Efficient Sampling Policy for Selecting a Subset With the Best

Gongbo Zhang ♦, Bin Chen ♦, Qing-Shan Jia ♦, Senior Member, IEEE, and Yijie Peng ♦, Member, IEEE

Abstract—In this article, we study the problem of selecting a subset with the best of a finite number of alternatives under a fixed simulation budget. Our work aims to maximize the posterior probability of correctly selecting such a subset. We formulate the dynamic sampling decision as a stochastic control problem in a Bayesian setting. In an approximate dynamic programming paradigm, we propose a sequential sampling policy based on value function approximation. We analyze the asymptotic property of the proposed sampling policy. Numerical experiments demonstrate the efficiency of the proposed procedure.

Index Terms—Bayesian, ranking and selection (R&S), sequential sampling, stochastic control, subset selection.

I. INTRODUCTION

Simulation optimization has been widely used in analyzing modern complex systems, e.g., healthcare, transportation, manufacturing, and supply chain systems [1]. Simulation experiments are often expensive and time-consuming since a large number of simulation replications are required in order to achieve an accurate estimate of the system performance [2]. In this note, we consider the problem of selecting a subset of size \( m \) containing the best of \( k \) alternatives, \( 1 \leq m < k \), \( m, k \in \mathbb{Z}^+ \). The true performance of each alternative is unknown in practice, which may be estimated by Monte Carlo simulation under a finite sampling budget \( T \).

When \( m = 1 \), the problem reduces to the ranking and selection (R&S) for finding the best. R&S has been widely studied in simulation, which aims to allocate simulation replications to alternatives for efficiently selecting the best alternative. There are fixed-precision and fixed-budget procedures for R&S [3], where the former allocates samples to guarantee a probability of correct selection (PCS) up to a certain level, whereas the latter optimizes performance metrics under a fixed simulation budget constraint. A well-researched paradigm is an indifference zone (IZ) framework [4], [5], which aims to guarantee a PCS level in the least favorable configuration. The sampling procedures in the IZ framework tend to allocate more replications than necessary in practice. Well-known methods to enhance the efficiency for finding the best alternative include optimal computing budget allocation (OCBA) [6], expected value of information [7], [8], knowledge gradient [9], [10], expected improvement [11], [12], and asymptotically optimal allocation procedure (AOAP) [13]. OCBA was initially a two-stage procedure, and becomes fully sequential by combining with a certain sequential rule, e.g., “most starving rule” [2]. The fully sequential procedures usually lead to higher PCS under the specified simulation budget, or achieve the same level of PCS using fewer simulation replications than the two-stage procedures [6]. AOAP is derived under a stochastic control framework, and it is proved to achieve the asymptotically optimal sampling ratios for selecting the best with a normal sampling distribution [14]. Our problem softens the optimization problem by allowing a subset containing the best to be acceptable, so that decision can be made in a more flexible way. Instead of selecting a single alternative as the best, our problem formulation can dramatically reduce the computational cost and it is particularly useful for many engineering applications when finding the best is not necessary and simulation is expensive. For example, in an emergency department healthcare staffing problem, the number of staffing designs is large and circumstances of patients are complex; therefore, finding a subset with the best of designs can reduce the time for decision-making. In addition, selecting a subset with the best can efficiently screen out some alternatives at the first stage before seeking an optimal decision accurately at the second stage.

Compared with the R&S for selecting the best alternative, the literature of finding a subset with the best is still sparse. Koenig and Law [15] developed a two-stage IZ procedure to select a subset of size \( m \) containing the top-\( f \) of \( k \) alternatives, \( 1 \leq f \leq m < k \). Following the pioneer work [15] using the least favorable configuration, Chen [16] proposes a two-stage procedure to select a subset of size \( m \) containing at least \( c \) of the top-\( f \) of \( k \) alternatives with unknown means and unknown variances, \( 1 \leq c \leq f \). The most relevant work to us is [17], which studies the problem of selecting \( f \) good alternatives out of \( m \) acceptable alternatives from \( k \) alternatives under a fixed simulation budget constraint. Yet strictly speaking, the objective in [17] is to correctly select a subset of size \( m \) containing the top-\( f \) of \( k \) alternatives. By relaxing the objective function using a Bonferroni inequality and asymptotic analysis, Gao and Gao [17] proposed a sampling procedure that screens out some alternatives at the first stage. In contrast to [17], we formulate the sampling decision as a stochastic control problem. Although we do not screen out any alternative, it can be proved that following our procedure, the sampling ratios of \( (m-1) \) alternatives converge to 0 almost surely as the number of simulation budget goes to infinity, which explains why selecting a subset with the best could reduce the computational cost.

Our work is related to the literature on subset selection, which selects a subset of random size that contains the best alternative with a high probability. Different than our work, subset selection procedures guarantee the PCS level and lead to a small expected subset size [18],

---

Manuscript received 21 December 2021; revised 16 June 2022; accepted 14 September 2022. Date of publication 20 September 2022; date of current version 28 July 2023. An earlier version of this paper was presented in part at the 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE) [DOI: 10.1109/CASE48305.2020.9216838]. This work was supported by the National Natural Science Foundation of China under Grants 72022001, 71901003, 62073182, and 62125304. Recommended by Associate Editor R. Jain. (Corresponding authors: Bin Chen; Yijie Peng.)

Gongbo Zhang and Yijie Peng are with the Department of Management Science and Information Systems, Guanghua School of Management, Peking University, Beijing 100084, China (e-mail: gongbozhang@pku.edu.cn; pengyijie@pku.edu.cn).

Bin Chen is with the College of Systems Engineering, National University of Defense Technology, Changsha 410073, China (e-mail: nudtcgb9372@gmail.com).

Qing-Shan Jia is with the Center for Intelligent and Networked Systems (CFINS), Department of Automation, Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China (e-mail: jiaqqs@tsinghua.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2022.3207871.

Digital Object Identifier 10.1109/TAC.2022.3207871

---

IEEE TRANSACTIONS ON AUTOMATIC CONTROL, VOL. 68, NO. 8, AUGUST 2023

0018-9286 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
follows a prior distribution and the $\in = \in +1$)

alternatives, and the performance of each alternative is distinguishable. The objective of the problem is to select a subset with the best based on a stochastic framework [21], [22], [23]. Another line of literature is a selection problem and the optimal policy for the Bellman equation of MDP

$\Theta$, where $\hat{\in} = \in +1$)

$\in$ is the state at stage $\in$.
The proof of the proposition can be found in [27, Appendix]. We aim to find a dynamic sampling decision such that (1) is maximized. Since the closed-form expression of (1) is unknown, we provide a lower bound of (1)

\begin{equation}
\max_{i=1,\ldots,m} \Pr \left\{ \bigcap_{j=m+1}^{k} \left( \mu_{(i)T} > \mu_{(j)T} \right) \big| \mathcal{E}_T \right\} \tag{2}
\end{equation}

which would typically become tight as $T$ grows large.

In this work, we propose a dynamic allocation policy $\mathcal{A}_T$ to maximize (2). The dynamic sampling decision can be captured by a stochastic control problem. Under the Bayesian setting, we recursively define the expected payoff for $\mathcal{A}_T$ by

$V_i(\mathcal{E}_T; \mathcal{A}_T) \triangleq \max_{i=1,\ldots,m} \Pr \left( \bigcap_{j=m+1}^{k} \left( \mu_{(i)T} > \mu_{(j)T} \right) \bigg| \mathcal{E}_T \right)$,

and for $0 \leq t < T$

$V_t(\mathcal{E}_t; \mathcal{A}_T) \triangleq \mathbb{E} \left[ V_{t+1}(\mathcal{E}_{t+1} \cup \{X_{t+1}\}; \mathcal{A}_{t+1}) \big| \mathcal{E}_t \right]_{i=A_{t+1}(\mathcal{E}_t)}$.

Then, the optimal allocation policy $\mathcal{A}_T$ can be well defined by $\mathcal{A}_T = \arg \max_{\mathcal{A}_T} V_0(\zeta_0; \mathcal{A}_T)$. The stochastic control problem can be viewed as a Markov decision process (MDP) with $(T+1)$ stages, where $\zeta_0$ is the state at stage 0, and $\mathcal{E}_t$ is the state at stage $t$, $0 < t \leq T$. The actions correspond to $A_{t+1}$ for $0 \leq t < T$ and the terminal selection action for $t = T$. The transition is $\mathcal{E}_t \rightarrow \mathcal{E}_{t+1}$, where $\mathcal{E}_{t+1} \triangleq \mathcal{E}_t \cup \{X_{t+1}\}$. The only nonzero reward is the terminal reward $V_T(\mathcal{E}_T)$. By induction, the solution to the stochastic control problem and the optimal policy for the Bellman equation of MDP are equivalent [28]. We can recursively solve the $\mathcal{A}_T$ by the Bellman equation: $A_{t+1}(\mathcal{E}_t) = \arg \max_{i=1,\ldots,m} \mathbb{E} \left[ V_{t+1}(\mathcal{E}_{t+1} \cup \{X_{t+1}\}) \big| \mathcal{E}_t \right]_{i=A_{t+1}(\mathcal{E}_t)}$, $0 \leq t < T$. Finding $\mathcal{A}_T$ through backward induction suffers from the curse-of-dimensionality [13]. To address the computational difficulty, we adopt an AOAs paradigm [29], which makes dynamic decision based on VFA and keeps learning the value function with decisions moving forward.

Suppose $X_{t+1} \sim_{iid} N(\mu_{t+1}, \sigma_{t+1}^2)$, $i = 1, \ldots, k$, with unknown means and known variances. To make the proposed allocation policy amenable to practical implementation, we focus on the known variance case and use the sample estimate as a plug-in for the true value. If the normal assumption is not satisfied, a macro replication obtained from a batched mean follows an approximately normal distribution by the central limit theorem. Similar to [30], we make the following assumption.

Assumption 1: $\mu_{i, t} \in \{1, \ldots, k\}$, forms a conjugate normal prior distribution.

By [31], the conjugate prior for a normal sampling distribution $N(\mu_{i, t}, \sigma_i^2)$ with unknown mean and known variance is also a normal distribution $N(\mu_i^{(0)}, \sigma_i^{(0)})$. The posterior distribution of $\mu_i$ is $N(\mu_i^{(t)}, \sigma_i^{(t)})$, where

$\sigma_i^{(t)} = \left( \frac{1}{\sigma_i^{(0)}} \right)^2 + \left( \frac{t_i}{\sigma_i^2} \right)^{-1}$

$\mu_i^{(t)} = \left( \frac{\mu_i^{(0)}}{\sigma_i^{(0)}} \right)^2 + \left( \frac{t_i m_i^{(t)}}{\sigma_i^2} \right)$,

$m_i^{(t)} = \sum_{t=1}^{t_i} X_{i, t}$.
If $\sigma^{(0)} \to \infty$, we have $\mu_{i}^{(t)} = m_{i}^{(t)}$, and such a prior is uninformative. For a normal distribution with unknown variance, there is a normal-gamma conjugate prior [31]. By conjugacy, $\mathcal{E}_{t}$ is completely determined by the posterior hyperparameters, and the dimension of $\mathcal{E}_{t}$ is fixed at any step.

### III. Dynamic Sampling Policy

In this section, in order to derive a dynamic sampling procedure with an analytical form, we adopt the VFA technique in [13], which uses a single feature of the value function one-step ahead. Specifically, suppose any step $t$ could be the last step. The joint distribution of vector

$$
\left( \mu_{(i)}, \mu_{(m+1)}, \ldots, \mu_{(i)}, \mu_{(k)} \right), \quad i \in \{1, \ldots, m\}
$$

follows a joint normal distribution with mean vector

$$
\left( \mu_{(i)}, \mu_{(m+1)}, \ldots, \mu_{(i)}, \mu_{(k)} \right), \quad i \in \{1, \ldots, m\}
$$

and covariance matrix $\Sigma_{i}^{(t)} = \begin{pmatrix} \Gamma \Lambda^{(i)} \Gamma \end{pmatrix}$, where $'$ denotes the transpose operation of the matrix, $\Gamma \in \mathbb{R}^{(k-m+1) \times (k-m)}$

and $\Lambda^{(i)} \in \mathbb{R}^{(k-m+1) \times (k-m+1)}$ is a diagonal matrix,

$$
\Lambda^{(i)} \triangleq \text{diag}(\sigma^{(t)}_{(i)} \sigma^{(t)}_{(m+1)} \ldots \sigma^{(t)}_{(i)} \sigma^{(t)}_{(k)}).
$$

In addition, for $i \in \{1, \ldots, m\}$

$$
\text{Pr} \left\{ \mu_{(i)} > \mu_{(j)}, j = m + 1, \ldots, k \right\} \mathcal{E}_{t}
$$

$$
= \text{Pr} \left\{ \sum_{k=1}^{m} z_{k} \mathcal{N}_{(i)} > \mu_{(j)}, j = m + 1, \ldots, k \right\}
$$

$$
= \frac{1}{(2\pi)^{(k-m)/2}} \int_{\sum_{k=1}^{m} z_{k} \mathcal{N}_{(i)} > \mu_{(j)}, j = m + 1, \ldots, k} \exp\left( -\frac{1}{2} \sum_{k=1}^{m} z_{k}^{2} \mathcal{N}_{(i)} \right) dz_{1} \ldots dz_{k-m}, j = m + 1, \ldots, k (3)
$$

where $L^{(i)} \triangleq \left( \begin{array}{ccc} l_{1}^{(i)} & \cdots & l_{k-m}^{(i)} \\ \cdots & \cdots & \cdots \\ l_{k-m}^{(i)} & \cdots & l_{1}^{(i)} \end{array} \right) \in \mathbb{R}^{(k-m) \times (k-m)}$, $j = m + 1, \ldots, k$, $i = 1, \ldots, k$, $i = 1, \ldots, k$. $z_{k} \sim \mathcal{N}(0, 1)$. Therefore, (3) is an integral of the density of $(k-m)$-dimensional standard normal distribution over a region formed by some hyperplanes, and the value function becomes the maximum value of $m$ integrals. We visualize (3) and its approximation when $k = 5$, $m = 2$, and $i = 1$. In Fig. 1, (3) is the integration of the 3-D standard normal density over the shadowed area, and an approximation is captured by the size of the largest centrally tangent ball with the radius $d_{i,j}$. In general, the VFA is given by $\tilde{V}_{t}(\mathcal{E}_{t}) \triangleq \max_{i} d_{i}(\mathcal{E}_{t})$, such that $d_{i}(\mathcal{E}_{t}) = \min_{i} d_{i}(\mathcal{E}_{t+1}(\mathcal{E}_{t}), \ldots, d_{k}(\mathcal{E}_{t+1}))$, where

$$
d_{i,j}(\mathcal{E}_{t}) = \frac{\mu_{(i)}^{(t)} - \mu_{(j)}^{(t)}}{\sqrt{\sigma^{(t)}_{(i)}^{2} + \sigma^{(t)}_{(j)}^{2}}}, \quad i = 1, \ldots, m
$$

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

As $t_{h} \to \infty$, by the law of large numbers (LLN), $\lim_{t_{h} \to \infty} \mu_{k}^{(t)} = \mu_{k}$, a.s. and $\lim_{t_{h} \to \infty} \sigma_{k}^{(t)} = 0$, a.s. If alternatives $i$ and $j$ are sampled infinitely often as $t$ goes to infinity, we have $\lim_{t_{h} \to \infty} d_{i,j}(\mathcal{E}_{t}) = \infty$. The VFA is reasonable since the exponential decreasing rate of the normal density.

**Proposition 2**: The error of the integral of the $(k-m)$ dimensional standard normal density over a centered ball with a radius $d_{i}(\mathcal{E}_{t})$ as an approximation of (3) decreases to 0 in an exponential rate as $d_{i}(\mathcal{E}_{t}) \to \infty$, $i \in \{1, \ldots, m\}$.

The proof of Proposition 2 can be found in [27, Appendix]. If the $(t+1)$th replication is the last one, a VFA one-step look ahead is given as $\tilde{V}_{t+1}(\mathcal{E}_{t+1}) \triangleq \mathbb{E} \left[ \tilde{V}_{t+1}(\mathcal{E}_{t+1}) \right]$. With a certainty equivalence, which replaces stochastic quantities by their expected values [32], a further approximation to soften the computational difficulty of the expectation term can be given as $\tilde{V}_{t}(\mathcal{E}_{t+1}) \triangleq \mathbb{E} \left[ \tilde{V}_{t+1}(\mathcal{E}_{t+1}) \right]$. For $i, h, j = 1, \ldots, k$. We use $\tilde{V}_{t}(\mathcal{E}_{t+1})$ as the criterion for making dynamic sampling decisions. For $i, h, j = 1, \ldots, k$, and $i \neq h, j \neq \ell$

$$
\tilde{V}_{t}(\mathcal{E}_{t+1}) = \max_{i,j} \left( \min_{j} \frac{\left( \mu_{(i,j)} - \mu_{(j)}^{(t)} \right)^{2}}{\left( \sigma_{(i,j)}^{(t)} + \sigma_{(j)}^{(t)} \right)^{2}}, \min_{j} \frac{\left( \mu_{(i,h)} - \mu_{(j)}^{(t)} \right)^{2}}{\left( \sigma_{(i,h)}^{(t)} + \sigma_{(j)}^{(t)} \right)^{2}} \right)
$$

and

$$
(\sigma_{\ell}^{(t+1)})^{2} = \left( \frac{1}{\sigma_{\ell}^{(t+1)}}^{2} + \frac{t_{\ell}}{\sigma_{\ell}^{2}} \right)^{-1}, \quad \ell = 1, \ldots, k.
$$

A fully sequential allocation procedure is given by

$$
\tilde{A}_{t+1}(\mathcal{E}_{t}) \in \text{arg max}_{i=1, \ldots, k} \tilde{V}_{t}(\mathcal{E}_{t+1}).
$$

The VFA can be rewritten as $\max_{i=1, \ldots, m} \min_{j=m+1, \ldots, k} 1/c_{i,j}^{2}$ ($i, j$), where $c_{i}(i, j)$ is the posterior noise–signal ratio (coefficient of variation) of $\mu_{(i)} - \mu_{(j)}$. The larger is $c_{i}(i, j)$, the more difficult to correctly identify the sign of the difference of
\( \Phi \leq 0 \)

Let \( |B| < 0 \), with \( \mu \exists \), which contradicts \( +1 \). Following [14], \( \Phi \), with respect to \( - \in B \), \( \forall \mu > 0 \), \( \Phi = 0 \), . . . , \( k \). Let \( 1 \) for \( \Phi \), \( i / \forall \Phi \lim a \Phi \lim 1 \geq \lim \). The following theorem further establishes that only \((k - m + 1)\) alternatives will be sampled infinitely often almost surely as \( t \to \infty \).

**Theorem 2:** For any \( \Gamma \subset T \) and \( |\Gamma| > 1 \), suppose conditions (5)–(7) do not hold simultaneously. Then, we have \( |\Phi| = k - m + 1 \).

\[
\min_{j \in B} G_{ij}(r_i, r_j) = \min_{h \in B} G_{hj}(r_h, r_j) \quad (5)
\]

\[
\frac{\partial}{\partial x} J_i = B \quad (6)
\]

\[
\frac{\partial}{\partial x} \min_{j \in B} G_{ij}(x, r_j) \mid_{x=r_i} = \min_{h \in B} \frac{\partial}{\partial x} \min_{j \in B} G_{hj}(x, r_j) \quad \forall i \in \Gamma \quad (7)
\]

where \( \sum_{i \in \Gamma} r_i + \sum_{j \in B} r_j = 1 \).

**Proof:** Based on Theorem 1, the proof boils down to the question of verifying only one alternative in \( T \) will be sampled infinitely often almost surely following (4).

Let \( \gamma_h(t) = h/t, \sum_{k=1}^{h-1} t = 1, \ h = 1, \ldots, k \). By the LLN, \( \lim_{t \to \infty} \mu_h(t) = \mu_h, \ h \in \Phi \). Since the asymptotic sampling ratios will be determined by the increasing order of \( d_{ij}(E_t) \) with respect to \( t \), we replace \( \mu_h(t) \) and \( \sigma_{ij}^2(t) \) with \( \mu_h \) and \( \sigma_{ij}^2/t_h \) in \( d_{ij}(E_t) \) for simplicity of analysis. Note that \( 0 \leq \gamma_h(t) < 1, \ h = 1, \ldots, k, \) and \((r_1(t), \ldots, r_k(t))\) is a bounded sequence. By the Bolzano–Weierstrass theorem [33], there exists a subsequence of \((r_1(\cdot), \ldots, r_k(\cdot))\) converging to \((\tilde{r}_1, \ldots, \tilde{r}_k)\), where \( \sum_{k=1}^{h} \tilde{r}_i = 1, \tilde{r}_i > 0 \). Without loss of generality, we assume that \((r_1(t), \ldots, r_k(t))\) converges to \((\tilde{r}_1, \ldots, \tilde{r}_k)\); otherwise, the following argument is made over a subsequence. Note that

\[
\lim_{t \to \infty} \left[ \frac{(\mu_i - \mu_j)^2}{\sigma_i^2 / (t_i + 1) + \sigma_j^2 / t_j} - \frac{(\mu_i - \mu_j)^2}{\sigma_i^2 / t_i + \sigma_j^2 / t_j} \right] = \lim_{t \to \infty} t \left[ G_{ij} \left( r_{ij}(t) + 1 / t, r_{ij}(t) \right) - G_{ij} \left( r_{ij}(t), r_{ij}(t) \right) \right] = \lim_{t \to \infty} \frac{\partial G_{ij}}{\partial x} \left( x, r_{ij}(t) \right) \mid_{x=r_{ij}(t)} = 0
\]
which contradicts the sampling rule (4). We then claim that \( \exists j \in B \) such that \( \tilde{r}_j > 0 \); otherwise, \( \exists i \in T \cap \Phi \) such that \( \tilde{r}_i > 0 \), and
\[
\lim_{t \to \infty} \frac{\partial G_{ij}(x, r_j(t))}{\partial x}
\bigg|_{x = r_j(t)} = 0,
\lim_{t \to \infty} \frac{\partial G_{ij}(r_i(t), x)}{\partial x}
\bigg|_{x = r_i(t)} > 0
\]
which contradicts the sampling rule (4). In addition, we claim that \( \forall j \in B \) such that \( \tilde{r}_j > 0 \); otherwise, \( \exists \ell, q \in B \), and \( \ell \neq q \) such that \( \tilde{r}_\ell > 0, \tilde{r}_q = 0 \), and for \( i \in T \cap \Phi \) such that \( \tilde{r}_i > 0 \),
\[
\lim_{t \to \infty} G_{ij}(r_i(t), r_\ell(t)) > 0, \lim_{t \to \infty} G_{ij}(r_i(t), r_q(t)) = 0
\]
which contradicts the sampling rule (4). Finally, we claim that \( \forall i \in T \cap \Phi \) such that \( \tilde{r}_i > 0 \); otherwise, \( \exists h, p \in T \cap \Phi, h \neq p \), such that \( \tilde{r}_h > 0, \tilde{r}_p = 0 \), and for \( j \in B \)
\[
\lim_{t \to \infty} G_{hj}(\ell_h(t), r_j(t)) > 0, \lim_{t \to \infty} G_{hj}(\ell_h(t), r_p(t)) = 0
\]
which contradicts the sampling rule (4). Summarizing the aforementioned, we have \( \tilde{r}_i > 0, \ell \in \Phi \).

Following the sampling rule (4), we will show that \( \tilde{r}_i, \ell \in \Phi \) satisfy (5) and (6) with \( \Gamma = T \cap \Phi \). If \( |T \cap \Phi| = 1 \), then (5) holds automatically. If \( |T \cap \Phi| > 1 \) and (5) does not satisfy, then \( \exists i, \ell \in T \cap \Phi, i \neq h \), such that the inequality
\[
\min_{j \in B} G_{ij}(\tilde{r}_i, \ell_j) > \min_{j \in B} G_{ij}(\tilde{r}_h, \ell_j)
\]
strictly holds, and then, there exists \( T_0 > 0 \), such that after step \( t > T_0 \)
due to the continuity of the function \( G_{ij}(r_i, r_j) \) with respect to \( r_i, r_j, \mu_i, \mu_j, \sigma_i, \) and \( \sigma_j \). By the sampling rule (4), alternative \( i \) will be sampled and alternative \( h \) will stop receiving replications before the inequality reverses. Therefore, violation of condition (5) leads to a contradiction to that \( (r_i(t), \ldots, r_k(t)) \) converges to \( (\tilde{r}_i, \ldots, \tilde{r}_k) \), which implies that (5) must hold.

If (6) does not hold, then \( \exists \ell \in B \) such that for \( i \in T \cap \Phi \), the inequality
\[
G_{i\ell}(\tilde{r}_i, \ell_\ell) > \min_{j \in B} G_{ij}(\tilde{r}_i, \ell_j)
\]
strictly holds. By a similar argument, alternative \( \ell \) will stop receiving replications before the inequality reverses, and then (6) must hold.

If \( |T \cap \Phi| > 1 \), then by the sampling rule (4), \( \tilde{r}_i, \ell \in \Phi \), satisfy (8) with \( \Gamma = T \cap \Phi \); otherwise, \( \exists i, h \in T \cap \Phi \) such that
\[
\lim_{t \to \infty} \left[ \min_{j \in B} \frac{(\mu_i - \mu_j)^2}{\sigma_i^2/(t_1 + 1) + \sigma_j^2/t_j} - \min_{j \in B} \frac{(\mu_i - \mu_j)^2}{\sigma_i^2/t_1 + \sigma_j^2/t_j} \right]
\]
which contradicts the sampling rule (4). By assumption, (5)–(7) do not hold simultaneously with \( \Gamma = T \cap \Phi \) and \( |T \cap \Phi| > 1 \). Therefore, \( |\Phi| = k - m + 1 \), which proves the theorem.

**Remark 1:** Equations (5) and (6) balance the large deviations rates for some pairwise comparisons of alternatives. Equation (7) balances the marginal increments of the large deviations rates for some pairwise comparisons with respect to the number of replications allocated to alternative \( i \in \Gamma \). The condition of Theorem 2 can be checked straightforwardly in a special case when \( \sigma_i = \cdots = \sigma_k \). In general, the condition could be checked numerically.

**Corollary 1:** Following the sampling rule (4), the sampling ratio of each alternative satisfy:
\[
\lim_{t \to \infty} r_i(t) = \begin{cases} r_i & i \in \Phi \\ 0 & \text{else} \end{cases} \text{a.s.}
\]
where \( \sum_{i=1}^{k} r_i = 1, \) and for \( i \neq j \in T \cap \Phi, j \in B \)
\[
\frac{(\mu_i - \mu_j)^2}{\sigma_i^2/r_i + \sigma_j^2/r_j} = \frac{(\mu_i - \mu_j)^2}{\sigma_i^2/r_i + \sigma_j^2/r_j}
\]
and
\[
r_i' = \sigma_i \sqrt{\sum_{j \in B} r_j^2/\sigma_j^2}.
\]

**Remark 2:** The proof of Corollary 1 is essentially the same as the proof for [13, Th. 3], which establishes that the sampling ratios determined by (8) and (9) are asymptotically optimal for selecting the best alternative. The proposed sequential sampling policy (4) for selecting a subset with the best is denoted as AOA. AOA dynamically determines the alternative that stops receiving simulation replications based on collected sample information. We caution that the sets \( T \) and \( B \), and the alternative \( T \cap \Phi \) cannot be uniquely determined in our theory, since they depend on the realization of simulation replications. Compared with selecting the single best alternative, where all of the \( k \) alternatives are sampled infinitely often, the sampling ratios of \( (m - 1) \) alternatives converge to 0 as \( t \to \infty \), which significantly reduces the computational cost.

**Proposition 3:** The solution to (8) and (9) is unique.

**Proof:** Suppose \( \bar{r} = (\bar{r}_1, \ldots, \bar{r}_{m+1}, \ldots, \bar{r}_k) \) and \( r^* = (r^*_1, \ldots, r^*_m, \ldots, r^*_k) \) are two different solutions to (8) and (9). Following [14], \( \bar{r} \) and \( r^* \) are derived by the Karush–Kuhn–Tucker conditions of a concave maximization optimization problem. Then, \( \bar{r}^* = (1 - \lambda) r^* + \lambda r^* \), \( \lambda \in [0, 1] \), are solutions to (8) and (9) since the optimal set of a convex optimization problem is convex [34]. By the chain rule, (8) and (9) and \( \sum_{i=1}^{k} r_i = 1 \) lead to \( \Psi' \cdot \bar{R} = 0 \), where \( \Psi' \in \mathbb{R}^{(k-m+1) \times (k-m+1)} \) is defined in (10) shown at the bottom of the next page, where for \( i = T \cap \Phi, j \in B \)
\[
\phi_{ij} \triangleq \frac{\partial G_{ij}(r_i', y)}{\partial y}
\bigg|_{y = r_i'} \quad \psi_{ij} \triangleq \frac{\partial G_{ij}(y, r_j')}{\partial y}
\bigg|_{y = r_j'}
\]
\[
\bar{r}_i \triangleq 2r_i, \quad \bar{r}_i' \triangleq 2r_i', \quad i \in \Phi
\]
\[
\bar{R} \triangleq \begin{pmatrix} \bar{r}_{m+1} & \cdots & \bar{r}_{m+k} & \cdots & \bar{r}_k & \bar{r}_k' \end{pmatrix}' \in \mathbb{R}^{(k-m+1) \times 1}
\]
\[
O \triangleq \begin{pmatrix} 0, 0, \ldots, 0 \end{pmatrix}' \in \mathbb{R}^{(k-m+1) \times 1}.
\]

Following [14], \( \phi_{ij} > 0 \) and \( \psi_{ij} > 0, j \in B \), and we have \( \text{det}(\Psi') \neq 0 \) and \( \Psi' \) is invertible, which yields \( \bar{R} = 0 \). Then \( r^*_i = \tilde{r}_i, \quad i \in \Phi, \) which contradicts that \( r^* \) and \( \bar{r} \) are two different solutions to (8) and (9). Therefore, the solution to (8) and (9) is unique, which proves the proposition. ■

**Remark 3:** A closely related problem to soften the objective for finding the best is good alternative selection. Let \( F = \{(1), \ldots, (m)\} \) be the top-\( m \) subset. The posterior PCS for finding a good alternative
\( \langle i^* \rangle_T \) can be expressed as

\[
\text{IPCS}_T = \Pr \{ \langle i^* \rangle_T \in \mathcal{F}^m \mid \mathcal{E}_T \} = \Pr \left\{ \bigcup_{J \subseteq \{1, \ldots, k\}} \bigcap_{j \in J} \left\{ \mu(i^*)_T > \mu(j)_T \right\} \left| \mathcal{E}_T \right. \right\}.
\]

(11)

The derivation of a dynamic allocation procedure for finding a good alternative based on (11) is essentially the same as the proposed sequential sampling rule (4); however, the former needs to enumerate all of the possible cases of the set \( J \) [in general, there are \( C_{k-1}^{m-1} \) cases, where \( C_{k-1}^{m-1} \) denotes the number of combinations by choosing \( (m - 1) \) elements from \( (k - 1) \) elements], leading to a much higher computational complexity than our method. In [27, Appendix], there are numerical examples to illustrate that for selecting a good alternative, the proposed allocation procedure (4) has a comparable performance to the allocation procedure derived based on (11).

IV. NUMERICAL EXPERIMENTS

In this section, we provide numerical experiments to demonstrate the efficiency of the proposed AOAs policy. The detail of implementation for AOAs is shown in [27, Appendix]. The equal allocation (EA), optimal computing budget allocation for \( r \) good designs (OCBA-rgm) [17], and OCBA-rgm given the true parameter (OCBA-rgmt) are implemented for comparisons. EA equally allocates simulation replications to each alternative, i.e., roughly \( T/k \) replications for each alternative. OCBA-rgm derives simulation replications based on some asymptotic conditions, and \((m - r)\) alternatives stop receiving replications at each step. The details of implementation for OCBA-rgm can be found in [27, Appendix]. OCBA-rgmt allocates simulation replications according to the OCBA-rgm rule under the perfect information (assuming the true parameters are known), whereas OCBA-rgm allocates simulation replications according to the OCBA-rgm rule with parameters estimated sequentially by available sample information. OCBA-rgmt is compared to show that the difference between the asymptotic property and finite-sample performance of a sequential allocation procedure. The efficiency of each procedure is measured by the IPCS, which is reported as a function of sampling budget \( T \), i.e., \( \text{IPCS}_T = E[\{i \in \mathcal{F}_T^m \}] \). In each numerical experiment, we set \( n_0 = 10 \), and IPCS is estimated by 100,000 independent macro experiments. The code for the numerical experiments in this note is available online.\(^1\)

Experiment 1: The allocation procedures are tested in a synthetic experiment with 50 competing alternatives. The size of the selected subset is 5. In each macro experiment, the performance of each alternative is generated from a normal distribution \( N(0, ((51 - i)/10)^2) \), \( i = 1, \ldots, 50 \). The simulation replications are drawn independently from \( N(\mu_i, \sigma_i^2) \), where \( \sigma_i = 51 - i \). The total simulation budget is \( T = 1000 \).

\[
\Psi \triangleq \begin{pmatrix}
\phi_{i(m+1)} & -\phi_{i(m+2)} & \cdots & 0 & 0 & \psi_{i(m+1)} - \psi_{i(m+2)} \\
0 & \phi_{i(m+2)} & \cdots & 0 & 0 & \psi_{i(m+2)} - \psi_{i(m+3)} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & \phi_{i(k-1)} - \psi_{i(k)} & \psi_{i(k-1)} - \psi_{i(k)} \\
-\psi_{m+1} & -\psi_{m+2} & \cdots & -\psi_{k-1} & -\psi_k & \psi_i \\
1 & 1 & \cdots & 1 & 1 & 1
\end{pmatrix}.
\]

(10)

From Fig. 2, we can see that OCBA-rgm has a slight edge over OCBA-rgmt that is worse than EA. AOAs achieves the best performance among all allocation procedures. The superior performance of AOAs could be attributed to the reason that AOAs makes full use of sample information and is derived in a stochastic control framework with a support for finite sample performance.

We report the effects of \( m \) on different allocation procedures in Experiment 1. Table I reports the IPCS of four allocation procedures when \( T = 1000 \) for \( m = 5, 15, 25, 35, 45 \). From Table I, we can see that all allocation procedures obtain higher IPCS as \( m \) grows, since the size of the selected subset increases. EA performs better than OCBA-rgm when \( m = 5, 15, 25 \). OCBA-rgm outperforms OCBA-rgmt, and AOAs stands out among all allocation procedures. A figure presenting the information in Table I can be found in [27, Appendix].

Three additional synthetic experiments, which highlight the difference between the asymptotic property and finite-sample performance for sequential OCBA-rgm procedure, and an experiment of comparison of AOAs and AOAP, can be found in [27, Appendix]. Numerical results show that the advantage of AOAs appears to be more significant when the number of competing alternatives is large.

Experiment 2—An \((s, S)\)-type Inventory System Problem: We consider an \((s, S)\)-type inventory system problem. The goal is to determine the alternatives that have low expected average cost. The inventory

\(^1\)Online. Available: https://github.com/gongbozhang-pku/Good-Enough-Selection.
We present a sequential sampling procedure for selecting the subset with the best of a finite set of alternatives. Given a simulation budget, the goal is to maximize the posterior PCS for selecting a subset containing the best alternative. We formulate the dynamic sampling decision under a stochastic control framework, and derive an efficient sequential allocation procedure named as AOAs by maximizing a VFA one-step look ahead.

The proposed sequential AOAs procedure is proved to be consistent, and we establish the asymptotic sampling ratios of AOAs, resulting in some useful insights. A set of experiments indicate that our proposed allocation procedure is significantly more efficient than existing sampling procedures, especially when the number of competing alternatives is large, under various sizes of the selected subset. How to apply the proposed allocation procedure to improve the computational efficiency in reinforcement learning [36], [37] deserves future research.

V. CONCLUSION

TABLE II

| Alternative | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-------------|---|---|---|---|---|---|---|---|---|----|
| $s$          | 5 | 5 | 10 | 10 | 10 | 10 | 10 | 10 | 20 | 20 |
| $S$          | 45 | 50 | 45 | 55 | 55 | 60 | 65 | 70 | 40 | 45 |

An emergency department healthcare staffing problem can be found in [27, Appendix].

References

[1] M. C. Fu et al., *Handbook of Simulation Optimization*, Berlin, Germany: Springer, 2015.
[2] C.-H. Chen and L. H. Lee, *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*. Singapore: World Sci., 2011.
[3] L. J. Hong, W. Fan, and J. Luo, “Review on ranking and selection: A new perspective,” vol. 8, pp. 321–343, 2021.
[4] Y. Ranott, “On two-stage selection procedures and related probability-inequalities,” *Commun. Statist.–Theory Methods*, vol. 7, no. 8, pp. 799–811, 1978.
[5] S.-H. Kim and B. L. Nelson, “A fully sequential procedure for indifference-zone selection in simulation,” *ACM Trans. Model. Comput. Simul.*, vol. 11, no. 3, pp. 251–273, 2001.
[6] C.-H. Chen, J. Lin, E. Yücesan, and S. E. Chick, “Simulation budget allocation for further enhancing the efficiency of ordinal optimization,” *J. Discrete Event Dyn. Syst.*, vol. 10, no. 3, pp. 251–270, 2000.
[7] S. E. Chick and K. Inoue, “New two-stage and sequential procedures for selecting the best simulated system,” *Operations Res.*, vol. 49, no. 5, pp. 732–743, 2001.
[8] S. E. Chick, J. Branke, and C. Schmidt, “Sequential sampling to myopically maximize the expected value of information,” *INFORMS J. Comput.*, vol. 22, no. 1, pp. 71–80, 2010.
[9] S. S. Gupta and K. J. Miescke, “Bayesian look ahead one-stage sampling allocations for selection of the best population,” *J. Stat. Plan. Infer.,* vol. 54, no. 2, pp. 229–244, 1996.
[10] P. Frazier, W. Powell, and S. Dayanik, “The knowledge-gradient policy for correlated normal beliefs,” *INFORMS J. Comput.*, vol. 21, no. 4, pp. 599–613, 2009.
[11] D. R. Jones, M. Schonlau, and W. J. Welch, “Efficient global optimization of expensive black-box functions,” *J. Glob. Optim.,* vol. 13, no. 4, pp. 455–492, 1998.
[12] I. O. Ryzhov, “On the convergence rates of expected improvement methods,” *Operations Res.*, vol. 64, no. 6, pp. 1515–1528, 2016.
[13] Y. Peng, E. K. Chong, C.-H. Chen, and M. C. Fu, “Ranking and selection as stochastic control,” *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2359–2373, Aug. 2018.
[14] P. Glynn and S. Junreja, “A large deviations perspective on ordinal optimization,” in *Proc. Winter Simul. Conf.*, 2004, pp. 577–585.
[15] L. W. Koenig and A. M. Law, “A procedure for selecting a subset of size m containing the t best of k independent normal populations, with applications to simulation,” *Commun. Statist. Simul. Comput.*, vol. 14, no. 3, pp. 719–734, 1985.
[16] E. Chen, “Subset selection procedures,” *J. Simul.*, vol. 3, no. 4, pp. 202–210, 2009.
[17] F. Gao and S. Gao, “A new strategy for selecting good enough designs using optimal computing budget allocation,” *J. Simul.*, vol. 12, no. 3, pp. 238–247, 2018.
[18] S. Gupta, “On some multiple decision (selection and ranking) rules,” *Technometrics*, vol. 7, no. 2, pp. 225–245, 1965.
[19] J. Boesel, B. L. Nelson, and S.-H. Kim, “Using ranking and selection to “clean up” after simulation optimization,” *Operations Res.*, vol. 51, no. 5, pp. 814–825, 2003.
[20] D. J. Eckman, M. Plumlee, and B. L. Nelson, “Revisiting subset selection,” in Proc. Winter Simul. Conf., 2020, pp. 2972–2983.

[21] W. Studden, “On selecting a subset of k populations containing the best,” Ann. Math. Statist., vol. 38, no. 4, pp. 1072–1078, 1967.

[22] J. J. Deely and S. S. Gupta, “On the properties of subset selection procedures,” Sankhyā: Indian J. Statist., Ser. A, vol. 30, no. 1, pp. 37–50, 1968.

[23] K.-J. Mieseke, “Bayesian subset selection for additive and linear loss function,” Commun. Statist.-Theory Methods, vol. 8, no. 12, pp. 1205–1226, 1979.

[24] W. Fan, L. J. Hong, and B. L. Nelson, “Indifference-zone-free selection of the best,” Operations Res., vol. 64, no. 6, pp. 1499–1514, 2016.

[25] D. J. Eckman and S. G. Henderson, “Guarantees on the probability of good selection,” in Proc. Winter Simul. Conf., 2018, pp. 351–365.

[26] Y.-C. Ho, C. G. Cassandras, C.-H. Chen, and L. Dai, “Ordinal optimisation and simulation,” J. Oper. Res. Soc., vol. 51, no. 4, pp. 490–500, 2000.

[27] G. Zhang, B. Chen, Q.-S. Jia, and Y. Peng, “Online appendix for ‘efficient sampling policy for selecting a subset with the best,’” 2021, [Online]. Available: https://ssrn.com/abstract=3973542

[28] D. P. Bertsekas, Approximate Dynamic Programming: Solving the Curses of Dimensionality. Hoboken, NJ, USA: Wiley, 2007.

[29] W. B. Powell, Approximate Dynamic Programming: Solving the Curses of Dimensionality. Hoboken, NJ, USA: Wiley, 2007.

[30] C.-H. Chen, “A lower bound for the correct subset-selection probability and its application to discrete-event system simulations,” IEEE Trans. Autom. Control, vol. 41, no. 8, pp. 1227–1231, Aug. 1996.

[31] M. H. DeGroot, Optimal Statistical Decisions. Hoboken, NJ, USA: Wiley, 2005.

[32] D. P. Bertsekas, “Dynamic programming and suboptimal control: A survey from ADP to MPC,” Eur. J. Control, vol. 11, no. 4-5, pp. 310–334, 2005.

[33] W. Rudin et al., Principles of Mathematical Analysis. New York, NY, USA: McGraw-Hill, 1964.

[34] S. Boyd, S. P. Boyd, and L. Vandenberghe, Convex Optimization, 7th ed., Cambridge, U.K.: Cambridge Univ. Press, 2009.

[35] A. M. Law, W. D. Kelton, and W. D. Kelton, Simulation Modeling and Analysis, New York, NY, USA: McGraw-Hill, 2000.

[36] Q.-S. Jia and J. Wu, “A structural property of charging scheduling policy for shared electric vehicles with wind power generation,” IEEE Trans. Control Syst. Technol., vol. 29, no. 6, pp. 2393–2405, Nov. 2021.

[37] Z. Jiang, Q.-S. Jia, and X. Guan, “A computing budget allocation method for minimizing EV charging cost using uncertain wind power,” IEEE Trans. Autom. Sci. Eng., vol. 18, no. 2, pp. 681–692, Apr. 2021.

[38] G. Zhang, C.-H. Chen, Q.-S. Jia, and Y. Peng, “Dynamic sampling allocation for selecting a good enough alternative,” in Proc. IEEE 16th Int. Conf. Automat. Sci. Eng., 2020, pp. 1319–1324.