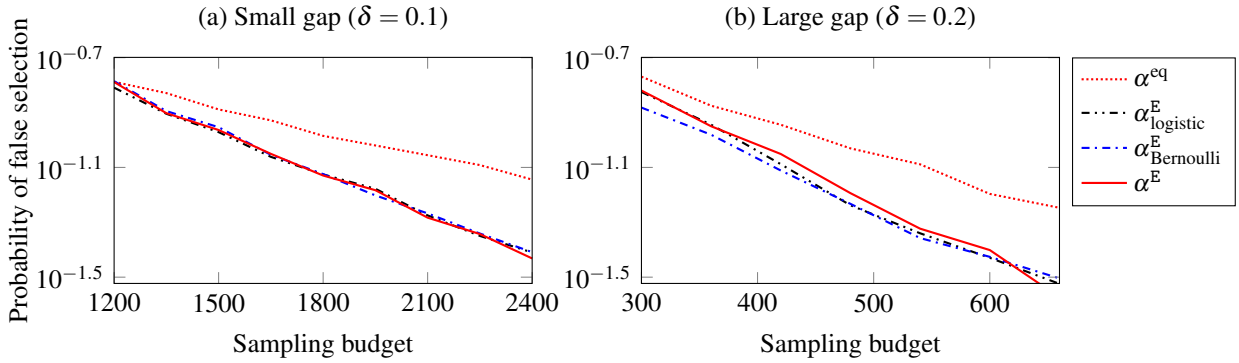

Figure 1: Probability of false selection under static allocation rules, plotted as a function of the sampling budget in a log-linear scale. The left and right panels correspond to the two configurations Ξ_1 and Ξ_2, respectively, with five static allocation rules characterized by $\alpha^{\mathrm{E}}, \alpha^{\mathrm{E}}_{\mathrm{Bernoulli}}, \alpha^{\mathrm{E}}_{\mathrm{logistic}}$, and α^{eq}.

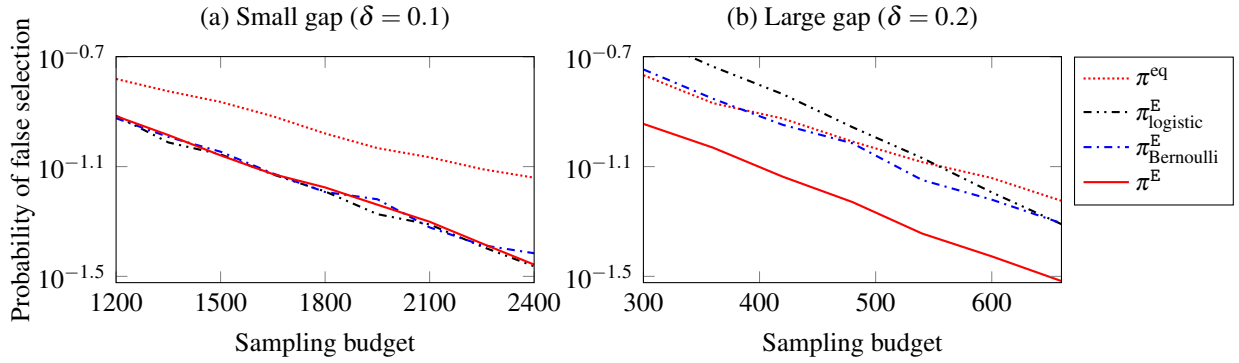

Figure 2: Probability of false selection under dynamic sampling policies, plotted as a function of the sampling budget in a log-linear scale. The left and right panels correspond to the two configurations Ξ_1 and Ξ_2, respectively, with four sampling policies: π^{E}, $\pi^{\mathrm{E}}_{\mathrm{Bernoulli}}$, $\pi^{\mathrm{E}}_{\mathrm{logistic}}$, and π^{eq}.

4.2 Bernoulli Distribution

We next repeat a similar experiment assuming that system performances follow Bernoulli distributions equipped with a logistic link function. We consider two system configurations, Ξ_3 and Ξ_4, each of which consists of six systems whose performances follow Bernoulli distributions equipped with a logistic link function. The underlying distributions under Ξ_3 and Ξ_4 are governed by parameters $\theta = (0, -0.125, -0.1, -0.075)'$

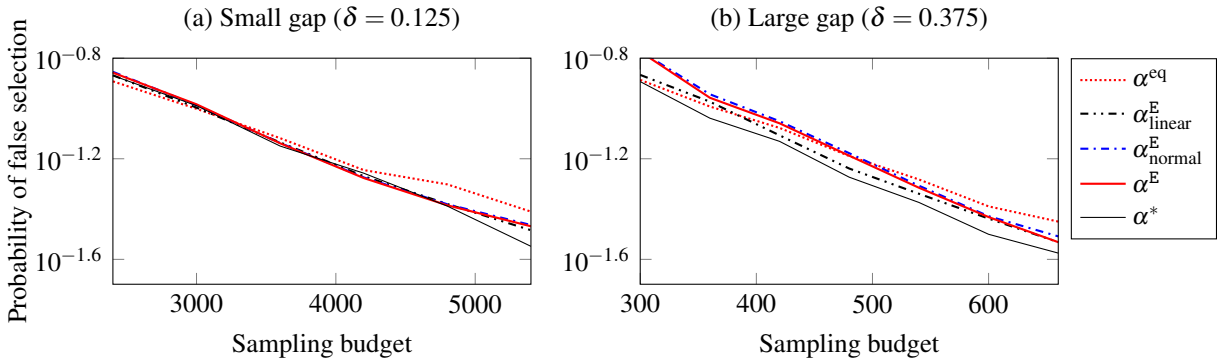

Figure 3: Probability of false selection under static allocation rules, plotted as a function of the sampling budget in a log-linear scale. The left and right panels correspond to the two configurations Ξ_3 and Ξ_4, respectively, with five static allocation rules characterized by $\alpha^*, \alpha^{\mathrm{E}}, \alpha^{\mathrm{E}}_{\mathrm{normal}}, \alpha^{\mathrm{E}}_{\mathrm{linear}}$, and α^{eq}.

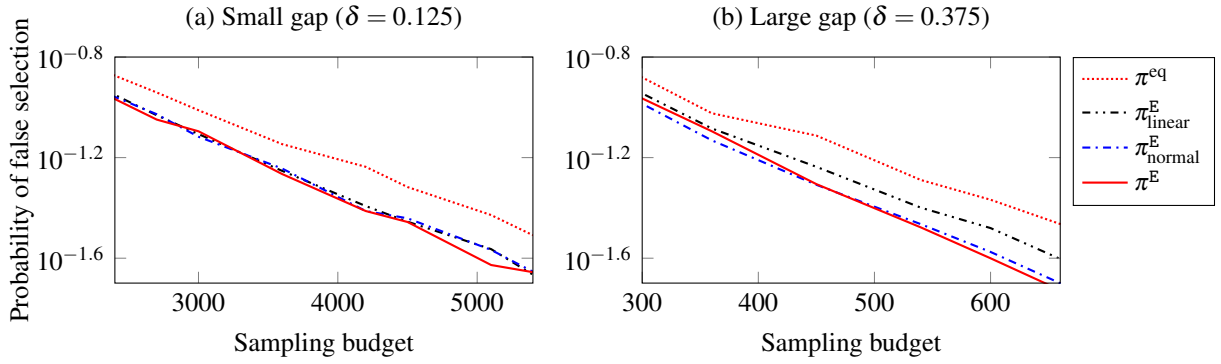

Figure 4: Probability of false selection under dynamic sampling policies, plotted as a function of the sampling budget in a log-linear scale. The left and right panels correspond to the two configurations Ξ_3 and Ξ_4, respectively, with four sampling policies: π^{E}, $\pi^{\mathrm{E}}_{\mathrm{normal}}$, $\pi^{\mathrm{E}}_{\mathrm{linear}}$, and π^{eq}.

and $\theta = (0, -0.375, -0.3, -0.225)'$, respectively. The common feature matrix is given by

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 2 & 2 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 2 \end{bmatrix}.$$

For configuration Ξ_3, the means are $\mu = (0.50, 0.47, 0.44, 0.43, 0.40, 0.38)'$, and those for configuration Ξ_4 are $\mu = (0.50, 0.41, 0.34, 0.29, 0.22, 0.18)'$, and the gap δ between the best and the second-best systems is 0.125 in configuration Ξ_3 and 0.375 in configuration Ξ_4.

For each configuration, we compare five allocation vectors $\alpha \in \{\alpha^*, \alpha^{\mathrm{E}}, \alpha^{\mathrm{E}}_{\mathrm{normal}}, \alpha^{\mathrm{E}}_{\mathrm{linear}}, \alpha^{\mathrm{eq}}\}$, where $\alpha^{\mathrm{E}}_{\mathrm{normal}}$ is the allocation when the distribution is misspecified as normal distributions and $\alpha^{\mathrm{E}}_{\mathrm{linear}}$ is the allocation when the link function g is misspecified as a truncated linear function, $\hat{g}(z) = \min\{z^+, 1\}$. In Figures 3 and 4, we report the estimated probability of false selection under static allocation rules and dynamic sampling policies, respectively, in which similar consequences can be obtained as in Figures 1 and 2. In Figure 4, we do not consider a dynamic sampling policy based on the true rate function $\rho(\cdot)$ for two reasons: the effect of maximizing $\rho^{\mathrm{E}}(\cdot)$ instead of $\rho(\cdot)$ is not significant in the case with a small gap, as is illustrated in Figure 3; and the computational complexity for maximizing $\rho(\cdot)$ (see Table 1) makes it difficult to implement such a policy within a reasonable amount of time.

5 CONCLUDING REMARKS

In this paper, we develop a mathematically rigorous framework for determining the optimal allocation of sampling budget when the underlying random variables are characterized by the generalized linear model. Perhaps the strongest conclusion from our results is that, if the average performance gap between the best and the second-best systems is small, the rate function for the probability of false selection is structured around the first two moments of the system performances. This structural property facilitates construction of dynamic sampling policies that are computationally efficient, representing a fertile area of future research.

REFERENCES

Abbasi-Yadkori, Y., D. Pál, and C. Szepesvári. 2011. "Improved Algorithms for Linear Stochastic Bandits". In *Advances in Neural Information Processing Systems 24*, edited by J. R. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, 2312–2320. Red Hook, New York: Curran Associates, Inc.

Arcones, M. A. 2006. "Large Deviations for M-estimators". *Annals of the Institute of Statistical Mathematics* 58(1):21–52.

Audibert, J., S. Bubeck, and R. Munos. 2010. "Best Arm Identification in Multi-Armed Bandits". In *Proceedings of the 23rd Conference on Learning Theory*. June 27th–29th, Haifa, Israel, 41–53.

Bonnans, J., and A. Shapiro. 1998. "Optimization Problems with Perturbations: A Guided Tour". *SIAM Review* 40(2):228–264.

Dembo, A., and O. Zeitouni. 2009. *Large Deviations Techniques and Applications*. Berlin, Heidelberg: Springer.

Even-Dar, E., S. Mannor, and Y. Mansour. 2006. "Action Elimination and Stopping Conditions for the Multi-Armed Bandit and Reinforcement Learning Problems". *Journal of Machine Learning Research* 7(Jun):1079–1105.

Ho, Y.-C., Q.-C. Zhao, and Q.-S. Jia. 2008. *Ordinal Optimization: Soft Optimization for Hard Problems*. New York: Springer.

Joutard, C. 2004. "Large Deviations for M-Estimators". *Mathematical Methods of Statistics*:179–200.

Kazerouni, A., and L. M. Wein. 2019. "Best Arm Identification in Generalized Linear Bandits". *arXiv preprint arXiv:1905.08224*.

Kullback, S. 1997. *Information Theory and Statistics*. Mineola, New York: Dover Publications.

Lai, T. L., and H. Robbins. 1985. "Asymptotically Efficient Adaptive Allocation Rules". *Advances in Applied Mathematics* 6(1):4–22.

Li, L., W. Chu, J. Langford, and R. Schapire. 2010. "A Contextual-Bandit Approach to Personalized News Article Recommendation". In *Proceedings of the 19th International Conference on World Wide Web*. April 26th–30th, Raleigh, North Carolina, USA, 661–670.

Martin, Y. C. 2010. *Quantitative Drug Design: A Critical Introduction*. Boca Raton, Florida: CRC Press.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. 2nd ed. London: Chapman & Hall.

Russo, D. 2020. "Simple Bayesian Algorithms for Best-Arm Identification". *Operations Research, Articles in Advance*. https://doi.org/10.1287/opre.2019.1911.

Shin, D., M. Broadie, and A. Zeevi. 2018. "Tractable Sampling Strategies for Ordinal Optimization". *Operations Research* 66(6):1693–1712.

Soare, M., A. Lazaric, and R. Munos. 2014. "Best-Arm Identification in Linear Bandits". In *Advances in Neural Information Processing Systems 27*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, 828–836. Cambridge, Massachusetts: MIT Press.

Xu, L., J. Honda, and M. Sugiyama. 2018. "A Fully Adaptive Algorithm for Pure Exploration in Linear Bandits". In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*. April 9th–11th, Playa Blanca, Lanzarote, Canary Islands, 843–851.

AUTHOR BIOGRAPHIES

DOHYUN AHN is an assistant professor in the Department of Systems Engineering and Engineering Management at The Chinese University of Hong Kong. He received his B.S, M.S., and Ph.D. degrees in Industrial & Systems Engineering, all from KAIST. His research interests include, but are not limited to, quantitative risk management, stochastic simulation, decision making via uncertainty quantification, and network analysis in finance and operations. His email address is dohyun.ahn@cuhk.edu.hk.

DONGWOOK SHIN is an assistant professor in the Department of Information Systems, Business Statistics, and Operations Management at the HKUST Business School. He received his Ph.D. degree from Columbia Business School. His research centers on optimal learning theory at the interface between machine learning and sequential decisions making, with applications to problems arising in the context of operations management. His email address is dwshin@ust.hk.