Hindawi Publishing Corporation
Discrete Dynamics in Nature and Society
Volume 2014, Article ID 195054, 9 pages
http://dx.doi.org/10.1155/2014/195054

Research Article

Efficient Simulation Budget Allocation for Ranking the Top m Designs

Hui Xiao[1] and Loo Hay Lee[2]

[1] *School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China*
[2] *Department of Industrial and Systems Engineering, National University of Singapore, Singapore 117576*

Correspondence should be addressed to Hui Xiao; msxh@swufe.edu.cn

Received 28 February 2014; Revised 22 May 2014; Accepted 4 June 2014; Published 18 June 2014

Academic Editor: Rigoberto Medina

We consider the problem of ranking the top m designs out of k alternatives. Using the optimal computing budget allocation framework, we formulate this problem as that of maximizing the probability of correctly ranking the top m designs subject to the constraint of a fixed limited simulation budget. We derive the convergence rate of the false ranking probability based on the large deviation theory. The asymptotically optimal allocation rule is obtained by maximizing this convergence rate function. To implement the simulation budget allocation rule, we suggest a heuristic sequential algorithm. Numerical experiments are conducted to compare the effectiveness of the proposed simulation budget allocation rule. The numerical results indicate that the proposed asymptotically optimal allocation rule performs the best comparing with other allocation rules.

1. Introduction

Discrete event system (DES) simulation has been widely used for analyzing and evaluating complex systems since the assumptions for deriving an analytical solution are rarely satisfied in real situation. While DES simulation has been successful in solving many practical problems in a variety of areas such as supply chain systems, healthcare systems, and manufacturing systems, the concerns on the efficiency have never ended [1]. To obtain a statistically significant value, a large number of simulation replications are needed for each design. The performance of each design is then estimated by its sample mean. The ultimate accuracy of this estimator cannot be improved faster than $(1/\sqrt{N})$, where N is the number of simulation replications.

Ordinal optimization (OO) which aims to obtain a good estimate through ordinal comparisons although the estimated value is still poor emerges as a way to improve the simulation efficiency [2]. If the goal of the simulation experiment is to identify the good designs instead of finding the accurate estimate of the true performance value, which is true in many real applications, OO can reduce the number of simulation replications significantly [3, 4]. Intuitively, a larger portion of the total simulation replications should be allocated to those designs that are critical in identifying the best design in order to achieve a high probability of correct selection. Based on this idea, an optimal computing budget allocation (OCBA) framework has been proposed to enhance the simulation efficiency further [5, 6]. OCBA focuses on the efficiency of simulation by intelligently allocating further replications based on both the mean and the variance. In parallel with OCBA, two other well-known ranking and selection procedures frequently used in simulation are the indifference zone (IZ) procedure and value information procedure (VIP). The IZ procedure focuses on finding a feasible way to guarantee that the prespecified probability of correct selection can be achieved [7]. The VIP uses the Bayesian posterior distribution to describe the evidence of correct selection and allocates further replications based on maximizing the value information [8, 9]. The three popular procedures of finding the best design are compared in [10]. Extensions of the research in OCBA and ranking and selection include subset selection [11–13], selecting the Pareto set for multiple objective functions [14], selecting the best design subject to stochastic constraints [15, 16], and complete ranking [17]. A detailed summary of the existing work in OCBA can be found in [18].

To the best of our knowledge, no previous research has considered the simulation budget allocation for ranking the top m designs out of k alternatives. The developed top m ranking procedure is most useful when the designs for comparison have multiple dimensions of performance measurements with some qualitative criteria such as environmental consideration and political feasibility. Top m ranking procedure can help the decision makers to identify the ranking of the top m designs based on the quantitative performance measurement. Final decision can be made by incorporating other qualitative performance measurements. Hence, the top m ranking gives decision makers a more flexible and people oriented way to support the decision making process. In addition, the top m ranking procedure can also be integrated into some evolutionary search algorithms, where the ranking information of the top solutions is needed in order to determine the search direction of next iteration. For example, the ranking of the top candidate solutions may be needed in each iteration of the genetic algorithm since better candidate solutions are given higher probabilities to reproduce in genetic algorithms.

In this paper, we consider the problem of ranking the top m designs out of total k alternatives, where the performance value of each design can only be estimated with noise via simulation. The objective of this paper is to determine how to allocate the simulation replications among the k designs such that the probability of correctly ranking the top m designs can be maximized. The organization of this paper is as follows. Section 2 provides the mathematical formulation of the top m ranking problem using OCBA framework. In Section 3, we derive the convergence rate function of the false ranking probability. Section 4 derives the asymptotically optimal allocation rule using large deviation theory. A sequential allocation algorithm is proposed in Section 5. Section 6 conducts numerical experiments to demonstrate the effectiveness of the proposed simulation budget allocation rule. Finally, we conclude our paper in Section 7.

2. Problem Formulation

Consider the problem of ranking the top m designs out of k alternatives. The performance of each design can only be estimated with noise via simulation. The mean performance is used as the ranking criterion. In order to have a steady mean performance value, a large number of simulation replications are needed because of the randomness of individual samples. Given that a total of n simulation replications are available, the objective is to find the best allocation of the total n replications to the k designs in order to maximize the probability of correctly ranking the top m designs.

Denote the mean performances of each design by $\mu_1, \ldots, \mu_k$ such that $\mu_1 < \cdots < \mu_i < \cdots < \mu_k$ and $\mu_{i+1} - \mu_i \geq \delta$, $i = 1, \ldots, k-1$, where δ is a positive number. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$ denote the proportion of the total computing budget n to be allocated to each design such that $\sum_{i=1}^{k} \alpha_i = 1$, $i = 1, \ldots, k$. Let $\overline{X}_i(\alpha_i n) = (\alpha_i n)^{-1} \sum_{j=1}^{\alpha_i n} X_{ij}$ denote the sample mean of design i, where $(X_{i1}, \ldots, X_{i,\alpha_i n})$ are the samples from population i. We ignore the case when $\alpha_i n$ is not an integer because we could let $\alpha_i n$ be $\lfloor \alpha_i n \rfloor$, the greatest integer less than $\alpha_i n$. The analysis remains unaffected. The objective of this paper is to find the optimal allocation strategy $\boldsymbol{\alpha}^* = (\alpha_1^*, \ldots, \alpha_k^*)$ such that the probability of correctly ranking the top m designs can be maximized with a fixed limited computing budget n.

Given that $\mu_1 < \cdots < \mu_i < \cdots < \mu_k$, the ranking of the top m designs is correctly identified if $\overline{X}_i(\alpha_i n) \leq \overline{X}_{i+1}(\alpha_{i+1} n)$ for all $i = 1, \ldots, m-1$ and if $\overline{X}_m(\alpha_m n) \leq \overline{X}_j(\alpha_j n)$ for all $j = m+1, \ldots, k$. Mathematically, we can write the probability of correctly ranking the top m designs as follows:

$$P(\mathrm{CR}_m) = P\left(\left\{\bigcap_{i=1}^{m-1}\left(\overline{X}_i(\alpha_i n) \leq \overline{X}_{i+1}(\alpha_{i+1} n)\right)\right\} \bigcap \left\{\bigcap_{j=m+1}^{k}\left(\overline{X}_m(\alpha_m n) \leq \overline{X}_j(\alpha_j n)\right)\right\}\right). \quad (1)$$

Hence, an optimal computing budget allocation problem can be formulated by maximizing the probability of correctly ranking the top m designs:

$$\max_{\alpha_1, \ldots, \alpha_k} \quad P(\mathrm{CR}_m)$$
$$\text{s.t.} \quad \alpha_1 + \cdots + \alpha_i + \cdots + \alpha_k = 1, \quad \alpha_i \geq 0, \ i = 1, \ldots, k. \quad (2)$$

Maximizing the probability of correctly ranking the top m designs is equivalent to minimizing the false ranking probability of the top m designs. Using asymptotical analysis, this is also equivalent to maximizing the convergence rate at which the false ranking probability goes to zero. In this paper, we use large deviation theory to derive this convergence rate function and reformulate the optimization model (2) by maximizing the convergence rate function. The assumptions needed in this paper are stated as follows.

Assumption 1. The performance of every design is independently simulated.

The independence of each design ensures that the samples $(X_{i1}, \ldots, X_{i,\alpha_i n})$ for each $i = 1, \ldots, k$ generated are independent. Thus, the results we obtained will not be affected by the correlations among different designs.

Define the cumulant generating function of sample mean $\overline{X}_i(n)$ to be $\Lambda_i^{(n)}(\theta) = \ln E(e^{\theta \overline{X}_i(n)})$. The effective domain of any function $f : D_f \to R^*$ is the set $\{x \in D_f : f(x) < \infty\}$, while the range is $R^* = R \cup \{+\infty\}$. Let $D_{\Lambda_i} = \{\theta \in R : \Lambda_i(\theta) < \infty\}$ and $F_i = \{\Lambda_i'(\theta) : \theta \in D_{\Lambda_i}^0\}$. For any set A, A^o denotes its interior and $\overline{A}$ denotes its closure.

Assumption 2. For each $i = 1, \ldots, k$,

(1) the limit $\Lambda_i(\theta) = \lim_{n \to \infty} (1/n) \Lambda_i^{(n)}(n\theta)$ is well defined as an extended real number for all θ;

(2) the origin belongs to $D^0_{\Lambda_i}$;

(3) $\Lambda_i(\cdot)$ is strictly convex and steep, that is, $\lim_{n\to\infty}|\Lambda'_i(\theta_n)| = \infty$, where $\{\theta_n\}$ is a sequence converging to the boundary point of D_{Λ_i};

(4) $[\mu_1, \mu_k] \subset \bigcap_{i=1}^k F_i^0$.

Assumption 2 implies that $\mu_i = \Lambda'_i(0)$ with $\overline{X}_i(n) \to \mu_i$ almost surely when $n \to \infty$. Furthermore, it indicates that the sample mean $\overline{X}_i(\alpha_i n)$ satisfies the large deviation principle. The last condition of Assumption 2 ensures that the sample mean of every design can take any value between μ_1 and μ_k and $P(\overline{X}_i(\alpha_i n) \geq \overline{X}_{i+1}(\alpha_{i+1}n)) > 0$ [19].

3. Rate Function of the False Ranking Probability

We now derive the probability of false ranking other than the correct ranking probability, followed by the corresponding derivation of its large deviation principle. Recall that the probability of correctly ranking the top m designs is defined in (1). The probability of falsely ranking the top m designs is simply its complement; that is,

$$\begin{aligned}
P(\mathrm{FR}_m) &= 1 - P(\mathrm{CR}_m) \\
&= P\left\{\left(\left\{\bigcap_{i=1}^{m-1}\left(\overline{X}_i(\alpha_i n) \leq \overline{X}_{i+1}(\alpha_{i+1}n)\right)\right\}\right.\right. \\
&\qquad \left.\left.\cap\left\{\bigcap_{j=m+1}^{k}\left(\overline{X}_m(\alpha_m n) \leq \overline{X}_j(\alpha_j n)\right)\right\}\right)^C\right\} \\
&= P\left\{\left(\bigcup_{i=1}^{m-1}\left(\overline{X}_i(\alpha_i n) \geq \overline{X}_{i+1}(\alpha_{i+1}n)\right)\right)\right. \\
&\qquad \left.\bigcup\left(\bigcup_{j=m+1}^{k}\left(\overline{X}_m(\alpha_m n) \geq \overline{X}_j(\alpha_j n)\right)\right)\right\},
\end{aligned} \tag{3}$$

where $(\cdot)^C$ represents the complement of $(\cdot)$.

$P(\mathrm{FR}_m)$ has a lower bound,

$$\begin{aligned}
\mathrm{lb} = \max\left\{\max_{1\leq i\leq m-1} P\left(\overline{X}_i(\alpha_i n) \geq \overline{X}_{i+1}(\alpha_{i+1}n)\right),\right. \\
\left.\max_{j=m+1,\ldots,k} P\left(\overline{X}_m(\alpha_m n) \geq \overline{X}_j(\alpha_j n)\right)\right\},
\end{aligned} \tag{4}$$

and an upper bound $\mathrm{ub} = (k-1)\times \mathrm{lb}$ such that, assuming the limit exists,

$$\lim_{n\to\infty}\frac{1}{n}\ln P(\mathrm{FR}_m) = \lim_{n\to\infty}\frac{1}{n}\ln \mathrm{lb}. \tag{5}$$

Theorem 3 formally states that the limit exists and the overall convergence rate function is the minimum rate function of each probability. The convergence rate can be understood as the speed at which false ranking probability goes to zero.

Theorem 3. *The rate function of $P(FR_m)$ is given by*

$$\begin{aligned}
&-\lim_{n\to\infty}\frac{1}{n}\ln P(FR_m) \\
&\quad = \min\left\{\min_{1\leq i\leq m-1}\inf_x\left(\alpha_i I_i(x) + \alpha_{i+1}I_{i+1}(x)\right),\right. \\
&\qquad\qquad \left.\min_{j=m+1,\ldots,k}\inf_x\left(\alpha_j I_j(x) + \alpha_m I_m(x)\right)\right\},
\end{aligned} \tag{6}$$

where $I(x) = \sup_{\theta\in R}(\theta x - \Lambda(\theta))$ is the Fenchel-Legendre transform and $\Lambda(\theta) = \ln E(e^{\theta X})$.

Proof. If there exists a function $G_i(\cdot)$ such that

$$\lim_{n\to\infty}\frac{1}{n}\ln P\left\{\overline{X}_i(\alpha_i n) \geq \overline{X}_{i+1}(\alpha_{i+1}n)\right\} = -G_i(\alpha_i, \alpha_{i+1}), \tag{7}$$

for $i = 1, \ldots, m-1$, and

$$\lim_{n\to\infty}\frac{1}{n}\ln P\left\{\overline{X}_m(\alpha_m n) \geq \overline{X}_j(\alpha_j n)\right\} = -G_j(\alpha_j, \alpha_m), \tag{8}$$

for $j = m+1, \ldots, k$.

Then, it can be concluded that

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\ln P(\mathrm{FR}_m) \\
&\quad = -\min\left\{\min_{1\leq i\leq m-1} G_i(\alpha_i, \alpha_{i+1}), \min_{j=m+1,\ldots,k} G_j(\alpha_j, \alpha_m)\right\}.
\end{aligned} \tag{9}$$

Now we are in the position to derive the assumed function $G_i(\cdot)$. Define $Y_n = (\overline{X}_i(\alpha_i n), \overline{X}_{i+1}(\alpha_{i+1}n))$. The cumulant moment generating function of Y_n can be written as

$$\begin{aligned}
\Lambda_n(\theta_i, \theta_{i+1}) &= \ln E\left(e^{\theta_i \overline{X}_i(\alpha_i n) + \theta_{i+1}\overline{X}_{i+1}(\alpha_{i+1}n)}\right) \\
&= \Lambda_i^{(\alpha_i n)}\left(\frac{\theta_i}{\alpha_i n}\right) + \Lambda_{i+1}^{(\alpha_{i+1}n)}\left(\frac{\theta_{i+1}}{\alpha_{i+1}n}\right).
\end{aligned} \tag{10}$$

Under Assumption 2,

$$\lim_{n\to\infty}\frac{1}{n}\Lambda_n(n\theta_i, n\theta_{i+1}) = \alpha_i\Lambda_i\left(\frac{\theta_i}{\alpha_i}\right) + \alpha_{i+1}\Lambda_{i+1}\left(\frac{\theta_{i+1}}{\alpha_{i+1}}\right). \tag{11}$$

By the Gärtner-Ellis Theorem [20], $\{Y_n, n = 1, 2, \ldots\}$ satisfies large deviation principle with good rate function which can be expressed as follows:

$$\begin{aligned} I(x_i, x_{i+1}) &= \sup_{\theta_i, \theta_{i+1}} \left(\theta_i x_i + \theta_{i+1} x_{i+1} - \alpha_i \Lambda_i \left(\frac{\theta_i}{\alpha_i} \right) \right. \\ &\qquad \left. - \alpha_{i+1} \Lambda_{i+1} \left(\frac{\theta_{i+1}}{\alpha_{i+1}} \right) \right) \\ &= \sup_{\theta_i} \left(\theta_i x_i - \alpha_i \Lambda_i \left(\frac{\theta_i}{\alpha_i} \right) \right) \\ &\quad + \sup_{\theta_{i+1}} \left(\theta_{i+1} x_{i+1} - \alpha_{i+1} \Lambda_{i+1} \left(\frac{\theta_{i+1}}{\alpha_{i+1}} \right) \right) \\ &= \alpha_i \sup_{\theta_i/\alpha_i} \left(\left(\frac{\theta_i}{\alpha_i} \right) x_i - \Lambda_i \left(\frac{\theta_i}{\alpha_i} \right) \right) \\ &\quad + \alpha_{i+1} \sup_{\theta_{i+1}/\alpha_{i+1}} \left(\left(\frac{\theta_{i+1}}{\alpha_{i+1}} \right) x_{i+1} - \Lambda_{i+1} \left(\frac{\theta_{i+1}}{\alpha_{i+1}} \right) \right) \\ &= \alpha_i I_i(x_i) + \alpha_{i+1} I_{i+1}(x_{i+1}). \end{aligned} \tag{12}$$

Hence, from large deviation principle,

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \ln P\left(\overline{X}_i(\alpha_i n) \geq \overline{X}_{i+1}(\alpha_{i+1} n)\right) \\ &\quad = -\inf_x \left(\alpha_i I_i(x) + \alpha_{i+1} I_{i+1}(x)\right). \end{aligned} \tag{13}$$

Similarly,

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \ln P\left(\overline{X}_m(\alpha_m n) \geq \overline{X}_j(\alpha_j n)\right) \\ &\quad = -\inf_x \left(\alpha_j I_j(x) + \alpha_m I_m(x)\right). \end{aligned} \tag{14}$$

Therefore, the convergence rate function of the false ranking probability can be expressed as follows:

$$\begin{aligned} &-\lim_{n\to\infty} \frac{1}{n} \ln P(\mathrm{FR}_m) \\ &\quad = \min\left\{ \min_{1\leq i\leq m-1} G_i(\alpha_i, \alpha_{i+1}), \min_{j=m+1,\ldots,k} G_j(\alpha_j, \alpha_m) \right\} \\ &\quad = \min\left\{ \min_{1\leq i\leq m-1} \inf_x \left(\alpha_i I_i(x) + \alpha_{i+1} I_{i+1}(x)\right), \right. \\ &\qquad\qquad \left. \min_{j=m+1,\ldots,k} \inf_x \left(\alpha_j I_j(x) + \alpha_m I_m(x)\right) \right\}. \end{aligned} \tag{15}$$

□

4. Asymptotically Optimal Allocation

The objective is to maximize the probability of correctly ranking the top m designs. This can be achieved by minimizing the false ranking probability. It is also equivalent to maximizing the convergence rate of $P(\mathrm{FR}_m)$ subject to $\sum_{i=1}^{k} \alpha_i = 1$ and $\alpha_i \geq 0$, for all $i = 1, \ldots, k$. The optimization model in (1) can be reexpressed as

$$\begin{aligned} \max \quad & \min\left\{ \min_{1\leq i\leq m-1} \inf_x \left(\alpha_i I_i(x) + \alpha_{i+1} I_{i+1}(x)\right), \right. \\ & \qquad \left. \min_{j=m+1,\ldots,k} \inf_x \left(\alpha_j I_j(x) + \alpha_m I_m(x)\right) \right\} \\ \text{s.t.} \quad & \sum_{i=1}^{k} \alpha_i = 1, \quad \alpha_i \geq 0, \; i \in \{1, \ldots, k\}. \end{aligned} \tag{16}$$

By [19], $\alpha_i I_i(x) + \alpha_{i+1} I_{i+1}(x)$ is a strictly increasing concave function. The infimum of concave functions is also concave. Likewise, the minimum of concave functions is a concave function too. Define $x(\alpha_i, \alpha_{i+1}) = \arg\inf_x(\alpha_i I_i(x) + \alpha_{i+1} I_{i+1}(x))$. As shown in [19], $x(\alpha_i, \alpha_{i+1})$ is the solution to $\alpha_i I_i'(x) + \alpha_{i+1} I_{i+1}'(x) = 0$ and

$$\begin{aligned} &\frac{\partial\left(\alpha_i I_i\left(x(\alpha_i, \alpha_{i+1})\right) + \alpha_{i+1} I_{i+1}\left(x(\alpha_i, \alpha_{i+1})\right)\right)}{\partial \alpha_i} \\ &\quad = I_i\left(x(\alpha_i, \alpha_{i+1})\right), \\ &\frac{\partial\left(\alpha_i I_i\left(x(\alpha_i, \alpha_{i+1})\right) + \alpha_{i+1} I_{i+1}\left(x(\alpha_i, \alpha_{i+1})\right)\right)}{\partial \alpha_{i+1}} \\ &\quad = I_{i+1}\left(x(\alpha_i, \alpha_{i+1})\right). \end{aligned} \tag{17}$$

The result can similarly be applied to $\alpha_j I_j(x) + \alpha_m I_m(x)$. Therefore, the optimization model (16) is a concave maximization problem and it can be reexpressed as follows:

$$\begin{aligned} \max \quad & z \\ \text{s.t.} \quad & \alpha_i I_i\left(x(\alpha_i, \alpha_{i+1})\right) + \alpha_{i+1} I_{i+1}\left(x(\alpha_i, \alpha_{i+1})\right) \geq z, \\ & \qquad\qquad i \in \{1, \ldots, m-1\} \\ & \alpha_j I_j\left(x(\alpha_j, \alpha_m)\right) + \alpha_m I_m\left(x(\alpha_j, \alpha_m)\right) \geq z, \\ & \qquad\qquad j \in \{m+1, \ldots, k\} \\ & \sum_{i=1}^{k} \alpha_i = 1, \quad \alpha_i \geq 0, \; i \in \{1, \ldots, k\}. \end{aligned} \tag{18}$$

Since model (18) is strictly concave and the functions of $\boldsymbol{\alpha}$ are continuous, a unique optimal solution must exist and the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for global optimality.

From the KKT conditions on problem (18), we define a new problem (19) by replacing some inequality signs and forcing α_i to be strictly positive:

$$\begin{aligned}
\max \quad & z \\
\text{s.t.} \quad & \alpha_1 I_1\left(x\left(\alpha_1,\alpha_2\right)\right)+\alpha_2 I_2\left(x\left(\alpha_1,\alpha_2\right)\right)=z \\
& \min\left\{\alpha_i I_i\left(x\left(\alpha_i,\alpha_{i+1}\right)\right)+\alpha_{i+1} I_{i+1}\left(x\left(\alpha_i,\alpha_{i+1}\right)\right),\right. \\
& \qquad \left.\alpha_{i-1} I_{i-1}\left(x\left(\alpha_{i-1},\alpha_i\right)\right)+\alpha_i I_i\left(x\left(\alpha_{i-1},\alpha_i\right)\right)\right\}=z, \\
& \qquad i\in\{2,\ldots,m-1\} \\
& \alpha_j I_j\left(x\left(\alpha_j,\alpha_m\right)\right)+\alpha_m I_m\left(x\left(\alpha_j,\alpha_m\right)\right)=z, \\
& \qquad j\in\{m+1,\ldots,k\} \\
& \sum_{i=1}^{k}\alpha_i=1, \quad \alpha_i>0,\ i\in\{1,\ldots,k\}.
\end{aligned} \tag{19}$$

Theorem 4. *Under Assumptions 1 and 2 in Section 2, problems* (18) *and* (19) *are equivalent; that is, a solution $\boldsymbol{\alpha}^*=(\alpha_1^*,\ldots,\alpha_k^*)$ is the optimal solution to* (18) *if and only if it is also an optimal solution to* (19).

Proof. We assume that a point satisfying the KKT condition of (18) is also feasible to (19). We first prove the forward and backward assertions. We then prove that the assumption that a point satisfying the KKT condition of (18) is also feasible to (19) is indeed correct.

Suppose $\boldsymbol{\alpha}^*$ is the optimal solution to (18). Since the feasible region of (19) is a subset of that of (18), if the optimal solution to (18) is feasible to (19), it must be optimal to (19). Since the KKT conditions are necessary and sufficient for optimality in (18), $\boldsymbol{\alpha}^*$ must satisfy the KKT conditions of (18). Hence, $\boldsymbol{\alpha}^*$ is feasible to (19). Therefore, if a point satisfies the KKT condition in (18), it must be optimal to (19).

Suppose the optimal solution to (18) is $\boldsymbol{\alpha}^*$ and the optimal solution to (19) is $\tilde{\boldsymbol{\alpha}}^*$, and $\boldsymbol{\alpha}^*\neq\tilde{\boldsymbol{\alpha}}^*$. Since the KKT conditions are necessary and sufficient condition to (18), thus, $\boldsymbol{\alpha}^*$ must satisfy the KKT conditions. Furthermore, the objective function of (19) is the same as that of (18), and the feasible region of (19) is a subset of that of (18). Therefore, $\boldsymbol{\alpha}^*$ must be infeasible to (19). However, we assumed that a point satisfying the KKT conditions of (18) must be feasible to (19). We have thus reached a contradiction. So we must have $\boldsymbol{\alpha}^*=\tilde{\boldsymbol{\alpha}}^*$.

We are now in the position to prove that a point satisfying the KKT conditions of (18) must be feasible to (19). If we let $\alpha_i=1/k$, we can have $z>0$ for problem (18). However, any $\alpha_i=0$ for $i=1,\ldots,m-1$ will lead to $\alpha_{i+1}\inf_x I_{i+1}(x)=\alpha_{i+1}I_{i+1}(\mu_{i+1})=0$. If $\alpha_j=0$ for $j=m+1,\ldots,k$, we will have $\alpha_m\inf_x I_m(x)=\alpha_m I_m(\mu_m)=0$. Therefore, the optimal solution must satisfy $\alpha_i>0$ for every $i=1,\ldots,k$.

Since the problem (18) is a concave optimization problem, the first order condition is also the optimality condition. According to the KKT conditions, there exist $\lambda_i\geq 0$, $\lambda_j\geq 0$, $i\in\{1,\ldots,m-1\}$, $j\in\{m+1,\ldots,k\}$, and $\gamma>0$ such that

$$\sum_{i=1}^{m-1}\lambda_i+\sum_{j=m+1}^{k}\lambda_j=1, \tag{20}$$

$$\lambda_1 I_1\left(x\left(\alpha_i^*,\alpha_{i+1}^*\right)\right)=\gamma, \tag{21}$$

$$\lambda_{i-1}I_i\left(x\left(\alpha_{i-1}^*,\alpha_i^*\right)\right)+\lambda_i I_i\left(x\left(\alpha_i^*,\alpha_{i+1}^*\right)\right)=\gamma, \quad i\in\{2,\ldots,m-1\}, \tag{22}$$

$$\lambda_{m-1}I_m\left(x\left(\alpha_{m-1}^*,\alpha_m^*\right)\right)+\sum_{j=m+1}^{k}\lambda_j I_m\left(x\left(\alpha_j^*,\alpha_m^*\right)\right)=\gamma, \tag{23}$$

$$\lambda_j I_j\left(x\left(\alpha_j^*,\alpha_m^*\right)\right)=\gamma, \quad j\in\{m+1,\ldots,k\}, \tag{24}$$

$$\lambda_i\left(\alpha_i^* I_i\left(x\left(\alpha_i^*,\alpha_{i+1}^*\right)\right)+\alpha_{i+1}^* I_{i+1}\left(x\left(\alpha_i^*,\alpha_{i+1}^*\right)\right)-z\right)=0, \quad i\in\{1,\ldots,m-1\}, \tag{25}$$

$$\lambda_j\left(\alpha_j^* I_j\left(x\left(\alpha_j^*,\alpha_m^*\right)\right)+\alpha_m^* I_m\left(x\left(\alpha_j^*,\alpha_m^*\right)\right)-z\right)=0, \quad j\in\{m+1,\ldots,k\}. \tag{26}$$

If $\lambda_1=0$, γ will be zero. Thus, we could conclude that all $\lambda_i,\lambda_j=0$, $j=m+1,\ldots,k$, by putting $\gamma=0$ into (21) to (24). Substituting $\lambda_1=0$ and $\gamma=0$ into (22) and (23) will result in $\lambda_i=0$, $i=2,\ldots,m-1$. However, this contradicts with (20) which requires at least one $\lambda_i,\lambda_j>0$. Thus, we conclude that $\lambda_1>0$ and $\gamma>0$ because $I_i(x(\alpha_i^*,\alpha_{i+1}^*))$, $i=1,\ldots,k-1$, is strictly positive. Similarly, we could conclude that $\lambda_j>0$, $j=m+1,\ldots,k$, from (24), and $\max\{\lambda_i,\lambda_{i+1}\}>0$, $i=1,\ldots,m-1$, from (22).

Based on the results that $\lambda_1>0$, $\lambda_j>0$, $j\in\{m+1,\ldots,k\}$, $\max\{\lambda_i,\lambda_{i+1}\}>0$, $i=2,\ldots,m-1$, and constraints of (18), we have the following equality from the complementary slackness condition in (25) and (26). For $i\in\{2,\ldots,m-1\}$, $j\in\{m+1,\ldots,k\}$,

$$\begin{aligned}
z &= \alpha_1^* I_1\left(x\left(\alpha_1^*,\alpha_2^*\right)\right)+\alpha_2^* I_2\left(x\left(\alpha_1^*,\alpha_2^*\right)\right) \\
&= \alpha_j^* I_j\left(x\left(\alpha_j^*,\alpha_m^*\right)\right)+\alpha_m^* I_m\left(x\left(\alpha_j^*,\alpha_m^*\right)\right) \\
&= \min\left\{\alpha_i^* I_i\left(x\left(\alpha_i^*,\alpha_{i+1}^*\right)\right)+\alpha_{i+1}^* I_{i+1}\left(x\left(\alpha_i^*,\alpha_{i+1}^*\right)\right),\right. \\
&\qquad \left.\alpha_{i-1}^* I_{i-1}\left(x\left(\alpha_{i-1}^*,\alpha_i^*\right)\right)+\alpha_i^* I_i\left(x\left(\alpha_{i-1}^*,\alpha_i^*\right)\right)\right\}.
\end{aligned} \tag{27}$$

So we have proved the assertion that a point satisfying the KKT conditions of (18) must be feasible to (19). This completes the proof of Theorem 4. □

INPUT: m, k, n, n_0, Δ ($n - kn_0$ is a multiple of Δ and $n_0 \geq 5$)
INITIALIZE: Perform n_0 simulation replications for all designs.
$l \leftarrow 0; N_1^l = \cdots = N_i^l = \cdots = N_k^l$
WHILE $\sum_{i=1}^k N_i^l < n$ DO
(1) Increase the computing budget by Δ
(2) Use (28) to determine the new budget allocation rule, that is, $N_i^{l+1}, i = 1, \ldots, k$.
The population distribution parameters are estimated by the sample statistics.
(3) Simulate additional $\max(0, N_i^{l+1} - N_i^l)$ for each design $i = 1, \ldots, k$.
END OF WHILE

Algorithm 1: Sequential allocation algorithm.

Therefore, we could conclude that the optimal allocation rule to rank the top m designs $\boldsymbol{\alpha}^* = (\alpha_1^*, \ldots, \alpha_k^*)$ solves

$$\begin{aligned} &\sum_{i=1}^k \alpha_i = 1, \quad \alpha_i > 0, \\ &\min \{\alpha_i^* I_i (x(\alpha_i^*, \alpha_{i+1}^*)) + \alpha_{i+1}^* I_{i+1} (x(\alpha_i^*, \alpha_{i+1}^*)), \\ &\quad \alpha_{i-1}^* I_{i-1} (x(\alpha_{i-1}^*, \alpha_i^*)) + \alpha_i^* I_i (x(\alpha_{i-1}^*, \alpha_i^*))\} \\ &= \alpha_1^* I_1 (x(\alpha_1^*, \alpha_2^*)) + \alpha_2^* I_2 (x(\alpha_1^*, \alpha_2^*)) \\ &= \alpha_j^* I_j \left(x\left(\alpha_j^*, \alpha_m^*\right)\right) + \alpha_m^* I_m \left(x\left(\alpha_j^*, \alpha_m^*\right)\right), \\ &\quad i = 2, \ldots, m-1; \quad j = m+1, \ldots, k. \end{aligned} \tag{28}$$

Suppose that the performance of each design follows the normal distribution; that is, $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, k$. Equation (28) can be rewritten as follows:

$$\begin{aligned} &\sum_{i=1}^k \alpha_i = 1, \quad \alpha_i > 0, \\ &\min \left\{ \frac{(\mu_i - \mu_{i+1})^2}{2(\sigma_i^2/\alpha_i^* + \sigma_{i+1}^2/\alpha_{i+1}^*)}, \frac{(\mu_i - \mu_{i-1})^2}{2(\sigma_i^2/\alpha_i^* + \sigma_{i-1}^2/\alpha_{i-1}^*)} \right\} \\ &= \frac{(\mu_1 - \mu_2)^2}{2(\sigma_1^2/\alpha_1^* + \sigma_2^2/\alpha_2^*)} = \frac{(\mu_j - \mu_m)^2}{2\left(\sigma_j^2/\alpha_j^* + \sigma_m^2/\alpha_m^*\right)}, \\ &\quad i = 2, \ldots, m-1; \quad j = m+1, \ldots, k. \end{aligned} \tag{29}$$

5. Sequential Allocation

Based on the results from Theorem 4, we can compute the simulation budget allocation rule using (28) for design performances with any arbitrary distributions or using (29) for normally distributed performances if the parameters of the design performance distribution are given. However, no information on the design performance distribution is known before simulation experiments are conducted. To overcome this dilemma, we suggest a heuristic sequential allocation algorithm in order to implement the allocation rule. The algorithm for sequential allocation is summarized in Algorithm 1.

Define l to be the iteration number and define N_i^l, $i = 1, \ldots, k$ to be the total number of simulation replications that have been allocated to design i up to iteration l. n is the total number of simulation replications available. Δ is the number of incremental simulation replications for each iteration. n_0 is the initial number of simulation runs for each design.

As the simulation continues, design i will be ranked number i for all $i \leq m$. The ranking of the top m designs may change from iteration to iteration although it will converge to the true ranking when the total computating budget goes to infinity. When the ranking of the top m designs changes, the budget allocation in the loop will be applied immediately. Therefore, the actual proportion of budget for every system will converge to the optimal proportion when the number of iterations is sufficiently large.

Furthermore, we need to take note of n_0, the initial number of replications for every design. n_0 cannot be too small because the estimation of the rate function can be poor especially when the variance of the performance is large. On the other hand, if n_0 is too large, some designs will be allocated excessively compared with their optimal allocation numbers. When the total budget is very limited, designs that need more replications may suffer from large n_0 and this would eventually affect the simulation results. Other than the initial number of replications, the incremental budget Δ is also important in the implementation process. Large Δ results in wasting of budget, while small Δ will lead to expensive computation in the loop.

It is also worthy to note that there will be significant variation in the estimate of the performance value. Denote the empirical cumulant generating function and rate function of system i as $\Lambda_i^{(n)}(\theta) = \ln((1/n)\sum_{j=1}^n e^{\theta \overline{X}_{ij}})$ and $I_i^{(n)}(x) = \sup_\theta\{\theta x - \Lambda_i^{(n)}(\theta)\}$, respectively. In Algorithm 1, we are using $G_i^{(n)}(\alpha_i, \alpha_{i+1})$ to estimate $G_i(\alpha_i, \alpha_{i+1})$ for the optimal allocation rule. Thus, the true optimal allocation α_i^* is actually estimated by $\alpha_i^*(n)$. As argued in [19], the empirical estimation of the optimal allocation is consistent; that is, $\alpha_i^*(n) \rightarrow \alpha_i^*$, for all $i = 1, \ldots, k$ almost surely when $n \rightarrow +\infty$.

6. Numerical Experiments

In this section, we test the proposed simulation budget allocation rule for ranking the top m designs by comparing it with different allocation rules: equal allocation which

Table 1: Parameters for the numerical experiments.

Design	Equal spacing		Equal variance		Increasing spacing decreasing variance	
	Mean	Variance	Mean	Variance	Mean	Variance
I	2	400	1	100	1	400
II	4	361	2	100	2	361
III	6	324	4	100	4	324
IV	8	289	7	100	7	289
V	10	256	11	100	11	256
VI	12	225	16	100	16	225
VII	14	196	22	100	22	196
VIII	16	169	29	100	29	169
IX	18	144	37	100	37	144
X	20	121	46	100	46	121
XI	22	100	56	100	56	100
XII	24	81	67	100	67	81
XIII	26	64	79	100	79	64
XIV	28	49	92	100	92	49
XV	30	36	106	100	106	36
XVI	32	25	121	100	121	25
XVII	34	16	137	100	137	16
XVIII	36	9	154	100	154	9
XIX	38	4	172	100	172	4
XX	40	1	191	100	191	1

simulates each design equally and the OCBA-m procedure [12]. Although OCBA-m only considers the selection of the top m designs and do not aim to identify the ranking of the top m designs, we can use it here for benchmarking purpose. In all the experiments below, the performance of each design is assumed to follow the normal distribution. Therefore, the optimal allocation rule can be obtained by solving (29). The assumption of normal distribution is generally held in simulation experiments since the output is obtained from an average performance or batch means, so that the central limit theorem effect holds.

6.1. Computing Budget Allocation Rules

Equal Allocation. The simulation replications are allocated equally to each design; that is, $\alpha_i = 1/k$, $i = 1, \ldots, k$. This is the simplest allocation rule and it can serve as a benchmark for other allocation procedures.

Top m Ranking (OCBA-Rm). The simulation budget allocation is derived using (29) and implemented using the sequential allocation algorithm proposed in Section 5. The allocation rule is solved by using the solver "fminimax" in Matlab.

OCBA-m [12]. Sequential allocation algorithm is used to implement this allocation rule. The simulation budget allocation rule is determined by the following equation:

$$\frac{\alpha_1^{l+1}}{s_1^2/\delta_1^2} = \frac{\alpha_2^{l+1}}{s_2^2/\delta_2^2} = \cdots = \frac{\alpha_k^{l+1}}{s_k^2/\delta_k^2}, \tag{30}$$

where s_i^2 is the sample variance for design i, $\delta_i^2 = (\overline{X}_i - c)^2$, and $c = (s_{m+1}\overline{X}_m + s_m\overline{X}_{m+1})/(s_{m+1} + s_m)$ as defined in [12].

6.2. Numerical Results for Different Allocation Procedures. To compare the performance of the procedures, we carried out numerical experiments for the different allocation procedures discussed above. In comparing the procedures, the effectiveness of the procedures is measured by the probability of correctly ranking the top m designs ($P(\mathrm{CR}_m)$) which is estimated by the fraction of the times that the procedure successfully identifies the correct ranking of the top m designs out of 10,000 independent simulation runs.

Each of the procedures simulates each of the k designs for $n_0 = 20$ replications initially as recommended in [1, 12]. The simulation budget is increased by $\Delta = 40$ for each iteration. Table 1 summarizes the mean and variance for the three experiments that will be conducted in this section. In this experiment, the total number of designs is 20, that is, $k = 20$ and the objective is to identify the ranking of the top 5 designs, that is, $m = 5$.

Figure 1 shows the numerical comparison for the three allocation rules, where (a) is for the equal spacing scenario, (b) is for the equal variance scenario, and (c) is for the increasing spacing but decreasing variance scenario.

The experiment results show that the proposed simulation budget allocation rule OCBA-Rm performs the best in all three experiments. It is also interesting to note that OCBA-m, which performs significantly better than EA when the objective is selecting the top m designs, fares much worse

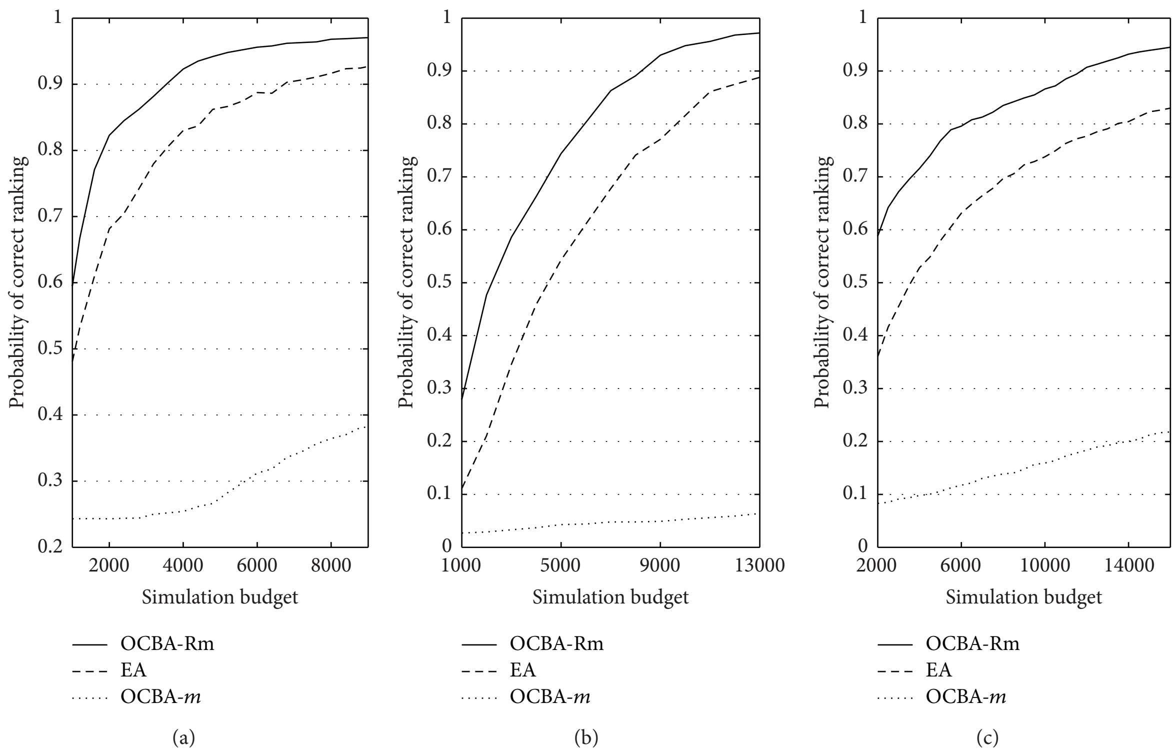

Figure 1: Probability of correctly ranking the top m designs under three different allocation rules.

than EA in all three experiments when the objective is to identify the ranking of the top m designs. This is because OCBA-m only focuses on distinguishing the set of the top m designs and the set of the nontop m designs without considering the ranking within the top m designs. Based on the numerical results, we can see that it is important to derive the OCBA-Rm when the ranking of the top m alternatives is needed. By using the proposed OCBA-Rm allocation rule, significant number of simulation budgets can be saved compared with EA and OCBA-m.

7. Conclusion

In this paper, we study the problem of simulation budget allocation of ranking the top m designs out of k alternatives. Based on the large deviation theory, we have derived the asymptotically optimal allocation rule to maximize the probability of correctly ranking the top m designs. In addition, a heuristic sequential allocation algorithm is suggested to implement the simulation budget allocation rule. Numerical experiments are conducted to compare the effectiveness of the proposed simulation budget allocation with some existing allocation procedures. The contribution of this paper is twofold. From a ranking and selection perspective, we offer a heuristic for ranking the top m designs out of k alternatives, where our empirical studies show that it can be more efficient than the existing methods. From the computing budget allocation perspective, our heuristic illustrates how the previous OCBA method for identifying the single best design can be modified to identify the ranking of the top m designs. In particular, we derive the asymptotically optimal allocation using large deviation theory, which allows us to remove the assumption that the performance of each design must be normally distributed.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank the editor and the anonymous reviewers for the constructive comments which have improved the quality of this paper.

References

[1] M. Law and W. D. Kelton, *Simulation Modeling & Analysis*, Industrial Engineering and Management Science, McGraw-Hill, New York, NY, USA, 2001.

[2] Y. C. Ho, R. S. Sreenivas, and P. Vakili, "Ordinal optimization of DEDS," *Discrete Event Dynamic Systems*, vol. 2, no. 1, pp. 61–88, 1992.

[3] L. Dai, "Convergence properties of ordinal comparison in the simulation of discrete event dynamic systems," *Journal of Optimization Theory and Applications*, vol. 91, no. 2, pp. 363–388, 1996.

[4] W. B. Gong, Y. C. Ho, and W. Zhai, "Stochastic comparison algorithm for discrete optimization with estimation," *SIAM Journal on Optimization*, vol. 10, no. 2, pp. 384–404, 2000.

[5] H. C. Chen, C. H. Chen, L. Dai, and E. Yucesan, "New development of optimal computing budget allocation for discrete event simulation," in *Proceedings of the Winter Simulation Conference*, pp. 334–341, December 1997.

[6] C. H. Chen, J. Lin, E. Yücesan, and S. E. Chick, "Simulation budget allocation for further enhancing the efficiency of ordinal optimization," *Discrete Event Dynamic Systems: Theory and Applications*, vol. 10, no. 3, pp. 251–270, 2000.

[7] S. H. Kim and B. L. Nelson, "Chapter 17: selecting the Best System," in *Handbooks in Operations Research and Management Science*, G. H. Shane and L. N. Barry, Eds., vol. 13, pp. 501–534, Elsevier, 2006.

[8] S. E. Chick and K. Inoue, "New two-stage and sequential procedures for selecting the best simulated system," *Operations Research*, vol. 49, no. 5, pp. 732–743, 2001.

[9] S. E. Chick and K. Inoue, "New procedures to select the best simulated system using common random numbers," *Management Science*, vol. 47, no. 8, pp. 1133–1149, 2001.

[10] J. Branke, S. E. Chick, and C. Schmidt, "Selecting a selection procedure," *Management Science*, vol. 53, no. 12, pp. 1916–1932, 2007.

[11] L. W. Koenig and A. M. Law, "A procedure for selecting a subset of size m containing the l best of k independent normal populations," *Communications in Statistics—Simulation and Computation*, vol. 14, pp. 719–734, 1985.

[12] C. H. Chen, D. He, M. Fu, and L. H. Lee, "Efficient simulation budget allocation for selecting an optimal subset," *INFORMS Journal on Computing*, vol. 20, no. 4, pp. 579–595, 2008.

[13] S. Zhang, L. H. Lee, E. P. Chew, C. H. Chen, and H. Y. Jen, "An improved simulation budget allocation procedure to efficiently select the optimal subset of many alternatives," in *Proceedings of the IEEE International Conference on Automation Science and Engineering: Green Automation Toward a Sustainable Society (CASE '12)*, pp. 230–236, August 2012.

[14] L. H. Lee, E. P. Chew, S. Teng, and D. Goldsman, "Finding the non-dominated Pareto set for multi-objective simulation models," *IIE Transactions*, vol. 42, no. 9, pp. 656–674, 2010.

[15] L. H. Lee, N. A. Pujowidianto, L. W. Li, C. H. Chen, and C. M. Yap, "Approximate simulation budget allocation for selecting the best design in the presence of stochastic constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 11, pp. 2940–2945, 2012.

[16] S. R. Hunter and R. Pasupathy, "Optimal sampling laws for stochastically constrained simulation optimization on finite sets," *INFORMS Journal on Computing*, vol. 25, no. 3, pp. 527–542, 2013.

[17] H. Xiao, L. H. Lee, and K. M. Ng, "Optimal computing budget allocation for complete ranking," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 2, pp. 516–524, 2014.

[18] C. H. Chen and L. H. Lee, *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*, World Scientific, Singapore, 2010.

[19] P. Glynn and S. Juneja, "A large deviations perspective on ordinal optimization," in *Proceedings of the Winter Simulation Conference*, pp. 577–584, December 2004.

[20] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer, New York, NY, USA, 1998.