Asymptotically Optimal Simulation Allocation under Dependent Sampling

Xiaoping Xiong
The Robert H. Smith School of Business, University of Maryland, College Park, MD 20742-1815, USA, xiaopingx@yahoo.com

Sandeep Juneja
School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India 400005, juneja@tifr.res.in

Michael C. Fu
The Robert H. Smith School of Business, Institute for Systems Research, University of Maryland, College Park, MD 20742-1815, USA, mfu@rhsmith.umd.edu

We consider the problem of optimal allocation of a simulation computing budget to maximize the probability of correct selection in the ordinal optimization setting where the output samples from different designs have a general distribution and may be mutually dependent, e.g., due to the use of common random numbers to generate them. Asymptotically, in terms of an increasing computational budget, this problem may be viewed in the large deviations framework, so that the rate function of the probability of false (incorrect) selection provides a good surrogate measure for this probability. We evaluate this rate function as a function of computation allocation to different designs and identify the equations satisfied by the optimal allocation in this asymptotic limit. By establishing several theoretical properties that the allocation must satisfy, we reduce the problem to a single-variable nonlinear optimization problem that can be solved by numerical methods. We also characterize the convergence rate of the optimality gap (bias) between the asymptotically optimal solution and the true optimal solution.

---

1. Introduction

In many simulation applications, the objective is to select a best design from among a finite set of competing designs that are evaluated using stochastic system models, where the decision is based on comparing performance measures that are expectations of random variables associated with each design model. In our setting, these expectations are unknown, and simulation is used to generate samples of the associated random variables. One then computes a sample mean for each design, and then selects the design with the best (w.l.o.g., the highest) sample mean. Under sufficient independence in the generated samples, the law of large numbers guarantees the convergence of the sample mean to the corresponding expectation, and the central limit theorem implies that the rate of convergence is proportional to the inverse of the square root of the number of samples generated. The traditional approach towards these types of problems is based on the classical statistical ranking & selection procedures for which there is an enormous body of literature, nearly all of it assuming independent Gaussian sampling distributions. In the simulation context considered here, there has been significant research progress made in the last decade or so on in developing procedures for the correlated setting (e.g., Yang and Nelson 1991; Nelson and Matejcik 1995, Goldsman and Nelson 1998; see also Chick and Inoue 2001ab for a Bayesian approach), but to the best of our knowledge, non-Gaussian distributions have not been addressed.

The approach of ordinal optimization (cf. Ho, Srinivas, and Vakili 1992, Dai and Chen 1997) is based on the observation that the probability of false (incorrect) selection decreases at an *exponential* rate as a function of the total number of generated samples (see Dai 1996, Dai and Chen 1997). Thus, while the number of samples needed to accurately *estimate* expectations may be large (as implied by the inverse square root convergence rate), far fewer samples may be needed to correctly *select* the best design. The optimal computing budget allocation (OCBA) framework introduced by Chen et al. (1997, 2000, 2005) exploits this observation by formulating and approximately solving an optimization problem that maximizes the probability of correct selection for a given total number of simulation replications. Thus, under the optimal allocation, little effort may be expended on obviously inferior designs, while more effort is focused on distinguishing between the top few competing designs. The formulation assumes that each sample from each design has a Gaussian distribution and is independent of all the other generated samples. Glynn and Juneja (2004) use the large deviations framework to rigorously analyze this optimization problem when the generated samples are allowed to have non-Gaussian distributions; however, they still assume independence between generated samples from different designs. In practice, common random numbers may be beneficially used to induce positive dependence between outputs from different designs to further improve the value of the probability of correct selection (see Dai and Chen 1997; Deng, Ho, and Hu 1992). In many applications, dependence between the designs may be endogenous, e.g., consider the case where the performance of different financial portfolios is based on the value of expected utility of each portfolio. Here the utility of any given portfolio is a function of underlying assets and these in turn may depend upon a few common factors.

Motivated by such applications, in this paper we consider the problem of optimal allocation of a computing budget when the outputs from different designs are dependent. Fu et al. (2004, 2006) consider this problem under the assumption that outputs from all the designs follow a multivariate Gaussian distribution. Based on Xiong (2005), we generalize their analysis to allow for non-Gaussian distributions. As noted by Glynn and Juneja (2004), this setting also has important practical consequences, because assuming Gaussian distributions for non-Gaussian sampling can lead to faulty allocations and significant performance degradation, even after batching.

A summary of our specific research contributions is as follows:

- We use the large deviations framework to identify the large deviations rate function associated with the probability of false selection as a function of the computational allocations to various designs. Thus, asymptotically, the problem of finding an optimal allocation to minimize the probability of false selection reduces to a deterministic optimization problem involving the corresponding rate function.

- We establish structural properties of the solution to the asymptotic optimization problem, which leads to a simplified, *single-variable* nonlinear optimization problem, for which we propose a numerical solution procedure.

- We explicitly characterize the convergence *rate* for the "optimality gap" (bias) between the optimal asymptotic allocations and the true (finite sample) optimal allocations. As far as we are aware, these are the first such results in ordinal optimization.

As is typical in the OCBA literature (cf. Chen et al. 1997, 2000; Fu et al. 2004, 2006), the analysis assumes that the large deviations rate function of the probability of false selection is known (e.g., if the joint distribution from samples from different designs is known). Since this is not the case in practice – as even the expected value of the samples is unknown – implementation involves a pilot phase where this rate function is estimated, from which the corresponding optimal allocations are determined, leading to some performance degradation due to estimation errors. However, analysis of this performance degradation is beyond the scope of this paper.

The rest of the paper is organized as follows. In Section 2, we formulate the problem and identify the large deviations rate function associated with the probability of incorrect selection. In Section 3 we develop a numerical solution methodology to solve the nonlinear optimization problem to determine asymptotically optimal allocations. The convergence rate of these allocations to the true (finite sample) optimal allocations is characterized in Section 4. In Section 5, we provide a brief conclusion to the paper. Some of the lengthy proofs are relegated to Appendices A and B.

2. Problem Formulation & Rate Function Identification

Suppose that we have k designs, with μ_i denoting the mean of design $i, i = 1, ..., k$, and the goal is to find the design with maximum mean. Without loss of generality, assume that

$$\mu_1 > \mu_2 \geq \ldots \geq \mu_k,$$

so the best design is design 1. In the setting of this paper, for each design i, μ_i is estimated from i.i.d. samples $J_{ij}, j = 1, 2, ...,$ where $E[J_{ij}] = \mu_i$. Samples *across designs*, however, may be dependent for a fixed (replication number) j, but are otherwise independent. Denote the sample mean for a design based on m replications by $\bar{J}_i(m) = \sum_{j=1}^{m} J_{ij}/m$. The design with the largest sample mean is chosen as the best design; thus, "correct selection" occurs if design 1 has the largest sample mean.

The problem of optimal allocation of a computing budget is to decide the number of samples for each design in order to maximize the probability of correct selection, or conversely to minimize the probability of "false selection." Let $p_i(n)$ denote the proportion of simulation replications (samples) allocated to design i out of the computing budget expressed as the total number of replications n, where $p_i(n) \geq 0$, $\sum_{i=1}^{k} p_i(n) = 1$, and each $p_i(n)n$ is an integer. Thus, $(J_{i,1}, J_{i,2}, \ldots, J_{i,p_i(n)n})$ denotes the samples for design i. Then false selection is made if $\bar{J}_1(p_1(n)n)$ is not the largest sample mean, so we define the probability of false selection by

$$P(FS) \triangleq P(\bar{J}_1(p_1(n)n) \leq \max_{2 \leq i \leq k} \bar{J}_i(p_i(n)n)). \tag{1}$$

The problem of interest is to find a probability vector $\tilde{\mathbf{p}}(\mathbf{n}) = (\tilde{p}_1(n), \ldots, \tilde{p}_k(n))$ that minimizes (1). As mentioned in the introduction, we first use the large deviations framework to identify the rate function of $P(FS)$ as $n \to \infty$. We then find a probability vector $\mathbf{p}^* \in [0, 1]^k$ to minimize this rate function, which acts as a surrogate to $P(FS)$. After providing a numerical procedure for finding $\mathbf{p}^*$, we prove convergence of $\tilde{\mathbf{p}}(\mathbf{n})$ to $\mathbf{p}^*$ and characterize the rate of convergence.

Note that (1) is lower bounded by

$$\max_{2\le i\le k} P(\bar{J}_1(p_1(n)n) \le \bar{J}_i(p_i(n)n))$$

and is upper bounded by

$$(k-1)\max_{2\le i\le k} P(\bar{J}_1(p_1(n)n) \le \bar{J}_i(p_i(n)n)).$$

Therefore, if for $2 \le i \le k$,

$$\lim_{n\to\infty} \frac{1}{n} \log P(\bar{J}_1(p_1(n)n) \le \bar{J}_i(p_i(n)n)) = -R_i(p_1, p_i) \tag{2}$$

for some rate function $R_i(\cdot,\cdot)$, then

$$\lim_{n\to\infty} \frac{1}{n} \log P(FS) = -\min_{2\le i\le k} R_i(p_1, p_i). \tag{3}$$

Thus, asymptotically our optimization problem reduces to finding a probability vector $p^* = (p_1^*, \ldots, p_k^*)$ that solves the optimization problem

$$\max_{(p_1,\ldots,p_k)} \min_{2\le i\le k} R_i(p_1, p_i). \tag{4}$$

We determine $R_i(p_1, p_i)$ using the Gartner-Ellis Theorem (see Dembo and Zeitouni 1998). Some notation and assumptions are needed before we do this.

2.1 Notation

Throughout, we use J_i to denote the generic design i random variable for $J_{ij}, j = 1, 2, ...$ Let $\Lambda_i : \Re^2 \to \Re$ denote the log-moment generating function of (J_1, J_i), i.e.,

$$\Lambda_i(x, y) = \log E[e^{xJ_1 + yJ_i}],$$

and $I_i : \Re^2 \to \Re^+$ denote the Fenchel-Legendre transform of Λ_i, i.e.,

$$I_i(x_1, x_i) = \sup_{\lambda_1,\lambda_i} (\lambda_1 x_1 + \lambda_i x_i - \Lambda_i(\lambda_1, \lambda_i)).$$

Let $\mathcal{D}_{\Lambda_i} \triangleq \{\lambda \in \mathbb{R}^2 : \Lambda_i(\lambda) < \infty\}$. Let A^o denote the interior of set A. It is well known that Λ_i is convex and C^∞ throughout $\mathcal{D}^o_{\Lambda_i}$ (see Dembo and Zeitouni 1998). Let $\mathcal{F}_i \triangleq \{\nabla\Lambda_i(\lambda) : \lambda \in \mathcal{D}_{\Lambda_i^o}\}$. For $x \in \mathcal{F}_i \subset \Re^2$, let λ_x denote the solution to

$$\nabla\Lambda_i(\lambda) = x.$$

Then,

$$I_i(x) = \lambda_x^T x - \Lambda_i(\lambda_x),$$

so that $I_i(x) < \infty$. It is also well known that I_i is convex and differentiable along the interior of $\mathcal{F}_i$, $I_i(\mu_1, \mu_i) = 0$, and $I_i(x) \geq 0$ for all $x \in \mathbb{R}^2$ for $2 \leq i \leq k$ (see Dembo and Zeitouni 1998).

For notational convenience, we introduce the following notational convention to denote partial derivatives for any function $f : \Re^2 \to \Re$:

$$f^{(j,k)}(x,y) \triangleq \frac{\partial^{j+k} f(x,y)}{\partial^j x \partial^k y}.$$

We make the following key assumption:

Assumption 1. *For each $i = 2, 3, ..., k$,*

(i) $(0,0) \in \mathcal{D}^o_{\Lambda_i}$*;*

(ii) $\exists \gamma_i > 0, \beta_i > 0$ *such that* $(-\gamma_i, 0), (0, \beta_i) \in \mathcal{D}^o_{\Lambda_i}$, $\Lambda_i^{(1,0)}(-\gamma_i, 0) = \mu_i$, $\Lambda_i^{(0,1)}(0, \beta_i) = \mu_1$*; and*

(iii) $J_1 - J_i$ *is a non-degenerate random variable.*

Condition (i) ensures that each J_i is light tailed, i.e., both its left and right tail decay at least exponentially fast. Condition (ii) implies $P(J_1 < \mu_i) > 0$, since if $P(J_1 \geq \mu_i) = 1$, no such γ_i exists; similarly, it also implies $P(J_i > \mu_1) > 0$. Condition (iii) is true in practically all cases of interest; even in those cases where it does not hold, it is easy to modify the analysis and the proposed algorithm to handle the situation.

2.2 Determining the Rate Function $R_i(p_1, p_i)$

Let $Z^{(n)}_{(i,p_1(n),p_i(n))} = \bar{J}_i(np_i(n)) - \bar{J}_1(np_1(n))$, and let $\Lambda^{(n)}_{(i,p_1(n),p_i(n))}(x) = \log E[e^{xZ^{(n)}_{(i,p_1(n),p_i(n))}}]$ denote its log-moment generating function. Since for any fixed j, $(J_{ij} : i = 1, ..., k)$ are i.i.d. vectors, we have

$$\Lambda^{(n)}_{(i,p_1(n),p_i(n))}(\lambda) = \begin{cases} np_i(n)\Lambda_i(-\frac{\lambda}{np_1(n)}, \frac{\lambda}{np_i(n)}) + n(p_1(n) - p_i(n))\Lambda_i(-\frac{\lambda}{np_1(n)}, 0) & p_1(n) \geq p_i(n), \\ np_1(n)\Lambda_i(-\frac{\lambda}{np_1(n)}, \frac{\lambda}{np_i(n)}) + n(p_i(n) - p_1(n))\Lambda_i(0, \frac{\lambda}{np_i(n)}) & p_i(n) \geq p_1(n). \end{cases}$$

Note $Z^{(n)}_{(i,p_1(n),p_i(n))}$ is only well defined when $\{np_i(n)\}$ are integers; and $\Lambda^{(n)}_{(i,p_1(n),p_i(n))}(\lambda)$ is well-defined when $\{np_i(n)\}$ are positive. We further assume that $p_i(n) \to p_i > 0$ as $n \to \infty$

for some probability vector $\{p_i\}$. We take this assumption as given for now, and we will show in Section 4 it always holds. It follows from this assumption that

$$\lim_{n\to\infty} \frac{\Lambda^{(n)}_{(i,p_1(n),p_i(n))}(n\lambda)}{n} = \Lambda_{(i,p_1,p_i)}(\lambda),$$

where

$$\Lambda_{(i,p_1,p_i)}(\lambda) = \begin{cases} p_i\Lambda_i(-\frac{\lambda}{p_1}, \frac{\lambda}{p_i}) + (p_1 - p_i)\Lambda_i(-\frac{\lambda}{p_1}, 0) & p_1 \geq p_i, \\ p_1\Lambda_i(-\frac{\lambda}{p_1}, \frac{\lambda}{p_i}) + (p_i - p_1)\Lambda_i(0, \frac{\lambda}{p_i}) & p_i \geq p_1. \end{cases} \tag{5}$$

Since $\mathcal{D}^o_{\Lambda_i}$ is non-empty, $\mathcal{D}^o_{\Lambda_{(i,p_1,p_i)}}$ is non-empty. Let $\mathcal{F}_{(i,p_1,p_i)} \triangleq \{\Lambda_{(i,p_1,p_i)}{}'(\lambda) : \lambda \in \mathcal{D}^o_{\Lambda_{(i,p_1,p_i)}}\}$. To characterize the rate function, we need the following additional assumption:

Assumption 2. *For each $i = 2, 3, ..., k$, $0 \in \mathcal{F}^o_{(i,p_1,p_i)}$ $\forall p_1, p_i > 0$.*

In other words, $\Lambda'_{(i,p_1,p_i)}$ has a zero in $\mathcal{F}^o_{(i,p_1,p_i)}$, which we will denote by

$$\lambda_i(p_1, p_i),$$

which we now show is unique and satisfies some other properties useful for our analysis.

Definition 1. *A function f is said to be homogeneous if $f(\alpha x) = \alpha f(x)$ for $\alpha > 0$.*

Lemma 1. *Under Assumptions 1 and 2, $\lambda_i(p_1, p_i)$ is a unique homogeneous function. Furthermore, it is positive for $p_1, p_i > 0$.*

Proof: Assumption 2 implies the existence of $\lambda_i(p_1, p_i)$ such that $\Lambda'_{(i,p_1,p_i)}(\lambda_i(p_1, p_i)) = 0$, and its uniqueness follows from the strict convexity of $\Lambda_{(i,p_1,p_i)}$, which we now establish.

$\Lambda_i(-\frac{\lambda}{p_1}, \frac{\lambda}{p_i})$ is a convex function of λ, as it is a log-moment generating function of $J_i/p_i - J_1/p_1$ evaluated at λ. Also, $\Lambda_i(0, x)$ and $\Lambda_i(x, 0)$ are strictly convex functions of x in the interior of their domain of finiteness (see Dembo and Zeitouni 1993, 2.2.24). Therefore, for $p_1 \neq p_i$, $\Lambda_{(i,p_1,p_i)}$ is strictly convex throughout $\mathcal{D}^o_{\Lambda_{(i,p_1,p_i)}}$; for $p_1 = p_i$, strict convexity follows from Assumption 1, which ensures that $J_1 - J_i$ is non-degenerate.

It is easy to check that $\Lambda'_{(i,p_1,p_i)}(0) = \mu_i - \mu_1 < 0$. Hence, $\lambda_i(p_1, p_i) > 0$.

By differentiating Equation (5) and setting equal to zero, we have

$$\frac{p_i}{p_1}\Lambda_i^{(1,0)}(-\frac{\lambda_i(p_1,p_i)}{p_1}, \frac{\lambda_i(p_1,p_i)}{p_i}) + \frac{p_1 - p_i}{p_1}\Lambda_i^{(1,0)}(-\frac{\lambda_i(p_1,p_i)}{p_1}, 0)$$
$$= \Lambda_i^{(0,1)}(-\frac{\lambda_i(p_1,p_i)}{p_1}, \frac{\lambda_i(p_1,p_i)}{p_i}), \qquad p_i \leq p_1,$$

$$\Lambda_i^{(1,0)}(-\frac{\lambda_i(p_1,p_i)}{p_1},\frac{\lambda_i(p_1,p_i)}{p_i})$$
$$= \frac{p_1}{p_i}\Lambda_i^{(0,1)}(-\frac{\lambda_i(p_1,p_i)}{p_1},\frac{\lambda_i(p_1,p_i)}{p_i}) + \frac{p_i-p_1}{p_i}\Lambda_i^{(1,0)}(0,\frac{\lambda_i(p_1,p_i)}{p_i}), \qquad p_1 \le p_i.$$

It follows that for any $\alpha > 0$, $\lambda_i(\alpha p_1, \alpha p_i) = \alpha\lambda_i(p_1, p_i)$. □

The lemma means that Assumption 2 is essentially equivalent to requiring the existence of $\lambda_i(1,p)$ satisfying $\Lambda'_{(i,1,p)}(\lambda_i(1,p)) = 0$ for each $p > 0$.

We can now state the main result characterizing the rate function.

Theorem 1. *Under Assumptions 1 and 2, $R_i(p_1,p_i) = I_{(i,p_1,p_i)}(0) = -\Lambda_{(i,p_1,p_i)}(\lambda_i(p_1,p_i))$ satisfies (2), so that (3) holds with this rate function.*

Proof: Under Assumptions 1 and 2, using the Gartner-Ellis Theorem (see Dembo and Zeitouni 1998, Theorem 2.3.6 and Lemma 2.3.9), it can be seen that for each $i \geq 2$, $(Z^{(n)}_{(i,p_1(n),p_i(n))} : n \geq 0)$ satisfies the large deviations principle, so that

$$\lim_{n\to\infty}\frac{1}{n}\log P(Z^{(n)}_{(i,p_1(n),p_i(n))} \geq 0) = -\inf_{x\geq 0} I_{(i,p_1,p_i)}(x), \tag{6}$$

where $I_{(i,p_1,p_i)}(x) = \sup_{\lambda\in\Re}(\lambda x - \Lambda_{(i,p_1,p_i)}(\lambda))$. Now, $I_{(i,p_1,p_i)}$ is strictly convex and C^∞ in the set $\mathcal{F}^o_{(i,p_1,p_i)}$ (see Dembo and Zeitouni 1998, 2.2.24), and equal to zero when $\mu_i - \mu_1 < 0$. Therefore,

$$R_i(p_1,p_i) = \inf_{x\geq 0} I_{(i,p_1,p_i)}(x) = I_{(i,p_1,p_i)}(0) = -\inf_{\lambda}\Lambda_{(i,p_1,p_i)}(\lambda) = -\Lambda_{(i,p_1,p_i)}(\lambda_i(p_1,p_i)).$$

□

Recall our optimization problem (4). If each $R_i(\cdot,\cdot)$ were a concave function, our optimization problem would be a concave optimization problem and first-order conditions would be necessary and sufficient for optimality. As noted in Glynn and Juneja (2004) and Xiong (2005), this is true when the samples of different designs are independently generated. However, when there is dependence amongst samples, $R_i(\cdot,\cdot)$ need not be a concave function making the optimization problem much harder to solve. In Section 3.2, this lack of concavity is illustrated in Example 3.

Note that if for each $i \geq 2$, J_1 and J_i have positive dependence, i.e.,

$$P(J_1 \leq x, J_i \leq y) \geq P(J_1 \leq x)P(J_i \leq y), \forall x,\ y, \tag{7}$$

then the optimization problem (4) gives a better solution compared to the case where each J_1 and J_i are independent or negatively dependent, where the latter corresponds to the reversal of the inequality in (7).

To see this, note that under positive dependence, for any $\lambda, \mu > 0$, $e^{-\lambda J_1}$ and $e^{\mu J_i}$ are negatively dependent, i.e.,

$$P(e^{-\lambda J_1} \leq x, e^{\mu J_i} \leq y) \leq P(e^{-\lambda J_1} \leq x)P(e^{\mu J_i} \leq y), \forall x,\ y.$$

This implies that (Lemma 2.1 in Jin et al. 2003)

$$Ee^{-\lambda J_1 + \mu J_i} \leq Ee^{-\lambda J_1} Ee^{\mu J_i}.$$

With simple algebra, it can be seen that $R_i(p_1, p_i)$ dominates $R_i(p_1, p_i)$ under independence for any non-negative p_1 and p_i, and the claim follows.

We now illustrate some properties of $R_i(\cdot, \cdot)$ that are useful in designing algorithms to numerically solve (4).

Lemma 2. *For all $i \geq 2$, $R_i(p_1, p_i)$ is increasing in p_i when $p_i \leq p_1$, and increasing in p_1 when $p_i \geq p_1$.*

Proof: Recall that $R_i(p_1, p_i) = -\Lambda_{(i,p_1,p_i)}(\lambda_i(p_1, p_i))$ and

$$\frac{\partial \Lambda_{(i,p_1,p_i)}(\lambda)}{\partial \lambda}\Big|_{\lambda=\lambda_i(p_1,p_i)} = 0.$$

If we fix $p_1 > 0$ and let $p_i < p_1$, we have

$$R_i^{(0,1)}(p_1, p_i) = -\frac{\partial \Lambda_{(i,p_1,p_i)}(\lambda)}{\partial p_i}\Big|_{\lambda=\lambda_i(p_1,p_i)}.$$

For notational convenience, we denote $\lambda_i(p_1, p_i)$ as λ_i in this proof. Then

$$R_i^{(0,1)}(p_1, p_i) = -\Lambda_i(-\frac{\lambda_i}{p_1}, \frac{\lambda_i}{p_i}) + \frac{\lambda_i}{p_i}\Lambda_i^{(0,1)}(-\frac{\lambda_i}{p_1}, \frac{\lambda_i}{p_i}) + \Lambda_i(-\frac{\lambda_i}{p_1}, 0). \tag{8}$$

Consider function $G(x) \equiv \Lambda_i(-\frac{\lambda_i}{p_1}, x)$. It is easily seen that $G''(x) > 0$ for any x. We may rewrite the right hand side of (8) as

$$G(0) - G(\frac{\lambda_i}{p_i}) + \frac{\lambda_i}{p_i}G'(\frac{\lambda_i}{p_i}) = \frac{\lambda_i^2}{2p_i^2}G''(\theta\frac{\lambda_i}{p_i}) \geq 0,$$

where the second equality is by a Taylor expansion of $G(0)$ around $\frac{\lambda_i}{p_i}$ and $\theta \in [0, 1]$. With this, we establish the first part. Applying similar arguments to $p_i > p_1$, the second part can be similarly established. □

Recall that $R_i(p_1, p_i)$ denotes the rate function associated with $P(FS)$ when only design 1 and i are considered and $p_1 n$ samples are allocated to design 1 while $p_i n$ samples are allocated to design i (here we assume that $p_1 n$ and $p_i n$ are integers to facilitate the discussion). Then $R_i(p_1, p_i)$ is a surrogate measure for $P(FS)$, in that the larger its value, asymptotically the smaller the value of $P(FS)$. Lemma 2 therefore is intuitively plausible as larger the number of samples of the design assigned less samples, asymptotically, lower the $P(FS)$. However, as our analysis indicates, this monotonicity may breakdown when the samples of design assigned more samples are further increased. For instance, $R_i(p_1, p_i)$ may decrease as p_i is increased when $p_i > p_1$. This may happen when J_1 and J_i are highly positively dependent, so that $J_1 - J_i$ has very little variability. Under such dependence, consider the case where $p_1 = p_i$. Then the probability

$$P\left(\frac{1}{p_1 n}\sum_{k=1}^{p_1 n} J_{1k} \le \frac{1}{p_i n}\sum_{k=1}^{p_i n} J_{ik}\right) = P\left(\frac{1}{p_1 n}\sum_{k=1}^{p_1 n}(J_{1k} - J_{ik}) \le 0\right)$$

may be quite small. However, if p_i is increased further, then this probability becomes

$$P\left(\frac{1}{p_1 n}\sum_{k=1}^{p_1 n} J_{1k} \le \left(\frac{p_1}{p_i}\right)\frac{1}{p_1 n}\sum_{k=1}^{p_1 n} J_{ik} + \left(\frac{p_i - p_1}{p_i}\right)\frac{1}{(p_i - p_1)n}\sum_{k=p_1 n+1}^{p_i n} J_{ik}\right). \tag{9}$$

Thus, as p_i increases, the weight given to the dependent samples reduces while that given to the independent samples increases. This suggests that in certain cases, $R_i(p_1, p_i)$ may decrease as p_i is increased from p_1, due to added noise from the independent samples $(p_i - p_1)n$ samples.

Equation (9) suggests that as $p_i \to \infty$, we may expect $R_i(p_1, p_i)$ to converge to the rate function of

$$P\left(\frac{1}{p_1(n)n}\sum_{k=1}^{p_1(n)n} J_{1k} \le \mu_i\right).$$

We show this in Lemma 4, and establish some other properties of $R_i(p_1, p_i)$ in Lemmas 3 and 4 useful to our analysis.

Lemma 3. *Under Assumptions 1 and 2, $R_i(\cdot, \cdot)$ is a homogeneous function.*

Proof: Follows from Lemma 1 and Theorem 1, applying the definition of $\Lambda_{(i,p_1,p_i)}(\lambda_i(p_1, p_i))$ in (5). □

Thus, it suffices to focus on $R_i(1, p)$ for $p > 0$ to understand the behavior of $R_i(p_1, p_i)$. For notational simplicity, let $\lambda_i(p)$ denote $\lambda_i(1, p)$ and $R_i(p)$ denote $R_i(1, p)$.

Let

$$f_i(\lambda, p) = -\Lambda_{(i,1,p)}(\lambda) = -\Lambda_i(-\lambda, \frac{\lambda}{p}) - (p-1)\Lambda_i(0, \frac{\lambda}{p}). \tag{10}$$

Then, $R_i(p) = f_i(\lambda_i(p), p)$. Also, as discussed before, $\lambda_i(p)$ is the unique solution to $f_i^{(1,0)}(\lambda_i(p), p) = 0$. Therefore, from the implicit function theorem, it follows that

$$\lambda_i'(p) = -\frac{f_i^{(1,1)}(\lambda_i(p), p)}{f_i^{(2,0)}(\lambda_i(p), p)}.$$

It is also easy to see that

$$R_i{}'(p) = f_i^{(0,1)}(\lambda_i(p), p).$$

Furthermore,

$$R_i{}''(p) = \frac{f_i^{(2,0)}(\lambda_i(p), p) f_i^{(0,2)}(\lambda_i(p), p) - f_i^{(1,1)}(\lambda_i(p), p)^2}{f_i^{(2,0)}(\lambda_i(p), p)}.$$

Lemma 2 and Lemma 3 imply that $R_i(p)$ is increasing when $p < 1$. In order to know the monotonicity of $R_i(p)$ on $(1, \infty)$, we need to analyze the first and second derivatives of $R_i(p)$ and its behavior as $p \to \infty$. So we put down the expressions of both derivatives above. The following Lemma gives the asymptotic behavior of $R_i(p)$ and $\lambda_i(p)$.

Lemma 4. *Under Assumptions 1 and 2,*

$$\lambda_i(p) \to \gamma_i \tag{11}$$

as $p \to \infty$, and

$$\frac{\lambda_i(p)}{p} \to \beta_i \tag{12}$$

as $p \to 0$. Furthermore,

$$R_i(\infty) = \lim_{p\to\infty} R_i(p) = -\Lambda_i(-\gamma_i, 0) - \gamma_i\mu_i = \inf_x I_i(\mu_i, x), \tag{13}$$

$$\lim_{p\to\infty} R_i{}'(p) = 0, \tag{14}$$

and $R_i(0^+) = 0$.

The proof of Lemma 4 is given in the Appendix. These properties will be used in the three examples presented later in the next section. Since under Assumptions 1 and 2, $R_i(0+) = 0$, we set $R_i(0) = 0$ henceforth.

3. Solving the Asymptotic Optimization Problem

In this section, we first reduce the asymptotic optimization problem (4) to a single variable optimization problem. We then develop a numerical procedure to solve this problem.

3.1 Transformation to a Single Variable Problem

Let $\Omega = \{2, 3, \ldots, k\}$. Setting $x_i = \frac{p_i}{p_1}$ for $i \in \Omega$, we can rewrite the budget constraint as $p_1(1 + \sum_{i\in\Omega} x_i) = 1$, and problem (4) as:

$$\max_{(x_j \geq 0 : j \in \Omega)} \frac{\min_{i\in\Omega} R_i(x_i)}{1 + \sum_{i\in\Omega} x_i}. \tag{15}$$

Recall that $R_i(\cdot)$ need not be a monotone function. Set $R_i^{-1}(z) \equiv \inf\{x > 0 | R_i(x) = z\}$. If $(x_i^* : i \in \Omega)$ solves (15), then the solution to (4) corresponds to $p_1^* = \frac{1}{1+\sum_{i\in\Omega} x_i^*}$ and $p_i^* = p_1^* x_i^*$. The following theorem characterizes such a solution.

Theorem 2. *Under Assumptions 1 and 2, if $(x_i^* : i \in \Omega)$ denotes an optimal solution to the optimization problem (15), then for each $i \in \Omega$, the following hold:*

1. $0 < x_i^* < \infty$.
2. $\exists z^* > 0$ *such that* $R_i(x_i^*) = z^*$ $\forall i \in \Omega$. *Furthermore,* $x_i^* = R_i^{-1}(z^*)$.
3. *If* $x_i^* \neq 1$, *then* $R_i'(x_i^*) \geq 0$.
4. *If* $z^* \leq R_i(1)$, *then* $x_i^* \leq 1$.

Proof: Under Assumptions 1 and 2, $R_i(0) = 0$, so it's clear that each $x_i^* > 0$. Similarly, since for any $x_i = \infty$, the objective function value equals zero, it follows that at an optimal point, each $x_i^* < \infty$.

To see conclusion 2, first note that $R_i(x_i^*)$ must be equal for each $i \in \Omega$. Otherwise, suppose that $R_i(x_i^*) < R_j(x_j^*)$ for some $i, j \in \Omega$. Then a small reduction in x_j^* does not change the numerator of the objective function, but it reduces the denominator, contradicting the optimality of $(x_i^* : i \in \Omega)$. Let $z^* = R_i(x_i^*)$. It is easy to see that $x_i^* = R_i^{-1}(z^*)$. Otherwise, $x_i^* > R_i^{-1}(z^*)$, and a better solution to (15) is obtained by replacing x_i^* with $R_i^{-1}(z^*)$.

To see conclusion 3, note that if $R_i'(x_i^*) < 0$, then there exists an $x < x_i^*$ such that $R_i(x) > R_i(x_i^*)$. Since $R_i(0) = 0$, and $R_i(\cdot)$ is a continuous function, it follows that there exists a $\tilde{x} < x < x_i^*$ such that $R_i(\tilde{x}) = R_i(x_i^*)$. This contradicts the fact that $x_i^* = R_i^{-1}(z^*)$.

Conclusion 4 now follows from Lemma 2. □

It can be seen that R_i^{-1} is continuous for Gaussian distributions (see Fu et al. 2006), but for non-Gaussian distributions, R_i^{-1} can be discontinuous when R_i is not monotone. This is illustrated in Example 3 (Figure 1: $a = 0.95, b = 1.05$ case) in the next section. The next lemma states some features of function R_i^{-1}.

Lemma 5. *Under Assumptions 1 and 2,*

1. R_i^{-1} *is strictly increasing.*
2. R_i^{-1} *is left continuous.*
3. *If* R_i^{-1} *is discontinuous at* z_0 *with* $x_0 = R_i^{-1}(z_0)$*, then* $R_i'(x_0) = 0$.

Proof.

1. Suppose $x_0 = R_i^{-1}(z_0) > 0$. Then for arbitrary $z_1 \in (0, z_0)$, there exists $x_1 \in (0, x_0)$ such that $R_i(x_1) = z_1$ by continuity of R_i. Hence, $R_i^{-1}(z_1) \le x_1 < x_0$.
2. Let $R_i^{-1}(z_0) = x_0$ and $z_n \uparrow z_0$. Then there exists $x_1 = \lim_{n\to\infty} R_i^{-1}(z_n)$. By continuity of R_i, we have $z_0 = \lim_{n\to\infty} z_n = \lim_{n\to\infty} R_i(R_i^{-1}(z_n)) = R_i(x_1)$. By monotonicity above, we have $x_1 \le x_0$. Then the definition of R_i^{-1} implies $x_1 = x_0$.
3. Discontinuity at z_0 implies there exists $x_1 > x_0$ such that $R_i^{-1}(z) \ge x_1$ whenever $z > z_0$. The above two features imply $R_i'(x_0) \ge 0$. Suppose $R_i'(x_0) > 0$. Then there exists $x_2 \in (x_0, x_1)$ such that $z_2 = R_i(x_2) > z_0$. Now we have $R_i^{-1}(z_2) \le x_2 < x_1$, a contradiction. □

Let $R_i^b \equiv \sup_{x\ge 0} R_i(x)$. Obviously, $z^* \le R^b \equiv \min_{i\in\Omega} R_i^b$. Now we can rewrite problem (15) as a single-variable nonlinear programming (NLP) problem

$$\min_{0\le z\le R_b} F(z) \tag{16}$$

where

$$F(z) = \frac{1 + \sum_{i=2}^{k} R_i^{-1}(z)}{z}. \tag{17}$$

Once we find an optimizer to this problem, z^*, we can easily obtain $\{p_i^*\}$ via

$$p_1^* = \frac{1}{1 + \sum_{i\in\Omega} R_i^{-1}(z^*)}, \quad p_i^* = p_1^* R_i^{-1}(z^*). \tag{18}$$

Let $\tilde{R}_i^b = \max(R_i(1), R_i(\infty))$. Then, clearly $R_i^b \geq \tilde{R}_i^b$. It is intuitively plausible that at least in typical cases, $R_i^b = \tilde{R}_i^b$. To see this, recall that $R_i(x)$ serves as a surrogate for the probability

$$P\left(\frac{1}{n}\sum_{k=1}^{n} J_{1k} \leq \frac{1}{xn}\sum_{k=1}^{xn} J_{ik}\right)$$

for large values of n. As proved in Lemma 2, $R_i(x)$ increases with x for $x < 1$. However, it may decrease with x for $x > 1$. It is intuitively plausible that either $R_i(x)$ increases with x for all x, in which case its highest value $R_i^b = R_i(\infty)$ computed in Lemma 4. Otherwise, it may have $R_i'(1+) < 0$, and it may either decrease with x for all $x > 1$, or that it first decreases and then increases with x. Thus, in all such cases, $R_i^b = \tilde{R}_i^b$. In particular, in these cases, $R_i(\cdot)$ is quasi-convex in the region $(1, \infty)$. Next, we illustrate this with some examples.

3.2 Examples

We present some illustrative examples where $R_i^b = \tilde{R}_i^b$. This subsection can be skipped without loss of continuity.

Example 1. Suppose that $\Lambda_i^{(0,1)}(-\mu, \lambda) \geq \Lambda_i^{(0,1)}(0, \lambda)$, for all $\mu \geq \lambda > 0$, and $(-\mu, \lambda) \in \mathcal{D}_{\Lambda_i}$. In the case of Gaussian and Bernoulli distributions, this condition corresponds to negative correlation between J_1 and J_i.

We know $R_i(p) = f_i(\lambda_i(p), p))$, where f_i is given by Equation (10). Then we have

$$\begin{aligned} R_i'(p) = f_i^{(1,0)}(\lambda_i(p), p) &= \frac{\lambda_i(p)}{p^2}\big[\Lambda_i^{(0,1)}(-\lambda_i(p), \frac{\lambda_i(p)}{p}) + (p-1)\Lambda_i^{(0,1)}(0, \frac{\lambda_i(p)}{p})\big] - \Lambda_i(0, \frac{\lambda_i(p)}{p}) \\ &\geq \frac{\lambda_i(p)}{p}\Lambda_i^{(0,1)}(0, \frac{\lambda_i(p)}{p}) - \Lambda_i(0, \frac{\lambda_i(p)}{p}) > 0, \end{aligned}$$

where the last step uses convexity of $\Lambda_i(0, \lambda)$. Hence we always have an increasing $R_i(p)$. In particular, $R_i^b = \tilde{R}_i^b$.

Example 2. Suppose that $\Lambda_i^{(0,2)}(-\mu, \lambda) \leq \Lambda_i^{(0,2)}(0, \lambda)$ and $\Lambda_i^{(0,3)}(0, \lambda) \leq 0$, for all $\mu \geq \lambda > 0$, and $(-\mu, \lambda) \in \mathcal{D}_{\Lambda_i}$. It can be seen that these conditions hold for a Gaussian distribution. In the appendix, we prove that R_i is quasi-convex, so $R_i^b = \tilde{R}_i^b$. A non-Gaussian distribution where both conditions hold (proof in the appendix) is the following:

$$J_1 \sim U(-1, 1), \quad J_i = cJ_1 - a + bU_i, \quad c \in [0, 0.5],$$

where $U_i \sim U(-1, 1)$ are i.i.d. and independent of J_1.

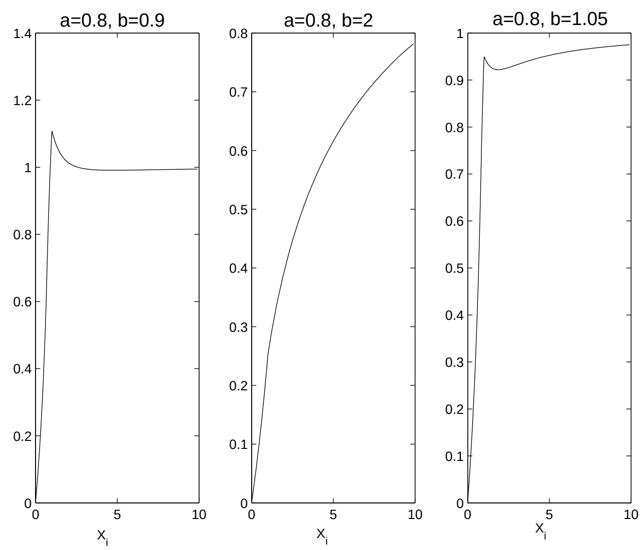

Figure 1: Function $R_i(x_i^*)$

Example 3. Consider J_1 and J_i as in Example 2, with $c = 1$, $0 < a < 1$ and $a < b$. In this case, it is difficult to establish quasi-convexity of R_i on $(1, \infty)$ analytically. However, numerically we observe this to be true for three set of parameters (refer to Figure 1; see also the appendix for more details):

- $1 > b > a$: $x_i^* \in (0, 1]$;
- $b \gg 1$: Since R_i is increasing to the right of 1, $x_i^* \in (0, \infty)$;
- $b > 1 + o(1)$: Since $R_i^{-1}(R_i(x)) \neq x$ for any $x \in (1, 4.7]$, $x_i^* \notin (1, 4.7]$.

3.3 Numerical Procedure for solving the NLP

We now develop an efficient numerical procedure to solve the one-dimensional NLP

$$\min_{0 \leq z \leq \tilde{R}_b} F(z), \tag{19}$$

where $\tilde{R}^b = \min_{i \in \Omega} \tilde{R}_i^b$ and F is given by (17). Note that even if $R^b > \tilde{R}^b$, the solution to (19) still provides a good solution to (16).

Straightforward algebra yields

$$M(z) \equiv z^2 F'(z) = \sum_{i=2}^{k} \frac{z}{R_i'(R_i^{-1}(z))} - 1 - \sum_{i=2}^{k} R_i^{-1}(z). \tag{20}$$

If F is smooth, we only need to locate all the roots of $M(z) = 0$ in interval $[0, \tilde{R}_b]$. But F is not differentiable at $R_i(1)$ unless J_i and J_1 are independent. Also note that F may not even be continuous when some R_i is not monotone, as then R_i^{-1} is discontinuous. So $M(z^*) = 0$ may not be satisfied or even well defined for each optimizer z^* to problem (19). Now we define two sets as follows:

$$\mathcal{S}_1 = \{R_i(1)|i \in \Omega\} \cap [0, \tilde{R}_b), \mathcal{S}_2 = \{z \in (0, \tilde{R}_b)|F(\cdot) \text{ discontinuous at } z\}.$$

If we let $k = 4$ and functions $\{R_i(\cdot), i = 2, 3, 4\}$ to be set as the three cases from left to right in Figure 1, then we can show $\tilde{R}_b = R_2(1)$, $\mathcal{S}_1 = \{R_3(1), R_4(1)\}$, and $\mathcal{S}_2 = \{R_4(1)\}$. Theorem 3 presents some necessary conditions that an optimizer z^* satisfies.

Theorem 3. *If z^* is an optimizer of problem (19), one of the following three conditions has to be satisfied:*

(i) $F'(z^*) = 0$, $z^* \notin \mathcal{S}_1$, $0 < z^* < \tilde{R}_b$;

(ii) $F'(z^*-) \leq 0$, $F'(z^*+) \geq 0$, $z^* \in \mathcal{S}_1 \setminus (\mathcal{S}_2 \cup \{\tilde{R}_b\})$;

(iii) $F'(z^*-) \leq 0$, $z^* \in \mathcal{S}_1 \cap (\mathcal{S}_2 \cup \{\tilde{R}_b\})$.

Proof: Note that $F(0+) = \infty$, so 0 cannot be an optimizer. Now consider the point $\tilde{R}^b$, and assume $\tilde{R}^b \notin \mathcal{S}_1$. If $R_i^{-1}(\tilde{R}^b) = \infty$ for some $i \in \Omega$, then the objective function $F(\tilde{R}^b-) = \infty$, and hence $\tilde{R}^b$ cannot be optimal. Now assume $R_i^{-1}(\tilde{R}^b)$ is finite for all $i \in \Omega$. Recall that $\tilde{R}^b = \min_{i\in\Omega} \tilde{R}^b_i$. Thus, $\tilde{R}^b = \tilde{R}^b_i$ for some i, and since $\tilde{R}^b \notin \mathcal{S}_1$, it follows that $\tilde{R}^b = R_i(\infty) > R_i(1)$. Therefore, $R_i^{-1}(\tilde{R}_b) = +\infty$. From (14), it follows that $R_i'(R_i^{-1}(\tilde{R}_b)) = 0$. Then we have $F'(\tilde{R}_b-) = \infty$, because $\frac{z}{R_i'(R_i^{-1}(z))} - R_i^{-1}(z) \to \infty$ as $z \downarrow \tilde{R}_b$. Therefore, $\tilde{R}_b$ cannot be optimal as long as $\tilde{R}_b \notin \mathcal{S}_1$.

Let $\omega \in \mathcal{S}_2 \setminus \mathcal{S}_1$. Lemma 5 implies that F is left continuous at ω, and $R_i'(R_i^{-1}(\omega)) = 0$ for some $i \in \Omega$. So we have $F'(\omega-) = \infty$, and thus ω cannot be optimal. Now if we assume $z^* \in (0, \tilde{R}_b) \setminus \mathcal{S}_1$, we have $z^* \notin \mathcal{S}_2$, i.e., F is continuous and differentiable at z^*. So $F'(z^*) = 0$ has to be true. This is the first situation in the theorem. For the case $z^* \in \mathcal{S}_1 \setminus (\mathcal{S}_2 \cup \{\tilde{R}_b\})$, we know F' has both left and right limits at z^*, so $F'(z^*-) \leq 0$ and $F'(z^*+) \geq 0$ have to hold. This is the second situation in the theorem. For the case $z^* \in \mathcal{S}_1 \cap (\mathcal{S}_2 \cup \{\tilde{R}_b\})$, we know $F'(\cdot)$ only has a left limit, so $F'(z^*-) \leq 0$, which is the third condition in the theorem.

□

The following establishes a useful property of the optimal solution $\{p_i^*, i \leq k\}$.

Lemma 6. *Without loss of generality, assume* $R_2(1) \le R_3(1) \le \cdots \le R_k(1)$*. Then there exists* $2 \le i_1 \le i_2 \le k$ *such that*

$$\begin{cases} p_i^* > p_1^*, & 2 \le i < i_1, \\ p_i^* = p_1^*, & i_1 \le i < i_2, \\ p_i^* < p_1^*, & i_2 \le i, \end{cases}$$

where $p_i^* = p_j^* = p_1^*$ *for some* $i_1 \le i, j < i_2$ *only if* $R_i(1) = R_j(1)$*.*

Proof: Suppose $R_i(1) < z^*$ for some i. Then $\frac{p_i^*}{p_1^*} = R_i^{-1}(z^*) > 1 \Longrightarrow p_i^* > p_1^*$. The reverse equality holds if $R_i(1) > z^*$. Since $p_i^* = p_1^*$ if and only if $R_i(1) = z^*$, we can prove the second part of the claim. □

This feature is consistent with intuition. The order of $R_i(1)$ reflects the order of simulation efficiency. A small (large) value implies low (high) efficiency, which then asks for more (less) allocation.

In order to locate the global optimizer of problem (16), we need to solve the following two problems:

P1. Calculate $F(z)$, or more precisely, $R_i^{-1}(z)$, for any given $R_i(1) < z \le \tilde{R}_i^b$.

P2. Find all z^* satisfying conditions in Theorem 3.

If we assume $R_i(x)$ on $(1, \infty)$ only takes three possible shapes as illustrated in Example 3 of the previous section, P1 is trivial to solve by calculating $R_i^{-1}(z)$ via numerical procedures such as the bisection method.

P2 depends on our knowledge on convexity of function F. If all the samples follow Gaussian distributions, Fu et al. (2006) show that F is piecewise convex. However, this may not be true in general settings. Let V denote the optimal value of problem (16). For an arbitrary $\Delta > 0$, the following result provides a means to locate a solution with objective value in the interval $[V, V + \Delta]$. Let $F_0 = F(\min_{i\in\Omega} R_i(1))$.

Theorem 4. *Consider a sequence* $\{z_n\}$ *where* $z_1 = \frac{1}{F_0 - \Delta}$ *and*

$$z_{n+1} = \frac{F(z_n) z_n}{\min_{1\le j\le n} F(z_j) \wedge F_0 - \Delta}, \text{ where } a \wedge b = \min(a, b).$$

For $\Delta < V$*, there exists an* m *such that* $z_m \ge \tilde{R}_b$ *and* $z_{m-1} < \tilde{R}_b$*. In addition,*

$$\min_{1\le n\le m} F(z_n) \wedge F_0 \le V + \Delta.$$

Proof. Since $\Delta < V$,

$$z_{n+1} = \frac{F(z_n)z_n}{\min_{1\le j\le n} F(z_j) \wedge F_0 - \Delta} \ge \frac{F(z_n)z_n}{F(z_n) - \Delta} > z_n. \tag{21}$$

Suppose the sequence never exceeds $\tilde{R}_b$, then it has to converge to some value, say $r \le \tilde{R}_b$. Since $F(z)$ is left continuous from Lemma 5, we have $F(z_n) \to F(r)$. Letting $n \to \infty$ in (21), we have $r \ge \frac{F(r)r}{F(r)-\Delta} > r$, a contradiction.

Note that $G(z) \equiv zF(z) = 1 + \sum_{i=2}^{k} R_i^{-1}(z)$ is an increasing function of z. Let $z_0 = 0$. Then, for any $0 \le n < m$ and any $z \in (z_n, z_{n+1}]$, we have

$$\begin{aligned} F(z) = \frac{G(z)}{z} \quad &\ge \quad \frac{G(z_n)}{z_{n+1}} = \begin{cases} F_0 - \Delta, & n = 0 \\ \min\limits_{1\le j\le n} F(z_j) \wedge F_0 - \Delta, & n > 0 \end{cases} \\ &\ge \quad \min_{1\le n\le m} F(z_n) \wedge F_0 - \Delta. \end{aligned}$$

We complete the proof by noticing that $(0, \tilde{R}_b] \subset \bigcup_{n=0}^{m-1} (z_n, z_{n+1}]$. □

Based on Theorem 4, we propose the following explicit algorithm:

Algorithm for Solving the Approximation Problem.

- **Step 1.** Let $z^* = \min_{i\in\{2,\ldots,k\}} R_i(1,1)$ and $F^* = F(z^*)$, where F is given by (17) and R_i is the rate function. Specify a tolerance level $\Delta \in (0, F^*)$.
  Set $n = 1$ and $z_1 = \frac{1}{F^*-\Delta}$.
- **Step 2.** If $z_n > \tilde{R}_b \equiv \min_{i\in\{2,\ldots,k\}} \tilde{R}_i^b$, go to Step 4; otherwise, go to Step 3.
- **Step 3.** If $F(z_n) < F^*$, let $z^* = z_n$ and $F^* = F(z_n)$.
  Set $z_{n+1} = \frac{F(z_n)z_n}{F^*-\Delta}$ and $n = n + 1$, then go to step 2.
- **Step 4.** Calculate p_1^* from z^* via (18) and $p_i^* = p_1^* R_i^{-1}(z^*)$ for $i = 2, ..., k$.
- **Return** $\{p_i^*\}$.

4. Asymptotic Analysis

Since the procedure described in the previous section gives a solution $\mathbf{p}^* = (p_1^*, \ldots, p_k^*)$ to the approximate problem, instead of the solution $\tilde{\mathbf{p}}(\mathbf{n}) = (\tilde{p}_1(n), \ldots, \tilde{p}_k(n))$ to the original problem of minimizing the probability of false selection given by (1), we prove that $\tilde{\mathbf{p}}(\mathbf{n}) \to \mathbf{p}^*$ and characterize the rate of convergence under additional assumptions. Lemma 7 proves that $\mathbf{p}^*$ dominates any solution $\mathbf{p}(\mathbf{n}) = (p_1(n), \ldots, p_k(n))$, where $\mathbf{p}(\mathbf{n})$ converges to some limit $\mathbf{p}$. Any vector $\mathbf{p} \in \Re^k$ is referred to as an allocation vector if $p_i \ge 0$ for each i, $\sum_{i\le k} p_i = 1$ and

np_i is an integer for each i. Let $P_p(FS)$ denotes the probability of false selection under the allocation vector $\mathbf{p}$. The following result establishes that a finite budget allocation based on p^* dominates all other linear allocations in a certain sense asymptotically.

Lemma 7. *Consider a sequence of allocation vectors* $\mathbf{p(n)} = (p_1(n), \ldots, p_k(n))$ *for* $n \geq 1$*, and suppose that there exists a vector* $\mathbf{p}$ *such that*

$$\mathbf{p(n)} \to \mathbf{p}.$$

Under Assumptions 1 and 2, if $\mathbf{p} \neq \mathbf{p}^*$*, then*

$$\lim_{n\to\infty} \frac{\log(P_{p(n)}(FS)/P_{p^*(n)}(FS))}{n} > 0, \tag{22}$$

where $\mathbf{p}^*(\mathbf{n})$ *is an allocation vector such that* $\mathbf{p}^*(\mathbf{n}) \to \mathbf{p}^*$*.*

Proof: We split the proof into three cases: (i) $p_1 > 0$ and $p_i > 0$ for all $i \in \Omega$; (ii) $p_1 = 0$; (iii) $p_1 > 0$ and $p_i = 0$ for some $i \in \Omega$.

From Theorem 1, (i) follows. Now consider (ii). Without loss of generality, we further assume $p_2 > 0$. Assumption 1 implies that there exists $\varepsilon > 0$ such that $P(J_1 \leq \mu_2 - \varepsilon) > 0$. Letting $\alpha_2(n) = \frac{p_2(n)}{p_1(n)}$, for $n \geq n_0$ sufficiently large so that $p_2(n) > p_1(n)$ for all $n \geq n_0$,

$$\begin{aligned}
&\log P\left(\bar{J}_2(p_2(n)n) \geq \bar{J}_1(p_1(n)n)\right) \\
&\quad= \log P\left(\sum_{m=1}^{p_1(n)n} (J_{2m}/\alpha_2(n) - J_{1m}) + \sum_{m=p_1(n)n+1}^{p_2(n)n} J_{2m}/\alpha_2(n) > 0\right) \\
&\quad\geq \log P\left(\sum_{m=p_1(n)n+1}^{p_2(n)n} J_{2m}/\alpha_2(n) \geq p_1(n)n(\mu_2 - \varepsilon)\right) \\
&\qquad + \log P\left(\sum_{m=1}^{p_1(n)n} (J_{2m}/\alpha_2(n) - J_{1m}) \geq -p_1(n)n(\mu_2 - \varepsilon)\right) \\
&\quad\geq \log P\left(\frac{1}{p_2(n)n} \sum_{m=p_1(n)n+1}^{p_2(n)n} J_{2m} \geq (\mu_2 - \varepsilon)\right) \\
&\qquad + p_1(n)n \log P\left((J_2/\alpha_2(n) - J_1) \geq -(\mu_2 - \varepsilon)\right).
\end{aligned}$$

By the strong law of large numbers, the first term converges to zero, and

$$P\left((J_2/\alpha_2(n) - J_1) \geq -(\mu_2 - \varepsilon)\right) \longrightarrow P\left(-J_1 \geq -(\mu_2 - \varepsilon)\right).$$

Hence, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log P(\bar{J}_2(p_2(n)n) \geq \bar{J}_1(p_1(n)n) \geq p_1 \log P(J_1 \leq \mu_2 - \varepsilon) = 0,$$

which then implies that

$$\lim_{n\to\infty} \frac{1}{n} P_{p(n)}(FS) = 0 > \lim_{n\to\infty} \frac{1}{n} P_{p^*(n)}(FS).$$

Case (iii) can be treated by switching the roles of $p_2(n)$ and $p_1(n)$ in the arguments for case (ii). □

The following convergence result follows from Lemma 7:

Theorem 5. *Under Assumptions 1 and 2,*

$$\lim_{n\to\infty} \tilde{\mathbf{p}}(\mathbf{n}) = \mathbf{p}^*.$$

Proof. Suppose that $\tilde{\mathbf{p}}(\mathbf{n})$ does not converge to $\mathbf{p}^*$. Then we can find an increasing and divergent sequence $\{n_m, m = 1, \cdots\}$ and a vector $\mathbf{p} \neq \mathbf{p}^*$ such that

$$\lim_{n\to\infty} \tilde{\mathbf{p}}(\mathbf{n_m}) = \mathbf{p}.$$

Then from Lemma 7, we have that

$$\lim_{n\to\infty} \frac{\log(P_{\tilde{p}(n_m)}(FS)/P_{p^*(n)}(FS))}{n} > 0,$$

implying $P_{\tilde{p}(n_m)}(FS) > P_{p^*(n)}(FS)$ for n sufficiently large, which is a contradiction. □

Theorem 5 implies that $\tilde{\mathbf{p}}(\mathbf{n})$ differs from $\mathbf{p}^*$ by an $o(1)$ amount. In order to get a better idea of the rate of convergence of $\tilde{\mathbf{p}}(\mathbf{n})$ to $\mathbf{p}^*$, we need an additional assumption.

Assumption 3. *J_i, J_1, and $J_i - J_1$ are non-lattice random variables.*

Recall that a function $f(n)$ is said to be $\Theta(g(n))$ if there exist positive constants $K_1 < K_2$ such that $K_1 g(n) \leq f(n) \leq K_2 g(n)$ for all n sufficiently large. We have the following two lemmas, whose proofs are given in the Appendix.

Lemma 8. *Under Assumptions 1, 2, and 3,*

$$P_{\alpha(n)}(FS) = \Theta\big(\frac{1}{\sqrt{n}} \exp[-n \min_{2\leq i\leq k} R_i(\alpha_1(n), \alpha_i(n))]\big),$$

for any allocation vector $\alpha(\mathbf{n})$ that converges to a vector α as $n \to \infty$.

Lemma 9. *Under Assumptions 1, 2, and 3,*

$$\sup_{n>0} |nR_i(\widetilde{p}_1(n), \widetilde{p}_i(n)) - nR_i(p_1^*, p_i^*)| < \infty, \ \forall \ i \in \Omega.$$

Recall that the optimizer z^* to problem (16) has to satisfy one of three conditions in Theorem 3. Based on these conditions, we introduce the following definition.

Definition 2. *A solution z^* to problem (16) is called **Zero Fit** if any of the following conditions hold:*

(1) $F'(z^*) = 0$ and $z^* \in (0, R^b) \setminus \mathcal{S}_1$;

(2) $F'(z^*-)F'(z^*+) = 0$ and $z^* \in \mathcal{S}_1 \setminus (\mathcal{S}_2 \cup \{R^b\})$;

(3) $F'(z^*-) = 0$ and $z^* \in \mathcal{S}_1 \cap (\mathcal{S}_2 \cup \{R^b\})$.

The following theorem gives a more precise order of $\widetilde{p}_i(n) - p_i^*$ in terms of the budget n. Its proof is given in the Appendix.

Theorem 6. *Under Assumptions 1, 2, and 3, suppose there do not exist $1 < i < j \leq k$ such that $z^* = R_i(1) = R_j(1)$, then*

(i) If z^* is not Zero Fit, $\widetilde{p}_i(n) - p_i^* \sim O(\frac{1}{n})$ for any $1 \leq i \leq k$;

(ii) If z^* is Zero Fit and $F''(z^*\pm) \neq 0$, $\widetilde{p}_i(n) - p_i^* \sim O(\frac{1}{\sqrt{n}})$ for any $1 \leq i \leq k$.

5. Conclusions

In this paper, we considered the ordinal optimization problem when the output from different populations is dependent and has a general (not necessarily Gaussian) distribution. By using the large deviations framework to identify the large deviations rate function of the probability of false selection, we obtain an optimization problem that is far more complex than the one obtained in the independent and/or Gaussian setting. However, by analyzing the structure of this problem, we are able to reduce it to a single variable nonlinear optimization problem and develop a numerical algorithm to approximately solve it. In addition to proving that the solution to this approximate optimization problem is in fact the asymptotic limit of the solution to the original optimal budget allocation problem, we are also able to characterize the convergence *rate* of the approximate solution under additional conditions. Both the analysis used and the results obtained are novel to this general ordinal optimization setting.

Acknowledgments

Michael Fu and Xiaoping Xiong were supported in part by the National Science Foundation under Grants DMI-9988867 and DMI-0323220, and by the Air Force Office of Scientific Research under Grant FA95500410210.

Appendix

In this appendix, we provide the proofs of Lemmas 4, 8, 9, Theorem 6, and further details of the analysis for Examples 2 and 3 from Section 3.2.

Proof of Lemma 4: First consider $p > 1$. For $\epsilon > 0$ and sufficiently small,

$$\Lambda'_{(i,1,p)}(\gamma_i + \epsilon) =$$

$$\Lambda_i^{(0,1)}(0, \frac{\gamma_i + \epsilon}{p}) - \Lambda_i^{(1,0)}(-(\gamma_i + \epsilon), \frac{\gamma_i + \epsilon}{p}) + \frac{1}{p}\left(\Lambda_i^{(0,1)}(-(\gamma_i + \epsilon), \frac{\gamma_i + \epsilon}{p}) - \Lambda_i^{(0,1)}(0, \frac{\gamma_i + \epsilon}{p})\right).$$

Note that $\Lambda_i^{(0,1)}(0, \frac{\gamma_i+\epsilon}{p}) \to \mu_i$ as $p \to \infty$, and

$$\Lambda_i^{(1,0)}(-(\gamma_i + \epsilon), \frac{\gamma_i + \epsilon}{p}) \to \Lambda_i^{(1,0)}(-(\gamma_i + \epsilon), 0) < \mu_i.$$

Therefore, $\Lambda'_{(i,1,p)}(\gamma_i + \epsilon) > 0$ for all p sufficiently large. Similarly it follows that $\Lambda'_{(i,p_1,p_i)}(\gamma_i - \epsilon) < 0$ for all p sufficiently large. Since ϵ is arbitrary, (11) follows.

To see (12), consider $p < 1$. For $\epsilon > 0$ and sufficiently small,

$$\Lambda'_{(i,1,p)}((\beta_i + \epsilon)p) =$$

$$\Lambda_i^{(0,1)}(-(\beta_i + \epsilon)p, \beta_i + \epsilon) - p\Lambda_i^{(1,0)}(-(\beta_i + \epsilon)p, \beta_i + \epsilon) - (1 - p)\Lambda_i^{(1,0)}(-(\beta_i + \epsilon)p, 0).$$

Note that as $p \to 0$,

$$\Lambda_i^{(0,1)}(-(\beta_i + \epsilon)p, \beta_i + \epsilon) \to \Lambda_i^{(0,1)}(0, \beta_i + \epsilon) > \mu_1,$$

and $\Lambda_i^{(1,0)}(-(\beta_i + \epsilon)p, 0) \to \mu_1$. Hence, $\Lambda'_{(i,1,p)}((\beta_i + \epsilon)p) > 0$ for p sufficiently large. Similarly, it can be seen that $\Lambda'_{(i,1,p)}((\beta_i - \epsilon)p) < 0$ for p sufficiently large, and (12) follows.

To see (13), note that for $p > 1$,

$$R_i(p) = -\Lambda_i(-\lambda_i(p), \frac{\lambda_i(p)}{p}) - (p - 1)\Lambda_i(0, \frac{\lambda_i(p)}{p}).$$

Now as $p \to \infty$,

$$\Lambda_i(-\lambda_i(p), \frac{\lambda_i(p)}{p}) \to \Lambda_i(-\gamma_i, 0).$$

Using Taylor's Theorem,

$$\Lambda_i(0, \frac{\lambda_i(p)}{p}) = \frac{\lambda_i(p)}{p}\Lambda_i^{(0,1)}(0,0) + o(1/p).$$

Therefore,

$$R_i(\infty) = -\Lambda_i(-\gamma_i, 0) - \gamma_i\mu_i.$$

To see that RHS equals $\inf_x I_i(\mu_i, x)$, note that the latter equals

$$\inf_x \sup_{\lambda_1,\lambda_2} (\lambda_1\mu_i + \lambda_2 x - \Lambda_i(\lambda_1, \lambda_2)).$$

Using the min-max theorem, this in turn equals

$$\begin{aligned} &\sup_{\lambda_1,\lambda_2} \inf_x (\lambda_1\mu_i + \lambda_2 x - \Lambda_i(\lambda_1, \lambda_2)) \\ &= \sup_{\lambda_1} (\lambda_1\mu_i - \Lambda_i(\lambda_1, 0)). \end{aligned}$$

The result now follows since $\Lambda_i'(-\gamma_i, 0) = \mu_i$.

To see (14), note that $R_i'(p) = f_i^{(0,1)}(\lambda_i(p), p)$ equals

$$\frac{\lambda_i(p)}{p^2}\Lambda_i^{(0,1)}(-\lambda_i(p), \frac{\lambda_i(p)}{p}) - \Lambda_i(0, \frac{\lambda_i(p)}{p}) + \frac{(p-1)\lambda_i(p)}{p^2}\Lambda_i^{(0,1)}(0, \frac{\lambda_i(p)}{p}). \tag{23}$$

From the fact that $f_i^{(1,0)}(\lambda_i(p), p) = 0$, it follows that

$$\frac{1}{p}\Lambda_i^{(0,1)}(-\lambda_i(p), \frac{\lambda_i(p)}{p}) = \Lambda_i^{(1,0)}(-\lambda_i(p), \frac{\lambda_i(p)}{p}) - \frac{p-1}{p}\Lambda_i^{(0,1)}(0, \frac{\lambda_i(p)}{p}).$$

Plugging this in (23),

$$R_i'(p) = \frac{\lambda_i(p)}{p}\Lambda_i^{(1,0)}(-\lambda_i(p), \frac{\lambda_i(p)}{p}) - \Lambda_i(0, \frac{\lambda_i(p)}{p}).$$

Equation (14) follows from this and (11).

To see that $R_i(0^+) = 0$, note that for $p < 1$,

$$R_i(p) = -p\Lambda_i(-\lambda_i(p), \frac{\lambda_i(p)}{p}) - (1-p)\Lambda_i(-\lambda_i(p), 0).$$

The result now follows from (12). □

Theorem 3.3 in Chaganty and Sethuraman 1993 is useful in proving Lemma 8. We reproduce its statement here to help the reader follow the proof of Lemma 8.

Theorem 7 (Chaganty and Sethuraman 1993). *Let T_n be a sequence of non-lattice valued random variables with m.g.f. $\phi_n(z) = E\exp(zT_n)$, which is nonvanishing and analytic in the region $\Psi = \{z \in \mathcal{C}, |z| < a\}$, where $a > 0$ and $\mathcal{C}$ is the set of all complex numbers. Let $\{a_n\}$ be a sequence of real numbers. Let $\psi_n(z) = \frac{\log \phi_n(z)}{a_n}$ and $\gamma_n(u) = \sup_{|s|<a, s\in\mathcal{R}}[us - \psi_n(s)]$, for $u \in \mathcal{R}$. Let $\{m_n\}$ be a bounded sequence of real numbers and $\{\tau_n\}$ be a sequence satisfying $\psi_n'(\tau_n) = m_n$ and $0 < \tau_n < a_0 < a$, for all n. Suppose that $a_n \to \infty$ such that $\tau_n\sqrt{a_n} \to \infty$. Further, assume that T_n satisfies the following three conditions:*

(a) There exists $\beta < \infty$ such that $|\psi_n(z)| < \beta$ for all n and $z \in \Psi$;

(b) There exists $\gamma > 0$ such that $\psi_n''(\tau_n) \geq \gamma$ for all n;

(c) There exists $\delta_0 > 0$ such that, for any given δ and δ' such that $0 < \delta < \delta_0 < \delta'$,

$$\sup_{\delta < |t| < \delta' \tau_n} \left|\frac{\phi_n(\tau_n + t\sqrt{-1})}{\phi_n(\tau_n)}\right| = o\left(\frac{1}{\sqrt{n}}\right). \tag{24}$$

Then

$$P\left(\frac{T_n}{a_n} \geq m_n\right) \sim \frac{1}{\tau_n\sqrt{2\pi a_n \psi_n''(\tau_n)}} \exp(-a_n\gamma_n(m_m)).$$

Proof of Lemma 8: In Theorem 7, set $T_n = n\bar{J}_i(\alpha_i(n)) - n\bar{J}_1(\alpha_1(n))$, $m_n = 0$, $a_n = n$, $\psi_n(\lambda) = \Lambda_{(i,\alpha_1(n),\alpha_i(n))}(\lambda)$, $\phi_n(\lambda) = \exp\{n\psi_n(\lambda)\}$, and τ_n be the solution to $\psi_n'(\lambda) = 0$. We also let $\psi(\lambda) = \Lambda_{(i,\alpha_1,\alpha_i)}(\lambda)$ and τ be the solution to $\psi'(\lambda) = 0$. Then there exists $a > \tau$ such that $\psi_n(\lambda)$ is well defined on $[0, a]$ for any sufficiently large n. We denote as Ψ the set $\{z \in \mathcal{C} : |z| \leq a, Re(z) \geq 0\}$. Now we only need to verify the conditions (a), (b), (c) in Theorem 7 above.

Obviously we have $\tau_n \to \tau$ and $\psi_n(z) \to \psi(z)$, $\forall\ z \in \Psi$. Since ψ_n and ψ are uniformly continuous on compact set Ψ, $\psi_n(z)$ uniformly converges to $\psi(z)$ and thus (a) holds. The convexity of $\psi(\lambda)$ and $\psi_n''(\tau_n) \to \psi(\tau)$ implies condition (b). To proceed, we denote as $\chi(\lambda; \xi)$ the moment-generating function, i.e., $\chi(\lambda; \xi) = Ee^{\lambda\xi}$. Obviously, $\frac{\chi(\lambda + t\sqrt{-1}; \xi)}{\chi(\lambda; \xi)}$ is the characteristic function of a random variable, say η, hence, we have

$$\left|\frac{\chi(\lambda + t\sqrt{-1}; \xi)}{\chi(\lambda; \xi)}\right| = \left|\chi(t\sqrt{-1}; \eta)\right| \leq 1.$$

In particular, if ξ is non-lattice, so is η. Now we suppose $\alpha_1 > \alpha_i$,

$$
\begin{aligned}
\left|\frac{\phi_n(\tau_n+t\sqrt{-1})}{\phi_n(\tau_n)}\right| &= \left|\frac{\chi(\tau_n+t\sqrt{-1};\frac{J_i}{\alpha_i(n)}-\frac{J_1}{\alpha_1(n)})}{\chi(\tau_n;\frac{J_i}{\alpha_i(n)}-\frac{J_1}{\alpha_1(n)})}\right|^{n\alpha_i(n)}\cdot\left|\frac{\chi(\tau_n+t\sqrt{-1};-\frac{J_1}{\alpha_1(n)})}{\chi(\tau_n;-\frac{J_1}{\alpha_1(n)})}\right|^{n(\alpha_1(n)-\alpha_i(n))} \\
&\leq \left|\frac{\chi(\tau_n+t\sqrt{-1};-\frac{J_1}{\alpha_1(n)})}{\chi(\tau_n;-\frac{J_1}{\alpha_1(n)})}\right|^{n(\alpha_1(n)-\alpha_i(n))}.
\end{aligned}
$$

Assumption 3 implies that the RHS can be taken as c.f. of a non-lattice random variable. To simplify, we denote as $\chi_n(t)$ the quantity in the brackets $|\ |$ on RHS of above formula. Since $\tau_n \to \tau$, we have $\tau' = \sup_n \tau_n < \infty$. We choose $\delta_0 = 1$ and fix δ and δ', we have

$$\sup_{\delta\leq|t|\leq\delta'\tau'} |\chi_n(t)| < 1.$$

Noticing that $\chi_n(t)$ uniformly converges to $\frac{\chi(\tau+t\sqrt{-1};-\frac{J_1}{\alpha_1})}{\chi(\tau;-\frac{J_1}{\alpha_1})}$ on a compact set $[\delta, \delta'\tau']$, we also have

$$x = \sup_n \sup_{\delta\leq|t|\leq\delta'\tau'} |\chi_n(t)| < 1,$$

which further implies that

$$\sup_{\delta\leq|t|\leq\delta'\tau_n} \left|\frac{\phi_n(\tau_n+t\sqrt{-1})}{\phi_n(\tau_n)}\right| \leq x^{n(\alpha_1(n)-\alpha_i(n))} = O(\exp[n\log(x)(\alpha_1-\alpha_i)]) = o(\frac{1}{\sqrt{n}}).$$

Similarly we can show (24) also holds when $\alpha_i \leq \alpha_1$, which completes the proof. □

Proof of Lemma 9: To simplify notation, we let

$$
\begin{aligned}
\widetilde{R}(n) &= \min_{2\leq i\leq k} R_i(\widetilde{p}_1(n), \widetilde{p}_i(n));\ \ \widetilde{R}_i(n) = R_i(\widetilde{p}_1(n), \widetilde{p}_i(n)); \\
R^*(n) &= \min_{2\leq i\leq k} R_i(p_1^*(n), p_i^*(n));\ \ R_i^*(n) = R_i(p_1^*(n), p_i^*(n));
\end{aligned}
$$

where $(p_i^*(n) : i \leq k)$ is obtained by suitably rounding of each p_i^* to a multiple of $1/n$ so that $|p_i^*(n) - p_i^*| \leq O(1/n)$, each $p_i^*(n) > 0$ and $\sum_{i=1}^k p_i^*(n) = 1$. If $R^* = \min_{2\leq i\leq k} R_i(p_1^*, p_i^*);\ R_i^* = R_i(p_1^*, p_i^*)$, we have

$$|R^*(n) - R^*| = O(1/n), \tag{25}$$

$$|R_i^*(n) - R_i^*| = O(1/n). \tag{26}$$

To show (26), we represent partial derivative with superscripts and write

$$\begin{aligned}
\left|R_i^*(n) - R_i^*\right| &= \left|R_i(p_1^*(n), p_i^*(n)) - R_i(p_1^*, p_i^*)\right| \\
&\leq \left|R_i(p_1^*(n), p_i^*(n)) - R_i(p_1^*, p_i^*(n))\right| + \left|R_i(p_1^*, p_i^*(n)) - R_i(p_1^*, p_i^*)\right| \\
&= \left|R_i^{(1,0)}(p_1^\theta, p_i^*(n))(p_1^*(n) - p_1^*)\right| + \left|R_i^{(0,1)}(p_1^*, p_i^\theta)(p_i^*(n) - p_i^*)\right| \\
&\leq K\left|(p_1^*(n) - p_1^*)\right| + K\left|(p_i^*(n) - p_i^*)\right| \\
&= O(\frac{1}{n}),
\end{aligned}$$

where the third step uses the mean value theorem with p_1^θ between $p_1^*(n)$ and p_1^* and p_i^θ between $p_i^*(n)$ and p_i^*, the fourth step uses boundedness of partial derivatives, and the last uses $|p_i^*(n) - p_i^*| \leq O(\frac{1}{n})$. To show (25), we notice that $R_i^* = R^*$ and thus,

$$|R^*(n) - R^*| = |\min_{i\in\Omega}[R_i^*(n) - R^*]| = |\min_{i\in\Omega}[R_i^*(n) - R_i^*]| \leq \max_{i\in\Omega}|R_i^*(n) - R_i^*| \approx O(\frac{1}{n}).$$

From Lemma 8 it follows that

$$\begin{aligned}
P_{\tilde{p}(n)}(FS) &= \Theta\big(\frac{1}{\sqrt{n}}\exp(-n\widetilde{R}(n)\big), \\
P_{p^*(n)}(FS) &= \Theta\big(\frac{1}{\sqrt{n}}\exp(-nR^*(n)\big).
\end{aligned}$$

Hence, we have

$$C_1 - \frac{1}{2}\log n - nR^*(n) \geq \log P_{p^*(n)}(FS) \geq \log P_{\tilde{p}(n)}(FS) \geq C_2 - \frac{1}{2}\log n - n\widetilde{R}(n), \quad (27)$$

where C_1 and C_2 are constants independent of n. We can replace $R^*(n)$ with R^* on the left hand side by using (25). The fact that R^* is the value of problem (4) implies $R^* \geq \widetilde{R}(n)$. Combining this with the above inequality and (25), we see that

$$\sup_{n>0} |\, n\widetilde{R}(n) - nR^*| < C_1 - C_2 + O(1) = K < \infty, \quad (28)$$

where K is another constant. Obviously $\widetilde{R}_i(n) \geq \widetilde{R}(n)$ for all $i \in \Omega$. In addition, Part 2 of Theorem 2 implies $R_i^* = R^*$ for all $i \in \Omega$. Then $n\widetilde{R}_i(n) - nR_i^* \geq n\widetilde{R}(n) - nR^* \geq -K$ by (28).

We define a new vector $\mathbf{p(n)}$ with:

$$p_i(n) = \begin{cases} \widetilde{p}_i(n) + 2K\frac{\widetilde{p}_i(n)}{nR^*}, & i \neq 2, \\ \widetilde{p}_i(n) - 2K\frac{1-\widetilde{p}_i(n)}{nR^*}, & i = 2. \end{cases} \quad (29)$$

Obviously, $\mathbf{p}(\mathbf{n})$ satisfies the budget constraint. Given sufficiently large n, each $p_i(n)$ is positive and $\widetilde{R}_i(n) \geq \widetilde{R}(n) > R^* - \frac{K}{n} > R^*/2$ for all $i \in \Omega$. We recall that the optimality of $\{p_i^*\}$ to problem (4) is independent of the value of n. So we always have, for arbitrary n,

$$\min_{i\in\Omega} R_i(p_1(n), p_i(n)) \leq \min_{i\in\Omega} R_i(p_1^*, p_i^*) = R^*. \tag{30}$$

Note that $\frac{p_i}{p_1} = \frac{\widetilde{p}_i(n)}{\widetilde{p}_1(n)}$ for $i > 2$. Combining this with homogeneity of R_i, and noting (29) we have, for $i > 2$,

$$\begin{aligned} nR_i(p_1(n), p_i(n)) &= \frac{p_1(n)}{\widetilde{p}_1(n)} nR_i(\widetilde{p}_1(n), \widetilde{p}_i(n)) = n\widetilde{R}_i(n) + 2K\frac{\widetilde{R}_i(n)}{R^*} \\ &> n\widetilde{R}_i(n) + K \geq n\widetilde{R}(n) + K \geq nR^*, \end{aligned} \tag{31}$$

where the last step uses (28).

Now $nR_i(p_1(n), p_i(n)) > nR^*$ for all $i > 2$. If this inequality also holds for $i = 2$, we will lead to a contradiction to (30). Hence, $nR_2(p_1(n), p_2(n)) \leq nR^* = nR_2^*$. Using superscripts to denote partial derivatives of $R_2(\cdot, \cdot)$, we have

$$\begin{aligned} n\widetilde{R}_2(n) - nR_2^* &= nR_2(\widetilde{p}_1(n), \widetilde{p}_2(n)) - nR_2^* \\ &\leq nR_2(\widetilde{p}_1(n), \widetilde{p}_2(n)) - nR_2(p_1(n), p_2(n)) \\ &= [nR_2(\widetilde{p}_1(n), \widetilde{p}_2(n)) - nR_2(p_1(n), \widetilde{p}_2(n))] \\ &\qquad + [nR_2(p_1(n), \widetilde{p}_2(n)) - nR_2(p_1(n), p_2(n))] \\ &= nR_2^{(1,0)}(p_1^\theta, \widetilde{p}_2(n))[\widetilde{p}_1(n) - p_1(n)] + nR_2^{(0,1)}(p_1(n), p_2^\theta)[\widetilde{p}_2(n) - p_2(n)] \\ &= -2K\frac{\widetilde{p}_1(n)}{R^*} R_2^{(1,0)}(p_1^\theta, \widetilde{p}_2(n)) + 2K\frac{1 - \widetilde{p}_2(n)}{R^*} R_2^{(0,1)}(p_1(n), p_2^\theta), \end{aligned}$$

where the second step uses $nR_2(p_1(n), p_2(n)) \leq nR_2^*$, the fourth step uses mean value theorem with p_1^θ falling in between $\widetilde{p}_1(n)$ and $p_1(n)$ and p_2^θ falling in between $\widetilde{p}_2(n)$ and $p_2(n)$. Since $\widetilde{p}_i(n)$ and $p_i(n)$ all converge to p_i^*, it is easy to verify boundedness of each term on the RHS. So $n\widetilde{R}_2(n) - nR_2^*$ is also bounded from above, and thus part (ii) is true when $i = 2$. Obviously, we can extend the arguments to all $i \in \Omega$ and complete the proof.

Additional notation is needed to help in proving Theorem 6. We write $n\widetilde{p}_i(n) - np_i^*$ as $\Delta_i(n)$, $\frac{\widetilde{p}_i(n)}{\widetilde{p}_1(n)}$ as $\widetilde{q}_i(n)$, and $\frac{p_i^*}{p_1^*}$ as q_i^*. We will also denote as $X(n) \in UBD$ if some series $X(n)$, depending on n, is uniformly bounded from above and below for any large n. We first show the following lemma:

Lemma 10. *Under Assumptions 1, 2, and 3,* $\Delta_1(n) \in UBD \Longrightarrow \Delta_i(n) \in UBD, \ \forall\ i \in \Omega$.

Proof: Recall that for (p_1, p_i), and α, $R_i(\alpha p_1, \alpha p_i) = \alpha R_i(p_1, p_i)$, i.e., R_i is a homogeneous function. Furthermore $R_i(\alpha) = R_i(1, \alpha)$. Therefore,

$$nR_i(\widetilde{p}_1(n), \widetilde{p}_i(n)) - nR_i(p_1^*, p_i^*) = n\widetilde{p}_1(n)R_i(\widetilde{q}_i(n)) - np_1^* R_i(q_i^*). \tag{32}$$

Through a Taylor series expansion, we have

$$R_i(\widetilde{q}_i(n)) = R_i(q_i^*) + (\widetilde{q}_i(n) - q_i^*)R_i'(V_i(n)),$$

where $V_i(n)$ lies between $\widetilde{q}_i(n)$ and q_i^* and R_i' represents the right (left) derivative if $\widetilde{q}_i(n)$ is greater (less) than q_i^*. Then, the RHS in (32) may be re-expressed as

$$R_i(q_i^*)\Delta_1(n) + nR_i'(V_i(n))(\widetilde{p}_i(n) - q_i^*\widetilde{p}_1(n)) = [R_i(q_i^*) - q_i^* R_i'(V_i(n))]\Delta_1(n) + R_i'(V_i(n))\Delta_i(n). \tag{33}$$

We observe that the coefficients of $\Delta_1(n)$ and $\Delta_i(n)$ converge to $R_i(q_i^*) - q_i^* R_i'(q_i^*)$ and $R_i'(q_i^*)$, respectively, as $n \to \infty$.

Note that $nR_i(\widetilde{p}_1(n), \widetilde{p}_i(n)) - nR_i(p_1^*, p_i^*)$, and hence the right hand side of (33) is bounded by Lemma 8(ii). Obviously Lemma 10 is true if $R_i'(q_i^*) \neq 0$. Now suppose $R_i'(q_i^*) = 0$ for some $i \in \Omega$; then $q_i^* \geq 1$, because $R_i'(\cdot)$ is positive on $(0, 1]$ by Lemma 2. If we further suppose $q_i^* > 1$, then $R_i'(q_i^*) = 0$ implies $F'(z^*-) = \infty$, where F' can be inferred from (20). But since z^* is optimal to (19), Theorem 3 requires $F'(z^*-) \leq 0$, a contradiction. So we may only have $q_i^* = 1$, i.e., $p_i^* = p_1^*$. We recall the following:

- $V_i(n)$ falls between $\widetilde{q}_i(n)$ and $q_i^* = 1$.
- $R_i'(V_i(n)) \to R_i'(1+) = 0$ when $\widetilde{q}_i(n) > q_i^* = 1$, whereas $R_i'(V_i(n)) \to R_i'(1-) > 0$ when $\widetilde{q}_i(n) < q_i^* = 1$.

Therefore, assuming $R_i'(q_i^*) = 0$ implies $V_i(n)$ has to converge to $q_i^* = 1$ from the right, and thus $\widetilde{q}_i(n) \geq V_i(n) \geq q_i^* = 1$. Now we have

$$\Delta_i(n) = n\widetilde{p}_i(n) - np_i^* \geq n\widetilde{p}_1(n) - np_i^* = n\widetilde{p}_1(n) - np_1^* = \Delta_1(n).$$

We can complete the proof by using $\sum_{i=1}^k \Delta_i(n) = 0$. Thus,

$$\Delta_1(n) \in UBD \Longrightarrow \Delta_i(n) \in UBD.$$

□

Proof of Theorem 6: We show part (i) by considering two cases:

I. $R_i'(q_i^*) = 0$ for some $i \in \Omega$;

II. $R_i'(q_i^*) \neq 0$ for all $i \in \Omega$.

As argued in the proof of Lemma 10, $R_i'(q_i^*) = 0$ for some i only if $q_i^* = 1$. The condition of this theorem then implies there is at most one i, say $i = 2$, such that $R_2'(q_2^*) = 0$ in case I.

Recall that both sides of (33) are uniformly bounded for all i, i.e.,

$$[R_i(q_i^*) - q_i^* R_i'(V_i(n))]\Delta_1(n) + R_i'(V_i(n))\Delta_i(n) \in UBD. \tag{34}$$

Since $R_i'(V_i(n))$ converges to a non-zero value for $i > 2$, we have, for $i > 2$,

$$[\frac{R_i(q_i^*)}{R_i'(V_i(n))} - q_i^*]\Delta_1(n) + \Delta_i(n) \in UBD.$$

Hence,

$$\sum_{i=3}^{k}[\frac{R_i(q_i^*)}{R_i'(V_i(n))} - q_i^*]\Delta_1(n) + \sum_{i=3}^{k}\Delta_i(n) \in UBD.$$

Since $R_2'(V_2(n)) \to 0$, $R_2'(V_2(n)) \in UBD$ and

$$\sum_{i=3}^{k}[\frac{R_i(q_i^*)}{R_i'(V_i(n))} - q_i^*]R_2'(V_2(n))\Delta_1(n) + \sum_{i=3}^{k} R_2'(V_2(n))\Delta_i(n) \in UBD.$$

Combining this with (34) for $i = 2$, we have

$$\left[R_2(q_2^*) + R_2'(V_2(n))\left(\sum_{i=3}^{k}[\frac{R_i(q_i^*)}{R_i'(V_i)} - q_i^*] - q_2^*\right)\right]\Delta_1(n) + R_2'(V_2(n))\sum_{i=2}^{k}\Delta_i(n) \in UBD.$$

Since $\sum_{i=2}^{k}\Delta_i(n) = 0$ and the coefficient of $\Delta_1(n)$ in the above item converges to $R_2(q_2^*) > 0$, $\Delta_1(n)$ is bounded. Part (i) for case I of the theorem follows from Lemma 10.

For case II, dividing RHS of (33) by $R_i'(V_i(n))$ (this is positive for n sufficiently large) implies, for $i \in \Omega$,

$$[\frac{R_i(q_i^*)}{R_i'(V_i(n))} - q_i^*]\Delta_1(n) + \Delta_i(n) \in UBD. \tag{35}$$

Therefore, their sum over all $i \in \Omega$,

$$\sum_{i=2}^{k}[\frac{R_i(q_i^*)}{R_i'(V_i(n))} - q_i^*]\Delta_1(n) + \sum_{i=2}^{k}\Delta_i(n) = \left(\sum_{i=2}^{k}[\frac{R_i(q_i^*)}{R_i'(V_i(n))} - q_i^*] - 1\right)\Delta_1(n) \in UBD.$$

If we denote as $\widetilde{B}(n)$ the coefficient of $\Delta_1(n)$ in the RHS above, we have

$$\widetilde{B}(n)\Delta_1(n) \in UBD. \tag{36}$$

and

$$\widetilde{B}(n) \to B^* \equiv \sum_{i=2}^{k} \Big[\frac{R_i^*(q_i^*)}{R_i'(q_i^*)} - q_i^*\Big] - 1. \tag{37}$$

To show that $\Delta_1(n) \in UBD$, we now show that $B^* \neq 0$. Suppose $z^* = R_i(1)$ for some i, say $i = 2$. Then the condition of the theorem implies $z^* \neq R_j(1)$ for all $j > 2$. In other words, $q_j^* \neq 1$ for $j > 2$ and $q_2^* = 1$. Hence, for $j > 2$, R_j is differentiable at q_j^* and thus $R_j'(q_j^*) = R_j'(q_j^*+) = R_j'(q_j^*-)$. Now we further divide case II into two possibilities:

IIa: $\widetilde{q}_2(n) \geq 1$ and $R_2'(1+) < 0$.

We have from (20),

$$\begin{aligned}
(z^*)^2 F'(z^*-) &= \sum_{i=2}^{k} \frac{z^*}{R_i'(R_i^{-1}(z^*)-)} - 1 - \sum_{i=2}^{k} R_i^{-1}(z^*) \\
&= \sum_{i=2}^{k} \frac{z^*}{R_i'(q_i^*-)} - 1 - \sum_{i=2}^{k} q_i^* \\
&= \sum_{i=3}^{k} \frac{z^*}{R_i'(q_i^*)} + \frac{z^*}{R_2'(1-)} - 1 - \sum_{i=2}^{k} q_i^* \\
&> \sum_{i=3}^{k} \frac{z^*}{R_i'(q_i^*)} + \frac{z^*}{R_2'(1+)} - 1 - \sum_{i=2}^{k} q_i^* = B^*,
\end{aligned}$$

where the inequality uses $R_2'(1-) > 0 > R_2'(1+)$. Now $F'(z^*-) \leq 0$, by Theorem 3, implies $B^* < 0$, and thus $\Delta_1(n) \in UBD$ and $\Delta_i(n) \in UBD$.

IIb: $R_i'(1+) \geq 0$.

We have again from (20),

$$\begin{aligned}
(z^*)^2 F'(z^*\pm) &= \sum_{i=2}^{k} \frac{z^*}{R_i'(R_i^{-1}(z^*)\pm)} - 1 - \sum_{i=2}^{k} R_i^{-1}(z^*) \\
&= \sum_{i=2}^{k} \frac{z^*}{R_i'(q_i^*\pm)} - 1 - \sum_{i=2}^{k} q_i^* \\
&= \sum_{i=3}^{k} \frac{z^*}{R_i'(q_i^*)} + \frac{z^*}{R_2'(1\pm)} - 1 - \sum_{i=2}^{k} q_i^*.
\end{aligned}$$

Now we have either $B^* = (z^*)^2 F'(z^*+)$ when $\widetilde{q}_2(n) \geq 1$, or $B^* = (z^*)^2 F'(z^*-)$ when $\widetilde{q}_2(n) < 1$. By the definition of **Zero Fit**, that z^* is not Zero Fit implies $B^* \neq 0$, and thus $\Delta_1(n) \in UBD$ and $\Delta_i(n) \in UBD$ for case II.

To clarify the proof, we emphasize that in case IIa, we do not necessarily have $B^* = (z^*)^2 F'(z^*+)$ or $B^* = (z^*)^2 F'(z^*-)$, so $B^* \neq 0$ cannot be directly deduced from the definition of Not Zero Fit as in case IIb.

Now we assume z^* is Zero Fit. If $B^* \neq 0$, (36) and (37) then imply $\Delta_1(n) \in UBD$ and thus part (ii) will automatically hold. Now we can focus on $B^* = 0$. Note that

$$R_i'(V_i(n)) = \big(R_i(\widetilde{q}_i(n)) - R_i(q_i^*)\big)/(\widetilde{q}_i(n) - q_i^*). \tag{38}$$

If we use second-order Taylor expansion of $R_i(\widetilde{q}_i(n))$ around q_i^*, we have

$$R_i'(V_i(n)) - R_i'(q_i^*) = \frac{1}{2}(\widetilde{q}_i(n) - q_i^*)R_i''(U_i(n)), \tag{39}$$

where $U_i(n)$ lies between $\widetilde{q}_i(n)$ and q_i^*, and R_i'' represents the right (left) derivative if $\frac{\widetilde{p}_i(n)}{\widetilde{p}_1}$ is greater (less) than q_i^*. Simple algebra yields

$$\begin{aligned}\frac{\Delta_i(n) - q_i^*\Delta_1(n)}{n\widetilde{p}_1} &= \frac{n\widetilde{p}_i(n) - np_i^* - q_i^*(n\widetilde{p}_1(n) - np_1^*)}{n\widetilde{p}_1(n)} = \widetilde{q}_i(n) - \frac{np_i^* + nq_i^*\widetilde{p}_1(n) - np_1^*q_i^*}{n\widetilde{p}_1(n)} \\ &= \widetilde{q}_i(n) - \frac{np_i^* + nq_i^*\widetilde{p}_1(n) - np_i^*}{n\widetilde{p}_1(n)} = \widetilde{q}_i(n) - q_i^*. \end{aligned} \tag{40}$$

Letting $W_i(n) = [R_i(q_i^*)R_i''(U_i(n))]/[R_i'(V_i(n))R_i'(q_i^*)]$, we can rewrite (36) as

$$\begin{aligned} &\widetilde{B}(n)\Delta_1(n) = \widetilde{B}(n)\Delta_1(n) - B^*\Delta_1(n) \\ =\ &\Big(\sum_{i=2}^k \big[\frac{R_i(q_i^*)}{R_i'(V_i(n))} - q_i^*\big] - 1\Big)\Delta_1(n) + \Big(\sum_{i=2}^k \big[\frac{\frac{R_i^*}{p_1^*}}{R_i'(q_i^*)} - q_i^*\big] - 1\Big)\Delta_1(n) \\ =\ &\sum_{i=2}^k \big[\frac{R_i(q_i^*)}{R_i'(V_i(n))} - \frac{R_i(q_i^*)}{R_i'(q_i^*)}\big]\Delta_1(n) \\ =\ &\sum_{i=2}^k \big[R_i'(q_i^*) - R_i'(V_i(n))\big]\frac{R_i(q_i^*)}{R_i'(V_i(n))R_i'(q_i^*)}\Delta_1(n) \\ =\ &-\frac{1}{2}\sum_{i=2}^k \big[\widetilde{q}_i(n) - q_i^*\big]W_i(n)\Delta_1(n) \\ =\ &-\frac{1}{2}\sum_{i=2}^k \frac{\Delta_1(n)}{n\widetilde{p}_1}\big[\Delta_i(n) - q_i^*\Delta_1(n)\big]W_i(n) \\ =\ &-\frac{1}{2}\sum_{i=2}^k \frac{\Delta_1(n)}{n\widetilde{p}_1(n)}\big[\Delta_i(n) + (\frac{R_i(q_i^*)}{R_i'(V_i(n))} - q_i^*)\Delta_1(n)\big]W_i(n) + \frac{(\Delta_1(n))^2}{2n\widetilde{p}_1(n)}\sum_{i=2}^k \frac{R_i(q_i^*)}{R_i'(V_i(n))}W_i(n), \end{aligned}$$

where the first equality uses $B^* = 0$, the second uses (37), the fifth step uses (39), the sixth uses (40). Combining (35) and $\frac{\Delta_1(n)}{n\widetilde{p}_1(n)} = \frac{\widetilde{p}_1(n) - p_1^*}{\widetilde{p}_1(n)} \to 0$ implies that the first term on the right hand side converges to zero. So (36) implies boundedness of the second term. Actually the summation in the second term converges to

$$\sum_{i=2}^{k} \frac{(R_i(q_i^*))^2 R_i''(q_i^*)}{(R_i'(q_i^*))^3} = z^* \sum_{i=2}^{k} \frac{z^* R_i''(R_i^{-1}(z^*))}{(R_i'(R_i^{-1}(z^*)))^3} = -z^* M'(z^*),$$

where the first equality uses $R_i(q_i^*) = z^*$, and the second uses $M(\cdot)$ defined in (20), and thus

$$M'(z) \equiv -\sum_{i=2}^{k} \frac{z R_i''(R_i^{-1}(z))}{(R_i'(R_i^{-1}(z)))^3}.$$

The condition $F''(z^*\pm) \neq 0$ of part (ii) implies that $M'(z^*) \neq 0$, and thus $\frac{(\Delta_1(n))^2}{2np_1^*} \in UBD$, or $\Delta_1(n) = O(\sqrt{n})$. We can also conclude that $\Delta_i(n) = O(\sqrt{n})$ by using (35). Part (ii) is established. □

Proof that $R_i(\cdot)$ is quasi-convex in Example 2 of Section 3.2. It suffices to show $R_i''(p) \geq 0$ when $R_i'(p) = 0$. Recall that $\lambda_i(p)$ is the unique solution to $f_i^{(1,0)}(\lambda_i(p), p) = 0$, where f_i is given by (10). For simplification, in this proof only, we drop the subscript of f_i and Λ_i, and denote $\lambda_i(p)$ by λ. It is easy to show the denominator of $R_i''(p)$ is negative, and we only need to look at the sign of

$$A \equiv f^{(2,0)}(\lambda, p) * f^{(0,2)}(\lambda, p) - f^{(1,1)}(\lambda, p)^2.$$

We first give expressions for $f^{(2,0)}(\lambda, p)$, $f^{(0,2)}(\lambda, p)$ and $f^{(1,1)}(\lambda, p)$:

$$\begin{aligned}
f^{(2,0)}(\lambda, p) &= \frac{2\lambda}{p^3}[\Lambda^{(0,1)}(-\lambda, \frac{\lambda}{p}) - \Lambda^{(0,1)}(0, \frac{\lambda}{p})] + \frac{\lambda^2}{p^4}[\Lambda^{(0,2)}(-\lambda, \frac{\lambda}{p}) + (p-1)\Lambda^{(0,2)}(0, \frac{\lambda}{p})]; \\
f^{(0,2)}(\lambda, p) &= \Lambda^{(2,0)}(-\lambda, \frac{\lambda}{p}) - \frac{2}{p}\Lambda^{(1,1)}(-\lambda, \frac{\lambda}{p}) + \frac{1}{p^2}\Lambda^{(0,2)}(-\lambda, \frac{\lambda}{p}) + \frac{p-1}{p^2}\Lambda^{(0,2)}(0, \frac{\lambda}{p}); \\
f^{(1,1)}(\lambda, p) &= \frac{1}{p^2}[\Lambda^{(0,1)}(0, \frac{\lambda}{p}) - \Lambda^{(0,1)}(-\lambda, \frac{\lambda}{p}) + \lambda\Lambda^{(1,1)}(-\lambda, \frac{\lambda}{p}) - \frac{\lambda}{p}\Lambda^{(0,2)}(-\lambda, \frac{\lambda}{p}) \\
&\quad -(1 - \frac{1}{p})\lambda\Lambda^{(0,2)}(0, \frac{\lambda}{p})].
\end{aligned}$$

Simple algebra yields that

$$\begin{aligned}
p^4 A &= \lambda\Lambda^{(2,0)}(-\lambda, \frac{\lambda}{p})\big[\lambda\Lambda^{(0,2)}(-\lambda, \frac{\lambda}{p}) + (p-1)\lambda\Lambda^{(0,2)}(0, \frac{\lambda}{p})\big] \\
&\quad +2p\lambda\Lambda^{(2,0)}(-\lambda, \frac{\lambda}{p})\big[\Lambda^{(0,1)}(-\lambda, \frac{\lambda}{p}) - \Lambda^{(0,1)}(0, \frac{\lambda}{p})\big] \\
&\quad -\big[\Lambda^{(0,1)}(0, \frac{\lambda}{p}) - \Lambda^{(0,1)}(-\lambda, \frac{\lambda}{p}) - \lambda\Lambda^{(1,1)}(-\lambda, \frac{\lambda}{p})\big]^2 \\
&\leq p\lambda\Lambda^{(2,0)}(-\lambda, \frac{\lambda}{p})\big[2\Lambda^{(0,1)}(-\lambda, \frac{\lambda}{p}) - 2\Lambda^{(0,1)}(0, \frac{\lambda}{p}) + \lambda\Lambda^{(0,2)}(0, \frac{\lambda}{p})\big],
\end{aligned}$$

where the inequality uses condition $\Lambda_i^{(0,2)}(-\mu,\lambda) \le \Lambda_i^{(0,2)}(0,\lambda)$. We notice that $R_i'(p) = 0$ implies

$$0 = f^{(0,1)}(\lambda,p) = \frac{\lambda}{p^2}\Lambda^{(0,1)}(-\lambda,\frac{\lambda}{p}) - \Lambda(0,\frac{\lambda}{p}) + \frac{p-1}{p^2}\lambda\Lambda^{(0,1)}(0,\frac{\lambda}{p}).$$

Plugging this into the right hand side of the above inequality, we have

$$\begin{aligned} p^4 A &\le 2p^3\Lambda^{(2,0)}(-\lambda,\frac{\lambda}{p})\big[\Lambda(0,\frac{\lambda}{p}) - \frac{\lambda}{p}\Lambda^{(0,1)}(0,\frac{\lambda}{p}) + \frac{\lambda^2}{2p^2}\Lambda^{(0,2)}(0,\frac{\lambda}{p})\big] \\ &= \frac{\lambda^3}{3}\Lambda^{(2,0)}(-\lambda,\frac{\lambda}{p})\Lambda^{(0,3)}(0,\theta\frac{\lambda}{p}) \le 0, \end{aligned}$$

where the second equality uses mean value theorem for some $\theta \in [0,1]$, and the inequality uses condition $\Lambda_i^{(0,3)}(0,\lambda) \le 0$.

Now we verify the two conditions for the non-Gaussian distribution. If we denote as $U(\lambda)$ the log-moment generating function of uniform distribution $U(-1,1)$, the conditions are reduced to $U''(-\mu + c\lambda) \le U''(c\lambda)$ and $c^3U'''(c\lambda) + b^3U'''(b\lambda) \le 0$ for $\mu \ge \lambda > 0$. Since distribution $U(-1,1)$ is symmetric, we know $U''(\lambda)$ is an even function, and $U'''(\lambda)$ is an odd function. Once we can show $U'''(\lambda) \le 0$ for $\lambda > 0$, we have

$$c^3U'''(c\lambda) + b^3U'''(b\lambda) = |c|^3U'''(|c|\lambda) + |b|^3U'''(|b|\lambda) \le 0,$$

and

$$U''(-\mu + c\lambda) = U''(\mu - c\lambda) \le U''(\lambda - c\lambda) \le U''(c\lambda).$$

To show $U'''(\lambda) > 0$ for $\lambda > 0$, we define

$$h(\lambda) \equiv \frac{(e^\lambda - 1)^3}{e^\lambda(e^\lambda + 1)} - \frac{\lambda^3}{2} = e^\lambda - 4 - e^{-\lambda} + \frac{8}{e^\lambda + 1} - \frac{\lambda^3}{2}.$$

Now we calculate the fourth derivative of h for $\lambda > 0$:

$$\begin{aligned} h''''(\lambda) &= e^\lambda - e^{-\lambda} + 8\frac{e^{4\lambda} - 11e^{3\lambda} + 11e^{2\lambda} - e^\lambda}{(e^\lambda + 1)^5} \\ &= \frac{1 - e^{-\lambda}}{(e^\lambda + 1)^5}\big[(e^\lambda + 1)^6 + 8e^{2\lambda}(e^{2\lambda} - 10e^\lambda + 1)\big] \\ &> \frac{1 - e^{-\lambda}}{(e^\lambda + 1)^5}\big[2^6e^{3\lambda} + 8e^{2\lambda}(-8e^\lambda)\big] = 0, \end{aligned}$$

where the inequality uses $e^{2\lambda} + 1 > 2e^\lambda$ and $e^\lambda + 1 > 2e^{\lambda/2}$. Since $h(0) = h'(0) = h''(0) = h'''(0) = 0$, we can show $h(\lambda) > 0$ by taking a fourth-order Taylor expansion. Straightforward algebra yields

$$U'''(\frac{\lambda}{2}) = \frac{8e^\lambda(e^\lambda + 1)}{(e^\lambda - 1)^3} - \frac{16}{\lambda^3}.$$

It is trivial to see that $h(\lambda) > 0$ implies $U'''(\frac{\lambda}{2}) < 0$.

For the case where $c > 0.5$, we actually can show R_i is quasi-convex on $[2c, \infty)$. However, it is hard to analytically characterize the shape of R_i on $[1, 2c]$. □

More detailed analysis of $R_i(\cdot)$ in Example 3 of Section 3.2.

As in the previous proof, denote $U(\lambda)$ as the log-moment generating function of distribution $U(-1,1)$. In addition, define function $\widetilde{U}(\lambda) = \lambda U'(\lambda) - U(\lambda)$. Using Lemma 4, we have $R_i(\infty) = a\gamma_i - U(\gamma_i)$, where γ_i is, as introduced in Assumption 1, the root to equation $U'(\gamma_i) = a$. To make the notation more meaningful, we replace γ_i with λ_∞. Now we let λ_1 be the root to $U'(\lambda_1) = \frac{a}{b}$ and express $R_i(\infty)$, $R_i(1)$ and $R_i'(1)$ as follows:

$$R_i(\infty) = \widetilde{U}(\lambda_\infty), \quad R_i(1) = \widetilde{U}(\lambda_1), \text{ and } R_i'(1) = \widetilde{U}(\lambda_1) - U(\frac{\lambda_1}{b}).$$

We note that the convexity of U implies both U and $\widetilde{U}$ are increasing when $\lambda > 0$. It follows

$$b > 1 \Longrightarrow U'(\lambda_1) = \frac{a}{b} < a = U'(\lambda_\infty) \Longrightarrow \lambda_1 < \lambda_\infty \Longrightarrow R_i(1) = \widetilde{U}(\lambda_1) < \widetilde{U}(\lambda_\infty) = R_i(\infty).$$

We can also show $b < 1 \Longrightarrow R_i(1) > R_i(\infty)$. Now we can see at least three possible cases that are illustrated in Figure 1. □

References

Chaganty, N.R., J. Sethuraman. 1993. Strong large deviation and local limit theorems. *Annals of Probability* **21(3)** 1671-1690.

Chen, C.H., E. Yücesan, L. Dai, H.C. Chen. 2005. Efficient computation of optimal budget allocation for discrete event simulation experiment. *IIE Transactions*, forthcoming.

Chen, C.H., J. Lin, E. Yücesan, S.E. Chick. 2000. Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Discrete Event Dynamic Systems: Theory and Applications* **10** 251-270.

Chen, H.C., C.H. Chen, L. Dai, E. Yücesan. 1997. New development of optimal computing budget allocation for discrete event simulation. *Proceedings of the Winter Simulation Conference.* 334-341.

Chick, S.E., K. Inoue. 2001a. New two-stage and sequential procedures for selecting the best simulated system. *Operations Research* **49** 732-743.

Chick, S.E., K. Inoue. 2001b. New procedures to select the best simulated system using common random numbers. *Management Science* **47** 1133-1149.

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

Dai, L., C. Chen. 1997. Rate of convergence for ordinal comparison of dependent simulations in discrete event dynamic systems. *Journal of Optimization Theory and Applications* **94** 29-54.

Deng, M., Y.C. Ho, J.Q. Hu. 1992. Effect of correlated estimation errors in ordinal optimization. *Proceedings of the Winter Simulation Conference*. 231-240.

Fu, M.C., Hu, J.Q., Chen, C.H., Xiong, X. 2006. Simulation allocation for determining the best design in the presence of correlated sampling. *INFORMS Journal on Computing*, forthcoming.

Fu, M.C., Hu, J.Q., Chen, C.H., Xiong, X. 2004. Optimal computing budget allocation under correlated sampling. *Proceedings of the Winter Simulation Conference*, 595-603.

Glynn, P., Juneja, S. 2004. A large deviations perspective on ordinal optimization. *Proceedings of the Winter Simulation Conference*. 577-585.

Goldsman, D., B.L. Nelson. 1998. Comparing systems via simulation, Chapter 8 in J. Banks, ed. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. John Wiley & Sons, New York. 273-306.

Y. C. Ho, R. Sreenivas and P. Vakili. 1992. Ordinal Optimization of DEDS. Discrete Event Dynamic Systems:Theory and Applications, 2:61-88.

Jin, X., Fu, M.C., Xiong, X. 2003. Probabilistic Error Bounds for Simulation Quantile Estimators. *Management Science* **49** 230-246.

Nelson, B.L., F.J. Matejcik. 1995. Using common random numbers for indifference-zone selection and multiple comparisons in simulation. *Management Science* **41** 1935-1945.

Xiong, X. 2005. Stochastic Optimization: Algorithms and Convergence Results. PhD dissertation, University of Maryland.

Yang, W.N., B.L. Nelson. 1991. Using common random numbers and control variates in multiple-comparison procedures. *Operations Research* **39** 583-591.