Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

DYNAMIC SAMPLING POLICY FOR SUBSET SELECTION

Gongbo Zhang

Guanghua School of Management
Peking University
5 Yiheyuan Road
Beijing 100871, P. R. CHINA

Yijie Peng

Guanghua School of Management
Peking University
5 Yiheyuan Road
Beijing 100871, P. R. CHINA

Jianghua Zhang

School of Management
Shandong University
27 South Shanda Road
Jinan 250100, P. R. CHINA

Enlu Zhou

Industrial and Systems Engineering
Georgia Institute of Technology
755 Ferst Drive NW
Atlanta, GA 30332, USA

ABSTRACT

We consider a problem of selecting a subset of a finite number of competing alternatives via Monte Carlo simulation, where the subset contains the top-m alternatives. Under the Bayesian framework, we develop a dynamic sampling policy to efficiently learn and select the top-m alternatives. The proposed sampling policy is proved to be consistent, i.e., the selected alternatives will be the true top-m alternatives as the number of simulation budget goes to infinity. Numerical results show that the proposed sampling policy outperforms the existing ones.

1 INTRODUCTION

The ranking and selection (R&S) problem has been actively studied in simulation and optimization (see Bechhofer 1995; Chen and Lee 2011; Powell and Ryzhov 2012), which aims to find the single best alternative from a finite set of k alternatives with unknown means. The best alternative is defined as the one with the largest mean, i.e., $\langle 1\rangle \overset{\Delta}{=} \arg\max_{i=1,2,\cdots,k}\mu_i$, where μ_i, $i=1,2,\cdots,k$ is the unknown mean of each alternative that can only be estimated by Monte Carlo simulation. A more general problem than R&S is to select alternatives with top-m largest means from a finite set of k alternatives $(1 \le m < k, m, k \in \mathbb{N}^+)$. We aim to efficiently allocate simulation budget to alternatives so that the probability of correct selection (PCS) of top-m alternatives can be maximized. Instead of focusing on the best alternative, the subset selection problem provides more choices for decision makers, and more choices sometimes can lead to better outcome. For example, in an emergency evacuation problem, different evacuation routes can be regarded as alternatives, and in response to stampedes and disasters, evacuating evacuees by several well-chosen routes could reduce the evacuation time compared with evacuating evacuees via a single best route (Zhang et al. 2018).

There are two different ways to categorize R&S procedures. In terms of methodologies, there are frequentist and Bayesian branches. In terms of goals, there are fixed-precision and fixed-budget procedures. See Kim and Nelson (2006), Chick (2006) and Hong et al. (2020) for overviews. Fixed-precision procedures (e.g., Rinott 1978; Kim and Nelson 2001; Frazier 2014) allocate simulation replications to guarantee a

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

pre-specified PCS level, whereas fixed-budget procedures (e.g., Chen et al. 2000; Frazier et al. 2009; Chick et al. 2010) allocate simulation replications to maximize the PCS or minimize the expected opportunity cost (EOC) under a given simulation budget. A representative Bayesian procedure is optimal computing budget allocation (OCBA) (Chen et al. 2000; Chen and Lee 2011), which is derived in a static optimization framework. The OCBA is originally a two-stage method and then is implemented in a sequential manner by combining with some heuristic rules. It has been widely observed that the sequential sampling procedures that incrementally allocate simulation replications perform better than the two-stage procedures (Chen and Lee 2011). Peng et al. (2018) propose a stochastic control framework to formulate the sequential sampling policy under the Bayesian framework, and develop an efficient sampling procedure that maximizes a value function approximation (VFA) one-step look ahead.

We briefly review the existing literature for selecting top-m alternatives. Koenig and Law (1985) adopt an indifference zone paradigm in the frequentist branch and propose a two-stage procedure. Chen et al. (2008) generalize the OCBA sampling procedure in Chen et al. (2000) to subset selection problem. Gao and Chen (2015a) and Gao and Chen (2015b) propose sequential allocation procedures for selecting top-m alternatives to asymptotically minimize approximation of EOC and maximize approximation of PCS, respectively. Zhang et al. (2015) improve the sampling procedure in Chen et al. (2008) with a better approximation of PCS. Gao and Chen (2016) further extend the results in Gao and Chen (2015b) to general sampling distributions by using a large deviations principle and Bonferroni inequality to approximate PCS. We caution that the approximation of PCS derived by Bonferroni inequality could be loose when the simulation budget is small. Particularly, it could even lead to a negative lower bound for PCS which is a probability. The sequential sampling procedures in Gao and Chen (2015b) and Gao and Chen (2016) are developed to achieve certain asymptotic optimality conditions, however, recent studies show that the asymptotic property is inadequate to capture the finite-sample behavior of sequential sampling policies (see Peng et al. 2018; Wu and Zhou 2018; Zhang et al. 2020; Shin et al. 2021; Shi et al. 2021).

In our work, we formulate the sequential sampling decision for selecting top-m alternatives as a stochastic dynamic programming problem under the Bayesian framework. We develop a sequential sampling policy that maximizes a value function approximation one-step look ahead. The proposed sampling policy is proved to be consistent and is demonstrated to perform well empirically.

The rest of the paper is organised as follows. Section 2 formulates the proposed problem. The value function approximation and dynamic allocation scheme are proposed in Section 3. Section 4 provides numerical results. The last section concludes the paper.

2 PROBLEM FORMULATION

Suppose there are k alternatives and the performance of each alternative is measured by unknown mean μ_i, $i=1,2,\cdots,k$. We assume that the performance of each alternative can be distinguished, i.e., $\mu_i \neq \mu_j$, $i,j=1,2,\cdots,k$, $i\neq j$. The set of true top-m alternatives is denoted as $\mathscr{T}^m \stackrel{\Delta}{=} \{\langle 1\rangle, \langle 2\rangle, \cdots, \langle m\rangle\}$, where $\langle i\rangle$, $i=1,2,\cdots,k$ are indices ranked by system performances such that $\mu_{\langle 1\rangle} > \mu_{\langle 2\rangle} > \cdots > \mu_{\langle m\rangle} > \mu_{\langle m+1\rangle} > \cdots > \mu_{\langle k\rangle}$. μ_i, $i=1,2,\cdots,k$ can be estimated by sampling independent and identically (i.i.d.) simulation replications $X_{i,t}$, $t\leq T$, $t\in\mathbb{Z}^+$, $i=1,2,\cdots,k$, where $T<\infty$ is the number of total simulation replications.

Suppose $(X_{1,t}, X_{2,t}, \cdots, X_{k,t})$ follows a joint distribution $Q(\cdot;\theta)$ with density function (or probability mass function) $q(\cdot;\theta)$, where θ contains all unknown parameters in the sampling distribution, including the unknown means, i.e., $\mu\in\theta$. To quantify the uncertainty of parameter θ, under the Bayesian framework, we suppose that θ follows a prior distribution $F(\cdot;\xi_0)$ that reflects the prior knowledge on the unknown parameters, where ξ_0 contains all hyper-parameters. The sampling policy is a sequence of mappings $\mathscr{A}_T(\cdot) \stackrel{\Delta}{=} (A_1(\cdot), A_2(\cdot), \cdots A_T(\cdot))$ with $A_t(\cdot)\in\{1,2,\cdots,k\}$ that sequentially allocates each replication to an alternative based on collected information $\mathscr{E}_{t-1}$ throughout $(t-1)$ steps. Let $\mathscr{E}_t \stackrel{\Delta}{=} \{\zeta_0, X_1^{(t)}, X_2^{(t)}, \cdots X_k^{(t)}\}$ be a collected information set throughout t steps, and $X_i^{(t)} \stackrel{\Delta}{=} (X_{i,1}, X_{i,2}, \cdots X_{i,t_i})$, $i=1,2,\cdots,k$ be the sample observations of alternative i throughout t steps, where $t_i \overset{\Delta}{=} \sum_{\ell=1}^{t} A_{i,\ell}(\mathscr{E}_{\ell-1})$ with $A_{i,\ell}(\mathscr{E}_{\ell-1}) \overset{\Delta}{=} \mathbb{1}(A_\ell(\mathscr{E}_{\ell-1}) = i)$, $i = 1, 2, \cdots, k$ be the number of simulation replications allocated to alternative i throughout t steps, and $\mathbb{1}(\cdot)$ is an indicator function that equals to 1 when the event in the bracket is true and equals to 0 otherwise. The set of estimated top-m alternatives after allocating t simulation replications is denoted as $\widehat{\mathscr{F}}_t^m \overset{\Delta}{=} \{\langle 1\rangle_t, \langle 2\rangle_t, \cdots, \langle m\rangle_t\}$, where $\langle i\rangle_t$, $i = 1, 2, \cdots, k$ are indices ranked by posterior means such that $\mu_{\langle 1\rangle_t} > \mu_{\langle 2\rangle_t} > \cdots > \mu_{\langle m\rangle_t} > \mu_{\langle m+1\rangle_t} > \cdots > \mu_{\langle k\rangle_t}$. If the allocation procedure stops at t-th step, a correct selection occurs when $\widehat{\mathscr{F}}_t^m = \mathscr{F}^m$. The posterior PCS for selecting top-m alternatives can be expressed as

$$\mathrm{PCS}_t = \Pr\{\widehat{\mathscr{F}}_t^m = \mathscr{F}^m | \mathscr{E}_t\} = \Pr\{\cap_{i=1}^m \cap_{j=m+1}^k \left(\mu_{\langle i\rangle_t} > \mu_{\langle j\rangle_t}\right) | \mathscr{E}_t\} \ . \tag{1}$$

We aim to find an efficient sequential sampling policy such that the posterior PCS for selecting the top-m alternatives can be maximized. In this work, we propose a sequential sampling policy $\mathscr{A}_t(\cdot)$ to maximize the posterior PCS (1) looking one-step ahead: $\max_{\mathscr{A}_t(\cdot)} \Pr\left\{ \mu_{\langle i\rangle_t} > \mu_{\langle j\rangle_t} \middle| \mathscr{E}_{t-1} \right\}$, $i = 1, 2, \cdots, m$, $j = m+1, m+2, \cdots, k$. The sequential sampling policy is captured by a stochastic dynamic programming problem. The expected payoff for a sequential sampling policy $\mathscr{A}_T(\cdot)$ can be defined recursively by

$$V_T(\mathscr{E}_T; \mathscr{A}_T(\cdot)) \overset{\Delta}{=} \Pr\left\{ \mu_{\langle i\rangle_T} > \mu_{\langle j\rangle_T},\ i = 1, 2, \cdots, m,\ j = m+1, m+2, \cdots, k \middle| \mathscr{E}_T \right\} \ , \tag{2}$$

which is the posterior integrated PCS (IPCS), and for $0 \le t < T$,

$$V_t(\mathscr{E}_t; \mathscr{A}_T(\cdot)) \overset{\Delta}{=} \mathbb{E}\left[V_{t+1}\left(\mathscr{E}_t \cup \left\{X_{i,t_i+1}\right\}; \mathscr{A}_T(\cdot)\right) \middle| \mathscr{E}_t \right] \Big|_{i=A_{t+1}(\mathscr{E}_t)} \ ,$$

where X_{i,t_i+1} is the (t_i+1)-th simulation replication for the allocated alternative i, $i = A_{t+1}(\mathscr{E}_t)$. Then the optimal sequential sampling policy is well defined by $\mathscr{A}_T^*(\cdot) \overset{\Delta}{=} \arg\max_{\mathscr{A}_T(\cdot)} V_0(\zeta_0; \mathscr{A}_T(\cdot))$.

The stochastic dynamic programming problem is a Markov decision process (MDP) with $(T+1)$ stages, where ζ_0 is the state at stage 0 and the information set $\mathscr{E}_t$ is the state at stage t, $0 < t \le T$. The action is A_{t+1} and transition $\mathscr{E}_t \to \mathscr{E}_{t+1} \overset{\Delta}{=} \{\mathscr{E}_t \cup \{X_{i,t_i+1}\}\}|_{i=A_{t+1}}$, $0 \le t < T$. The only nonzero reward is the final reward $V_T(\mathscr{E}_T)$. Note that the dimension of state space grows as the step grows. We have the Bellman equation: $V_T(\mathscr{E}_T) \overset{\Delta}{=} \Pr\left\{ \mu_{\langle i\rangle_T} > \mu_{\langle j\rangle_T},\ i = 1, 2, \cdots, m,\ j = m+1, m+2, \cdots, k \middle| \mathscr{E}_T \right\}$, and for $0 \le t < T$, $V_t(\mathscr{E}_t) \overset{\Delta}{=} \mathbb{E}\left[V_{t+1}(\mathscr{E}_t \cup \{X_{i,t_i+1}\}) | \mathscr{E}_t\right]|_{i=A_{t+1}^*(\mathscr{E}_t)}$, where $A_{t+1}^*(\mathscr{E}_t) = \arg\max_{i=1,2,\cdots,k} \mathbb{E}\left[V_{t+1}(\mathscr{E}_t \cup \{X_{i,t_i+1}\}) | \mathscr{E}_t\right]$. The backward induction can be used to solve the stochastic dynamic programming problem, however, it suffers from curse-of-dimensionality when solving R&S problem (Peng et al. 2018). To derive a dynamic sampling policy with an analytical form, we adopt an approximate dynamic programming (ADP) paradigm which makes dynamic decision based on a VFA and keeps learning the VFA with decisions moving forward.

A common assumption is that simulation replications of different alternatives are independent, i.e., $Q(\cdot;\theta) = \prod_{i=1}^k Q_i(\cdot;\theta_i)$, where θ_i comprises all unknown parameters in the marginal distribution Q_i, $i = 1, 2, \cdots, k$. We assume that for each alternative, the simulation replications are i.i.d. normally distributed, i.e., $X_{i,t} \sim N\left(\mu_i, \sigma_i^2\right)$, $i = 1, 2, \cdots, k$. The variance σ_i^2 in the sampling distribution is assumed to be known in this work and sample variance is used as a plug-in for the true variance in practice. The posterior and predictive distribution can be calculated by Bayes rule. If the prior distribution is a conjugate prior of the sampling distribution, the posterior distribution $F(\cdot;\xi_t)$ lies in the parametric family of the prior distribution, and then $\mathscr{E}_t$ can be completely summarized by posterior hyper-parameters ξ_t. Suppose the prior distribution of μ_i is $N(\mu_i^{(0)}, (\sigma_i^{(0)})^2)$, which is the conjugate prior for the normal distribution with unknown mean and known variance. The posterior distribution of μ_i is $N(\mu_i^{(t)}, (\sigma_i^{(t)})^2)$ with posterior variance $(\sigma_i^{(t)})^2 = 1 \Big/ \left(\frac{1}{(\sigma_i^{(0)})^2} + \frac{t_i}{\sigma_i^2} \right)$ and posterior mean $\mu_i^{(t)} = (\sigma_i^{(t)})^2 \left(\frac{\mu_i^{(0)}}{(\sigma_i^{(0)})^2} + \frac{t_i m_i^{(t)}}{\sigma_i^2} \right)$, where $m_i^{(t)} = \sum_{\ell=1}^{t_i} X_{i,\ell} / t_i$ is the sample mean, and the predictive distribution of X_{i,t_i+1} is $N(\mu_i^{(t)}, \sigma_i^2 + (\sigma_i^{(t)})^2)$, $i = 1,2,\cdots,k$. Note that if $\sigma_i^{(0)} \to \infty$, then $\mu_i^{(t)} = m_i^{(t)}$ and $(\sigma_i^{(t)})^2 = \sigma_i^2 / t_i$, $i = 1,2,\cdots,k$, which only includes sample information without any hyper-parameters, and such a case is called uninformative. Then we have the information set $\mathscr{E}_t = \{\mu_1^{(t)}, \mu_2^{(t)}, \cdots, \mu_k^{(t)}, (\sigma_1^{(t)})^2, (\sigma_2^{(t)})^2, \cdots, (\sigma_k^{(t)})^2, \mathscr{E}_0\}$. For a normal distribution with unknown variance, there is a normal-gamma conjugate prior (DeGroot 2005), and the corresponding analysis is left for future research.

3 DYNAMIC SAMPLING POLICY

Conditioned on $\mathscr{E}_t$, μ_i follows a normal distribution with mean $\mu_i^{(t)}$ and variance $(\sigma_i^{(t)})^2$, $i = 1,2,\cdots,k$. Then, the joint distribution of vector $(\mu_{\langle 1\rangle_t} - \mu_{\langle m+1\rangle_t}, \cdots \mu_{\langle 1\rangle_t} - \mu_{\langle k\rangle_t}, \cdots, \mu_{\langle m\rangle_t} - \mu_{\langle m+1\rangle_t}, \cdots \mu_{\langle m\rangle_t} - \mu_{\langle k\rangle_t})$ follows a joint normal distribution with mean vector $(\mu^{(t)}_{\langle 1\rangle_t} - \mu^{(t)}_{\langle m+1\rangle_t}, \cdots \mu^{(t)}_{\langle 1\rangle_t} - \mu^{(t)}_{\langle k\rangle_t}, \cdots, \mu^{(t)}_{\langle m\rangle_t} - \mu^{(t)}_{\langle m+1\rangle_t}, \cdots \mu^{(t)}_{\langle m\rangle_t} - \mu^{(t)}_{\langle k\rangle_t})$, and covariance matrix $\Sigma = \Gamma'\Lambda\Gamma$, where $'$ denotes the transpose operation of the matrix, $\Lambda \in \mathbb{R}^{k\times k}$, $\Lambda \triangleq diag((\sigma^{(t)}_{\langle 1\rangle_t})^2, (\sigma^{(t)}_{\langle 2\rangle_t})^2, \cdots, (\sigma^{(t)}_{\langle m\rangle_t})^2, (\sigma^{(t)}_{\langle m+1\rangle_t})^2, (\sigma^{(t)}_{\langle m+2\rangle_t})^2, \cdots, (\sigma^{(t)}_{\langle k\rangle_t})^2)$ is a diagonal matrix, and $\Gamma \in \mathbb{R}^{k\times[m\times(k-m)]}$ is a block matrix,

$$\Gamma \triangleq \begin{pmatrix} R_1 & R_2 & \cdots & R_m \\ S_1 & S_2 & \cdots & S_m \end{pmatrix}_{k\times[m\times(k-m)]},$$

where $R_i \in \mathbb{R}^{m\times(k-m)}$, $i = 1,2,\cdots,m$ is a matrix in which i-th row is 1 and the remaining rows are 0, and $S_i \in \mathbb{R}^{(k-m)\times(k-m)}$, $i = 1,2,\cdots,m$ is a diagonal matrix, i.e.,

$$R_i \triangleq \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \to i\text{th row}, \qquad S_i \triangleq \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix}_{(k-m)\times(k-m)}.$$

With Cholesky decomposition of the covariance matrix $\Sigma = \Gamma'\Lambda\Gamma = U'U$, where $U \in \mathbb{R}^{[m\times(k-m)]\times[m\times(k-m)]}$ is an upper triangular matrix with elements u_{ij}, i.e., $u_{ij} = 0$ if $i > j$, $i,j = 1,2,\cdots,m\times(k-m)$, the value function (2) can be rewritten as $\frac{1}{\left(\sqrt{2\pi}\right)^{m\times(k-m)}} \times \int_{\sum_{\ell=1}^{q} u_{\ell,q} z_\ell \geq \mu^{(t)}_{\langle j\rangle_t} - \mu^{(t)}_{\langle i\rangle_t}} \prod_{h=1}^{m\times(k-m)} \exp\left(-\frac{z_h^2}{2}\right) dz_1 \cdots dz_{m\times(k-m)}$. A VFA one-step look ahead is an integral of the multivariate standard normal density over an area covered by $[m\times(k-m)]$-dimensional hyper-planes: $\sum_{\ell=1}^{q} u_{\ell,q} z_\ell = \mu^{(t)}_{\langle j\rangle_t} - \mu^{(t)}_{\langle i\rangle_t}$, $i = 1,2,\cdots,m$, $j = m+1, m+2,\cdots,k$, $q = 1,2,\cdots,m\times(k-m)$, and in the case with $k = 3$ and $m = 2$, the area is the shaded area in Figure 1. Since the standard normal density decays at an exponential pace with respect to the distance from the origin, the VFA can be further simplified as the integral over the area of the largest internal ball encompassed by the hyper-planes, i.e., the circle in Figure 1 for the case of $k = 3$ and $m = 2$. It can be shown that the error of the integral of the $[m\times(k-m)]$-dimensional standard normal density over the largest internal ball as an approximation of $\Pr\left\{ \mu_{\langle i\rangle_t} > \mu_{\langle j\rangle_t},\ i = 1,2,\cdots,m,\ j = m+1, m+2,\cdots,k \middle| \mathscr{E}_t \right\}$ decreases to 0 in an exponential rate as t goes to infinity. Owing to symmetry of normal density, maximizing the integral over the largest internal ball is equivalent to maximizing the size of the circle, which yields an analytical form for the solution to the stochastic dynamic programming. Let $d(\mathscr{E}_t) \triangleq \min_{i,j} d_{ij}(\mathscr{E}_t)$, $i = 1,2,\cdots,m$, $j = m+1, m+2,\cdots,k$ be the radius of the largest internal ball of the integration region, where $d_{ij}(\mathscr{E}_t) = \left(\mu^{(t)}_{\langle i\rangle_t} - \mu^{(t)}_{\langle j\rangle_t}\right) \Big/ \sqrt{\left(\sigma^{(t)}_{\langle i\rangle_t}\right)^2 + \left(\sigma^{(t)}_{\langle j\rangle_t}\right)^2}$, $i \in \{1,2,\cdots,m\}$, $j \in \{m+1, m+2,\cdots,k\}$ is the distance from the origin to each hyper-plane, and then the VFA is given by $\widetilde{V}_t(\mathscr{E}_t) \stackrel{\Delta}{=} d^2(\mathscr{E}_t)$. At any step t, we treat the $(t+1)$-th step as the last step, and then an approximation of the value function looking one step ahead at step t by allocating a replication to the alternative i, $i=1,2,\cdots,k$ can be given as

$$\widetilde{V}_t(\mathscr{E}_t;i) \stackrel{\Delta}{=} \mathbb{E}[\widetilde{V}_{t+1}(\mathscr{E}_t \cup \{X_{i,t_i+1}\})|\mathscr{E}_t] \ . \tag{3}$$

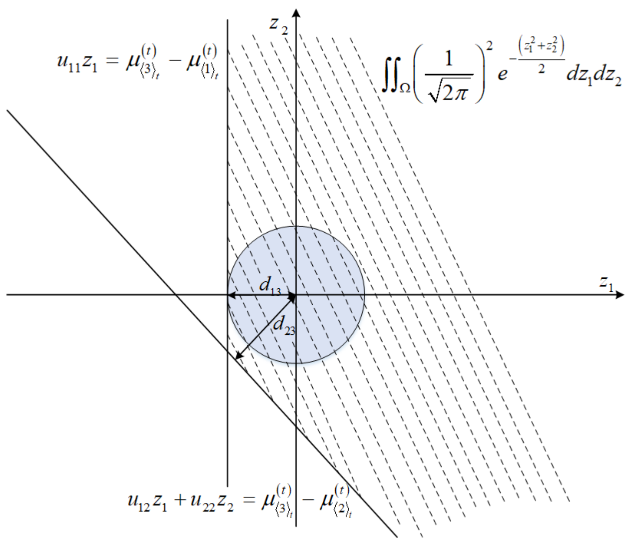


Figure 1: Area of integration for approximation is the ball, where dominant values of integrand $\exp\left(-\left(z_1^2+z_2^2\right)/2\right)$ are captured.

Further, since the expectation is computationally expensive to calculate, the certainty equivalence (Bertsekas 2005) is used as an approximation of (3), that is, replacing stochastic quantities by their expected values, which can be expressed as $\mathbb{E}[\widetilde{V}_{t+1}(\mathscr{E}_t \cup \{X_{i,t_i+1}\})|\mathscr{E}_t] = \widetilde{V}_{t+1}(\mathscr{E}_t \cup \mathbb{E}[X_{i,t_i+1}|\mathscr{E}_t])$. Then the VFA looking one-step ahead can be calculated, i.e., for $i,h=1,2,\cdots,m$ and $j,\ell=m+1,m+2,\cdots,k$,

$$\widehat{V}_t(\mathscr{E}_t;i) \stackrel{\Delta}{=} \widetilde{V}_{t+1}(\mathscr{E}_t \cup \mathbb{E}[X_{i,t_i+1}|\mathscr{E}_t]) = \min\left\{ \min_j \frac{(\mu^{(t)}_{\langle i\rangle_t} - \mu^{(t)}_{\langle j\rangle_t})^2}{(\sigma^{(t+1)}_{\langle i\rangle_t})^2 + (\sigma^{(t)}_{\langle j\rangle_t})^2}, \min_{\substack{h\neq i\\ j}} \frac{(\mu^{(t)}_{\langle h\rangle_t} - \mu^{(t)}_{\langle j\rangle_t})^2}{(\sigma^{(t)}_{\langle h\rangle_t})^2 + (\sigma^{(t)}_{\langle j\rangle_t})^2} \right\} \ ,$$

and

$$\widehat{V}_t(\mathscr{E}_t;j) \stackrel{\Delta}{=} \widetilde{V}_{t+1}\left(\mathscr{E}_t \cup \mathbb{E}\left[X_{j,t_j+1}|\mathscr{E}_t\right]\right) = \min\left\{ \min_i \frac{(\mu^{(t)}_{\langle i\rangle_t} - \mu^{(t)}_{\langle j\rangle_t})^2}{(\sigma^{(t)}_{\langle i\rangle_t})^2 + (\sigma^{(t+1)}_{\langle j\rangle_t})^2}, \min_{\substack{\ell\neq j\\ i}} \frac{(\mu^{(t)}_{\langle i\rangle_t} - \mu^{(t)}_{\langle \ell\rangle_t})^2}{(\sigma^{(t)}_{\langle i\rangle_t})^2 + (\sigma^{(t)}_{\langle \ell\rangle_t})^2} \right\} \ ,$$

where $(\sigma^{(t+1)}_{\langle \ell\rangle_t})^2 = 1/(1/(\sigma^{(0)}_{\langle \ell\rangle})^2 + (t_\ell+1)/\sigma^2_{\langle \ell\rangle})$, $\ell=1,2,\cdots,k$.

A dynamic sampling policy for selecting top-m alternatives (DSSm) that optimizes the VFA looking one-step ahead is given by

$$\widehat{A}_{t+1}(\mathscr{E}_t) = \arg\max_{i=1,2,\cdots,k} \widehat{V}_t(\mathscr{E}_t;i) \ . \tag{4}$$

The proposed sequential DSSm (4) uses the information on the posterior means and posterior variances of the estimated unknown performances of alternatives. Note that $(\mu^{(t)}_{\langle i\rangle_t} - \mu^{(t)}_{\langle j\rangle_t})^2$ and $(\sigma^{(t)}_{\langle i\rangle_t})^2 + (\sigma^{(t)}_{\langle j\rangle_t})^2$ are the squared mean and variance of the posterior distribution of the difference in performances of alternatives i and

j, respectively, $i=1,2,\cdots,m$, $j=m+1,m+2,\cdots,k$, and then the VFA can be rewritten as $\min\limits_{i,j} 1\big/c_v^2(i,j)$, where $c_v(i,j)$ is the posterior noise-signal ratio (coefficient of variation) of $\mu_{\langle i\rangle_t}-\mu_{\langle j\rangle_t}$, $i=1,2,\cdots,m$, $j=m+1,m+2,\cdots,k$. DSSm (4) minimizes the maximum of $c_v(i,j)$, which is reasonable since the larger is the value of $c_v(i,j)$, the higher the difficulty in comparing $\mu_{\langle i\rangle_t}$ and $\mu_{\langle j\rangle_t}$ from posterior information. At each step, DSSm (4) sequentially allocates a replication to an alternative to reduce the noise-signal ratio, focusing on the alternatives that are most difficult in comparison based on posterior information. DSSm (4) is proved to be consistent in the following theorem, i.e., as $t\to\infty$, the top-m estimated set $\mathscr{F}_m^{(t)}$ will be the true top-m set $\mathscr{F}_m$.

Theorem 1 The proposed sequential DSSm (4) is strongly consistent, i.e., $\lim\limits_{t\to\infty}\widehat{\mathscr{F}}_m^{(t)}=\mathscr{F}_m,\ a.s.$.

Proof. We only need to prove that every alternative will be sampled infinitely often almost surely, i.e., $\lim_{t\to\infty}t_i=\infty$ for each $i=1,2,\cdots,k$. Following DSSm (4), the consistency will follow by the law of large numbers (LLN).

Define $\mathscr{I}\stackrel{\Delta}{=}\{i\in\{1,2,\ldots,k\}:$ alternative $\langle i\rangle$ is sampled infinitely often a.s.$\}$. If $\{1,2,\cdots,m\}\cap\mathscr{I}=\emptyset$ and $\{m+1,m+2,\ldots,k\}\cap\mathscr{I}\neq\emptyset$, then $\forall\ i\in\{1,2,\cdots,m\}$, $\lim\limits_{t\to\infty}(\sigma_{\langle i\rangle}^{(t)})^2>0$, $\lim\limits_{t\to\infty}[(\sigma_{\langle i\rangle}^{(t)})^2-(\sigma_{\langle i\rangle}^{(t+1)})^2]>0$. $\exists\ h\in\{1,2,\cdots,m\}$ such that $\lim\limits_{t\to\infty}[\widehat{V}_t(\mathscr{E}_t;h)-\widetilde{V}_t(\mathscr{E}_t)]>0,\ a.s.$ and $\exists\ j\in\{m+1,m+2,\cdots,k\}$ such that $\lim\limits_{t\to\infty}(\sigma_{\langle j\rangle}^{(t)})^2=0$, $\lim\limits_{t\to\infty}[(\sigma_{\langle j\rangle}^{(t)})^2-(\sigma_{\langle j\rangle}^{(t+1)})^2]=0$, and $\lim\limits_{t\to\infty}[\widehat{V}_t(\mathscr{E}_t;j)-\widetilde{V}_t(\mathscr{E}_t)]=0,\ a.s.$, which leads to an contradiction to the sampling rule (4). Therefore, $\{1,2,\cdots,m\}\cap\mathscr{I}\neq\emptyset$.

If $\{m+1,m+2,\ldots,k\}\cap\mathscr{I}=\emptyset$, then $\forall\ j\in\{m+1,\cdots,k\}$, $\lim\limits_{t\to\infty}(\sigma_{\langle j\rangle}^{(t)})^2>0$, $\lim\limits_{t\to\infty}[(\sigma_{\langle j\rangle}^{(t)})^2-(\sigma_{\langle j\rangle}^{(t+1)})^2]>0$. $\exists\ j\in\{m+1,m+2,\ldots,k\}$ such that $\lim\limits_{t\to\infty}[\widehat{V}_t(\mathscr{E}_t;j)-\widetilde{V}_t(\mathscr{E}_t)]>0,\ a.s.$, and for $i\in\mathscr{I}\cap\{1,2,\cdots,m\}$, $\lim\limits_{t\to\infty}[\widehat{V}_t(\mathscr{E}_t;i)-\widetilde{V}_t(\mathscr{E}_t)]=0,\ a.s.$, which leads to an contradiction to the sampling rule (4). Therefore, $\{m+1,m+2,\ldots,k\}\cap\mathscr{I}\neq\emptyset$.

In addition, if $\{m+1,m+2,\ldots,k\}\setminus\mathscr{I}\neq\emptyset$, where $\setminus$ denotes the set minus, then for $i\in\{1,2,\cdots,m\}\cap\mathscr{I}$, $j\in\{m+1,m+2,\ldots,k\}\cap\mathscr{I}$, and $\ell\in\{m+1,m+2,\ldots,k\}\setminus\mathscr{I}$, $\lim\limits_{t\to\infty}d_{ij}(\mathscr{E}_t)=\infty$ and $\lim\limits_{t\to\infty}d_{i\ell}(\mathscr{E}_t)<\infty$. Thus, for $j\in\{m+1,m+2,\ldots,k\}\cap\mathscr{I}$, $\lim\limits_{t\to\infty}[\widehat{V}_t(\mathscr{E}_t;j)-\widetilde{V}_t(\mathscr{E}_t)]=0,\ a.s.$, and $\exists\ q\in\{m+1,m+2,\ldots,k\}\setminus\mathscr{I}$ such that $\lim\limits_{t\to\infty}[\widehat{V}_t(\mathscr{E}_t;q)-\widetilde{V}_t(\mathscr{E}_t)]>0,\ a.s.$, which leads to an contradiction to the sampling rule (4). Therefore, $\{m+1,m+2,\ldots,k\}\subset\mathscr{I}$.

Finally, if $\{1,2,\ldots,m\}\setminus\mathscr{I}\neq\emptyset$, then for $j\in\{m+1,\ldots,k\}$, $i\in\{1,2,\cdots,m\}\cap\mathscr{I}$, $h\in\{1,2,\ldots,m\}\setminus\mathscr{I}$, $\lim\limits_{t\to\infty}d_{ij}(\mathscr{E}_t)=\infty$, and $\lim\limits_{t\to\infty}d_{hj}(\mathscr{E}_t)<\infty$.

Thus, for $i\in\{1,2,\ldots,m\}\cap\mathscr{I}$, $\lim\limits_{t\to\infty}[\widehat{V}_t(\mathscr{E}_t;i)-\widetilde{V}_t(\mathscr{E}_t)]=0,\ a.s.$, and $\exists\ h\in\{1,2,\ldots,m\}\setminus\mathscr{I}$ such that $\lim\limits_{t\to\infty}[\widehat{V}_t(\mathscr{E}_t;h)-\widetilde{V}_t(\mathscr{E}_t)]>0,\ a.s.$, which leads to an contradiction to the sampling rule (4). Therefore, $\{1,2,\ldots,m\}\subset\mathscr{I}$. Summarizing the above, the theorem is proved. $\square$

Some existing works (see Chen et al. (2008), Zhang et al. (2015), Gao and Chen (2015b), Gao and Chen (2016)) formulate the sampling decision for selecting top-m alternatives as the solution to a static optimization problem:

$$\begin{aligned}\max_{r_1,r_2,\cdots,r_k}\quad & \Pr\left\{\bar{X}_{\langle i\rangle}(r_iT)>\bar{X}_{\langle j\rangle}(r_jT),\ i=1,2,\cdots,m,\ j=m+1,m+2,\cdots,k\,\middle|\,\theta\right\} \\ \text{s.t.}\quad & \textstyle\sum_{i=1}^k r_i=1\ ,\end{aligned}\tag{5}$$

where $r_i\stackrel{\Delta}{=}t_i\big/T$, $\sum_{i=1}^k r_i=1$ is the sampling ratio, r_iT, $i=1,2,\cdots,k$ is the number of simulation replications allocated to alternative i, and for $i=1,2,\cdots,k$, $\bar{X}_{\langle i\rangle}(r_iT)\stackrel{\Delta}{=}\sum_{\ell=1}^{r_iT}X_{\langle i\rangle,\ell}\Big/(r_iT)$, where a technicality that the number of simulation replications needs to be integer is ignored. Such a static optimization problem is difficult to solve, to reduce the complexity of solving (5), based on the large deviations theory for classical R&S problem (Glynn and Juneja 2004), a surrogate problem is to maximize the large deviations rate of the PFS:

$$\begin{aligned} \max_{r_1,r_2,\cdots,r_k} \quad & \lim_{T\to\infty} -\frac{1}{T}\log\left(1-\Pr\left\{\bar{X}_{\langle i\rangle}(r_iT) > \bar{X}_{\langle j\rangle}(r_jT),\ i=1,2,\cdots,m,\ j=m+1,m+2,\cdots,k \,\middle|\, \theta\right\}\right) \\ \text{s.t.} \quad & \textstyle\sum_{i=1}^{k} r_i = 1 \ , \end{aligned} \tag{6}$$

In the case of normal sampling distributions, the objective function in optimization problem (6) is equivalent to $\max\limits_{r_1,r_2,\cdots,r_k} \min\limits_{\substack{i=1,2,\cdots,m \\ j=m+1,m+2,\cdots,k}} G_{ij}(r_i,r_j;\theta)$, where $G_{ij}(r_i,r_j;\theta) = \frac{\left(\mu_{\langle i\rangle}-\mu_{\langle j\rangle}\right)^2}{2\left(\sigma^2_{\langle i\rangle}/r_i+\sigma^2_{\langle j\rangle}/r_j\right)}$, $i=1,2,\cdots,m,\ j=m+1,m+2,\cdots,k$. Solve the optimization problem (6) with normal underlying distributions, we have the optimal convergence rate of the large deviations of the PFS for selecting top-m alternatives satisfies, for $i,h=1,2,\cdots,m,\ j,\ell=m+1,m+2,\cdots,k$,

$$\min_{j=m+1,m+2,\cdots,k} \frac{\left(\mu_{\langle h\rangle}-\mu_{\langle j\rangle}\right)^2}{\sigma^2_{\langle h\rangle}\Big/r^*_h+\sigma^2_{\langle j\rangle}\Big/r^*_j} = \min_{i=1,2,\cdots,m} \frac{\left(\mu_{\langle i\rangle}-\mu_{\langle \ell\rangle}\right)^2}{\sigma^2_{\langle i\rangle}\Big/r^*_i+\sigma^2_{\langle \ell\rangle}\Big/r^*_\ell} \ , \tag{7}$$

$$\sum_{i=1}^{m} \frac{(r^*_i)^2}{\sigma^2_{\langle i\rangle}} = \sum_{j=m+1}^{k} \frac{(r^*_j)^2}{\sigma^2_{\langle j\rangle}} \ , \tag{8}$$

where $\sum_{\ell=1}^{k} r^*_\ell = 1,\ r^*_\ell > 0,\ \ell=1,2,\cdots,k$. Then we can investigate the asymptotic optimality of the sampling ratios of the proposed sequential DSSm (4), that is, as $t\to\infty$, the sampling ratio of each alternative sequentially achieves the optimal decreasing rate of the large deviations of PFS, i.e.,

$$\lim_{t\to\infty} r_i^{(t)} = r^*_i,\ \ a.s.,\ \ i=1,2,\cdots,k \ , \tag{9}$$

where $r_i^{(t)} = t_i/t$, $\sum_{i=1}^{k} r^*_i = 1$, $r^*_i \geq 0$, $i=1,2,\cdots,k$, and $r^* = \left(r^*_1,r^*_2,\cdots,r^*_k\right)$ satisfies (7) and (8). We leave the proof of (7), (8) and (9) to future work.

4 NUMERICAL EXPERIMENTS

In this section, we conduct numerical experiments to test the performances of different allocation procedures for selecting top-m alternatives. The proposed sequential DSSm procedure is compared with the Equal Allocation (EA); Optimal Computing Budget Allocation for selecting top-m alternatives (OCBAm) in Chen et al. (2008); Improved Optimal Computing Budget Allocation for selecting top-m alternatives (OCBAm+) in Zhang et al. (2015); Optimal Computing Budget Allocation for subset selection (OCBAss) in Gao and Chen (2015b) and Improved Optimal Computing Budget Allocation for Subset Selection (OCBASS) in Gao and Chen (2016):

- EA: The procedure allocates the same number of simulation replications to each alternative, i.e., $t_i = \lceil T/k \rceil$, $i=1,2,\cdots,k$, where $\lceil\cdot\rceil$ is a ceiling function.
- OCBAm: The procedure needs to determine a separating parameter c. At each step, it allocates Δ simulation replications to alternatives according to $\frac{r_i}{r_j} = \left(\frac{\sigma_{\langle i\rangle}/\delta_i}{\sigma_{\langle j\rangle}/\delta_j}\right)^2$, where $i,j=1,2,\cdots,k,\ i\neq j$, and $\delta_\ell = \mu_{\langle \ell\rangle} - c$, $\ell=1,2,\cdots,k$. The suggested choice of parameter c in Chen et al. (2008) depends on the number of allocated simulation replications t_i, $i=1,2,\cdots,k$, i.e.,

$$c = \frac{\bar{\sigma}_{\langle m+1\rangle}\mu_{\langle m\rangle} + \bar{\sigma}_{\langle m\rangle}\mu_{\langle m+1\rangle}}{\bar{\sigma}_{\langle m\rangle} + \bar{\sigma}_{\langle m+1\rangle}} \ , \tag{10}$$

where $\bar{\sigma}_{\langle i \rangle} = \sigma_{\langle i \rangle} \big/ \sqrt{t_i}$, $i = 1, 2, \cdots, k$. In *Experiment 1*, we show that such a parameter c could not lead to a good performance when sequentially implementing the OCBAm sampling rule. In the *Experiment 2-4*, we choose

$$c = \frac{\sigma^2_{\langle m+1 \rangle} \mu_{\langle m \rangle} + \sigma^2_{\langle m \rangle} \mu_{\langle m+1 \rangle}}{\sigma^2_{\langle m \rangle} + \sigma^2_{\langle m+1 \rangle}}, \tag{11}$$

for the OCBAm sampling rule, which could lead to a better performance than (10) as shown in *Experiment 1*. The allocated alternative is chosen by the "most starving" sequential rule in Chen and Lee (2011), $A_{t+1}(\mathscr{E}_t) = \arg \max_{i=1,2,\cdots,k} \{(t+1) r_i^{(t)} - t_i\}$.

- OCBAm+: At each step, the procedure allocates a replication according to an asymptotic sampling ratios.
- OCBAss: At each step, the procedure allocates Δ simulation replications to an alternative i^* in order to balance (7) and (8). However, Gao and Chen (2015b) approximate the objective function in the static optimization problem (5) with Bonferroni inequality and derive (7) and (8) by approximately maximizing the approximated PCS, which do not have a rigorous definition of the optimal large deviations rate of PFS for selecting top-m alternatives.
- OCBASS: The procedure allocates Δ simulation replications to an alternative i^* in order to balance (7). However, there are two alternatives that can be chosen as the allocated alternative according to the OCBASS sampling rule. We randomly choose an alternative among the two alternatives as the allocated alternative in the numerical experiments.

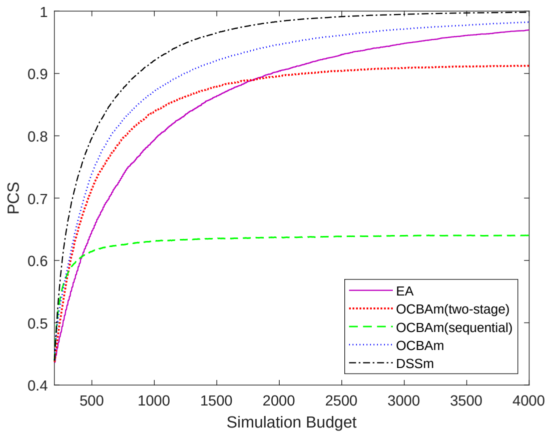
Experiment 1.1 and *Experiment 1.2* have deterministic performance of the alternatives, and the performance of different allocation procedure is measured by the classical PCS. In *Experiment 2* to *Experiment 4*, we set $\Delta = 1$ for OCBAm, OCBAss and OCBASS procedures. The initial number of simulation replications is $n_0 = 10$. In each macro experiment, the performance of each alternative is randomly generated from the normal conjugate prior. The statistical efficiency of each allocation procedure is measured by the IPCS, i.e., $\text{IPCS}_T \stackrel{\Delta}{=} \mathbb{E}[\mathbb{E}[\mathbb{1}\{\widehat{\mathscr{F}}_T^m = \mathscr{F}^m\} | \mathscr{E}_T, \mu]]$ estimated based on 100,000 independent macro experiments. The IPCS is reported as a function of the sampling budget T in each experiment.

Experiment 1: Comparisons of the choice of parameter c in OCBAm procedure. In this experiment, we compare the DSSm procedure with: EA; sequential OCBAm with parameter c in (10) (denoted by OCBAm(sequential)); sequential OCBAm with parameter c in (11) (denoted by OCBAm); two-stage OCBAm with parameter c in (10) (denoted by OCBAm(two-stage)) where the initial simulation replications are equally allocated to estimate unknown parameters and the rest simulation replications are allocated according to OCBAm rule. The initial number of simulation replications is $n_0 = 20$. IPCSs are estimated based on 100,000 independent macro experiments. Two numerical settings are tested in *Experiment 1*:

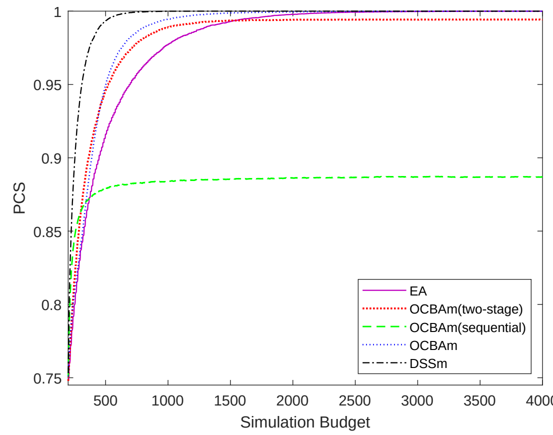
Experiment 1.1 We aim to select top-3 alternatives from 10 competing alternatives with distributions $N(i, i^2)$, $i = 1, 2, \cdots, 10$. The total simulation budget is $T = 4000$ and the first 200 simulation replications are equally allocated to each alternative. Figure 2 (a) shows the performances of different allocation procedures.

Experiment 1.2 We aim to select top-3 alternatives from 10 competing alternatives with distributions $N(i, (11-i)^2)$, $i = 1, 2, \cdots, 10$. The total simulation budget is $T = 4000$ and the first 200 simulation replications are equally allocated to each alternative. Figure 2 (b) shows the performances of different allocation procedures.

From Figure 2 (a), we can see that the PCS of OCBAm(sequential) flattens as the simulation budget grows large. The PCS of EA surpasses the PCS of OCBAm(two-stage) when the simulation budget reaches around 1900. The PCS of OCBAm increases at a fast pace, and the performance of OCBAm is significantly better than the performance of OCBAm(sequential). DSSm performs the best among all allocation procedures. From Figure 2 (b), we can see that the PCS of OCBAm(sequential) increases at a fast pace at the beginning and flattens as the simulation budget grows large. The PCS of OCBAm and



(a) PCS for *Experiment 1.1*



(b) PCS for *Experiment 1.2*

Figure 2: Comparison of PCSs for 5 allocation procedures in Experiment 1. (a) The distribution for each alternative is $N\left(i,(11-i)^2\right)$, $i=1,2,\cdots,10$. (b) The distribution for each alternative is $N\left(i,i^2\right)$, $i=1,2,\cdots,10$. The number of initial simulation replications is $n_0=20$ for each alternative. PCSs are estimated by 10^5 macro replications.

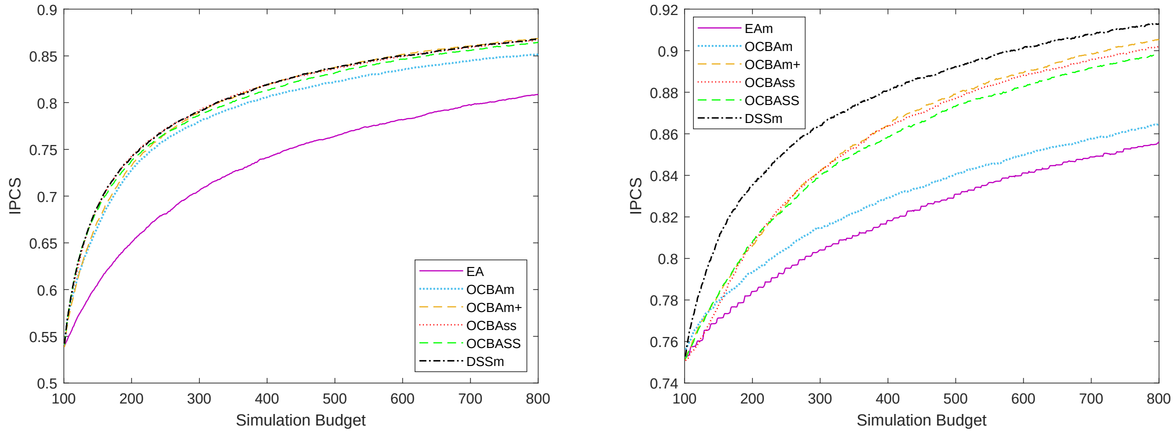
the PCS of EA surpasses the PCS of OCBAm(two-stage) when the simulation budget reaches around 470 and 1560, respectively. The performance of OCBAm is significantly better than the performance of OCBAm(sequential). DSSm performs the best among all allocation procedures. Numerical results show that the performance of sequential OCBAm with parameter c in (11) is significantly better than the performance of sequential OCBAm with parameter c in (10), which implies that the suggested choice of parameter c in Chen et al. (2008) could not lead to a good performance when sequentially implementing the OCBAm sampling rule.

Experiment 2: 10 alternatives with equal variances. We aim to select top-3 alternatives from 10 competing alternatives. In each macro experiment, μ_i, $i=1,2,\cdots.10$ are generated from the normal conjugate priors with hyper-parameters given by $\mu_i^{(0)}=0$, $\sigma_i^{(0)}=1$. True variances are $\sigma_i^2=1$, $i=1,2,\cdots,10$. The total simulation budget is $T=800$ and the first 100 simulation replications are equally allocated to each alternative. Figure 3 (a) shows the performances of different allocation procedures.

From Figure 3 (a), we can see that EA performs the worst among all allocation procedures. The IPCS of OCBAm increases at a faster pace than the IPCS of EA. OCBAss and OCBASS have a comparable performance at the beginning, and the former increases at a faster pace than the latter when the simulation budget reaches around 130. The IPCS of OCBAm+ increases at a slow pace at the beginning, and it surpasses the IPCS of OCBAm, OCBASS and OCBAss when the simulation budget reaches around 125, 255 and 410, respectively. DSSm performs the best among all allocation procedures at the beginning, and it has a comparable performance with OCBAm+ as the simulation budget grows that has a slight edge over the OCBAss.

Experiment 3: 10 alternatives with decreasing variances. We aim to select top-3 alternatives from 10 competing alternatives. In each macro experiment, μ_i, $i=1,2,\cdots,10$ are generated from the normal conjugate priors with hyper-parameters given by $\mu_i^{(0)}=i$, $\sigma_i^{(0)}=(11-i)\big/\sqrt{20}$. True variances are $\sigma_i^2=(11-i)^2$, $i=1,2,\cdots,10$. The total simulation budget is $T=800$ and the first 100 simulation replications are equally allocated to each alternative. Figure 3 (b) shows the performances of different allocation procedures.

From Figure 3 (b), we can see that EA becomes the worst as the simulation budget grows. The IPCS of OCBAss is the lowest at the beginning, and it surpasses EA, OCBAm and OCBASS when the simulation budget reaches around 135, 160 and 220, respectively. OCBAm+ and OCBASS have a



(a) IPCS for *Experiment 2*

(b) IPCS for *Experiment 3*

Figure 3: Comparison of IPCSs for 6 allocation procedures in Experiment 2 and Experiment 3. (a) Prior parameters are $\mu_i^{(0)} = 0$ and $\sigma_i^{(0)} = 1$. True parameters are $\sigma_i^2 = 1$, $i = 1, 2, \cdots, 10$. (b) Prior parameters are $\mu_i^{(0)} = i$ and $\sigma_i^{(0)} = (11 - i)/\sqrt{20}$. True parameters are $\sigma_i^2 = (11 - i)^2$, $i = 1, 2, \cdots, 50$. The number of initial simulation replications is $n_0 = 10$ for each alternative. IPCSs are estimated by 10^5 macro replications.

comparable performance at the beginning, and the former surpasses the latter that is better than OCBAm when the simulation budget reaches around 230. OCBAm+ has a slight edge over OCBAss as the simulation budget grows. DSSm performs the best among all allocation procedures.

Experiment 4: 10 alternatives with increasing variances. We aim to select top-3 alternatives from 10 competing alternatives. In each macro experiment, μ_i, $i = 1, 2, \cdots, 10$ are generated from the normal conjugate priors with hyper-parameters given by $\mu_i^{(0)} = i$, $\sigma_i^{(0)} = i/10$. True variances are $\sigma_i^2 = i^2$, $i = 1, 2, \cdots, 10$. The total simulation budget is $T = 1500$ and the first 100 simulation replications are equally allocated to each alternative. Figure 4 shows the performances of different allocation procedures.

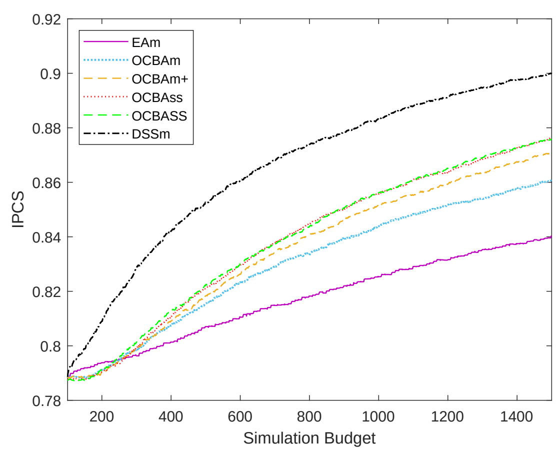


Figure 4: Comparison of IPCSs for 6 allocation procedures in Experiment 4. Prior parameters are $\mu_i^{(0)} = i$ and $\sigma_i^{(0)} = i/10$. True parameters are $\sigma_i^2 = i^2$, $i = 1, 2, \cdots, 10$. The number of initial simulation replications is $n_0 = 10$ for each alternative. IPCSs are estimated by 10^5 macro replications.

From Figure 4, we can see that the IPCSs of OCBAm, OCBAm+, OCBAss and OCBASS decrease slightly at the beginning, and then quickly catch up with the IPCS of EA and surpass the latter when the simulation budget reaches around 240, which could be attributed to the reason that the information of the induced correlations is ignored or not taken fully into account by those allocation procedures in such a low-confidence scenario (i.e., variances of competing alternatives are large relative to the differences in means and the sample size) (Peng et al. 2015). OCBAm+ has an edge over OCBAm as the simulation budget grows. The IPCS of OCBAss surpasses the IPCS of OCBASS when the simulation budget reaches around 650 and the former has a slight edge over the latter as the simulation budget increases. DSSm performs the best among all allocation procedures.

5 CONCLUSIONS

The paper studies the dynamic sampling allocation problem for subset selection. We formulate the dynamic sampling decision as a stochastic dynamic programming problem. Under the Bayesian framework, we propose an efficient sampling procedure named as DSSm, which maximizes the VFA one-step look ahead and is proved to be consistent. Numerical experiments demonstrate that DSSm is more efficient than other tested allocation procedures. Future research includes the asymptotic analysis for the sampling ratios of the proposed sequential sampling procedure. The optimal decreasing rate of the large deviations of the PFS for subset selection could also be a future work.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 71901003 and 72022001.

REFERENCES

Bechhofer, R. G. 1995. *Design and Analysis of Experiment for Statistical Selection, Screening, and Multiple Comparisons*. New York: John Wiley and Sons.

Bertsekas, D. P. 2005. "Dynamic Programming and Suboptimal Control: A Survey from ADP to MPC". *European Journal of Control* 11(4-5):310–334.

Chen, C.-H., D. He, M. Fu, and L. H. Lee. 2008. "Efficient Simulation Budget Allocation for Selecting an Optimal Subset". *INFORMS Journal on Computing* 20(4):579–595.

Chen, C.-H., and L. H. Lee. 2011. *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*, Volume 1. Singapore: World Scientific.

Chen, C.-H., J. Lin, E. Yücesan, and S. E. Chick. 2000. "Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization". *Journal of Discrete Event Dynamic Systems* 10(3):251–270.

Chick, S. E. 2006. "Subjective Probability and Bayesian methodology". *Handbooks in Operations Research and Management Science* 13:225–257.

Chick, S. E., J. Branke, and C. Schmidt. 2010. "Sequential Sampling to Myopically Maximize the Expected Value of Information". *INFORMS Journal on Computing* 22(1):71–80.

DeGroot, M. H. 2005. *Optimal Statistical Decisions*, Volume 82. John Wiley & Sons.

Frazier, P., W. Powell, and S. Dayanik. 2009. "The Knowledge-gradient Policy for Correlated Normal Beliefs". *INFORMS Journal on Computing* 21(4):599–613.

Frazier, P. I. 2014. "A Fully Sequential Elimination Procedure for Indifference-zone Ranking and Selection with Tight Bounds on Probability of Correct Selection". *Operations Research* 62(4):926–942.

Gao, S., and W. Chen. 2015a. "Efficient Subset Selection for the Expected Opportunity Cost". *Automatica* 59:19–26.

Gao, S., and W. Chen. 2015b. "A Note on the Subset Selection for Simulation Optimization". In *Proceedings of the 2015 Winter Simulation Conference*, edited by I. M. T. M. K. R. C. M. L. Yilmaz, W K V Chan and M. D. Rossetti, 3768–3776. Huntington Beach, CA: Institute of Electrical and Electronics Engineers, Inc.

Gao, S., and W. Chen. 2016. "A New Budget Allocation Framework for Selecting Top Simulated Designs". *IIE Transactions* 48(9):855–863.

Glynn, P. W., and S. Juneja. 2004. "A Large Deviations Perspective on Ordinal Optimization". In *Proceedings of the 2004 Winter Simulation Conference*, edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 577–585. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hong, L. J., W. Fan, and J. Luo. 2020. "Review on Ranking and Selection: A New Perspective". *arXiv preprint arXiv:2008.00249*.

Kim, S.-H., and B. L. Nelson. 2001. "A Fully Sequential Procedure for Indifference-zone Selection in Simulation". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 11(3):251–273.

Kim, S.-H., and B. L. Nelson. 2006. "Selecting the Best System". *Handbooks in Operations Research and Management Science* 13:501–534.

Koenig, L. W., and A. M. Law. 1985. "A Procedure for Selecting a Subset of Size m Containing the l Best of k Independent Normal Populations, with Applications to Simulation". *Communications in Statistics-Simulation and Computation* 14(3):719–734.

Peng, Y., C.-H. Chen, M. C. Fu, and J.-Q. Hu. 2015. "Non-monotonicity of Probability of Correct Selection". In *Proceedings of the 2015 Winter Simulation Conference*, edited by I. M. T. M. K. R. C. M. L. Yilmaz, W K V Chan and M. D. Rossetti, 3678–3689. Huntington Beach, CA: Institute of Electrical and Electronics Engineers, Inc.

Peng, Y., C.-H. Chen, M. C. Fu, and J.-Q. Hu. 2018. "Gradient-based Myopic Allocation Policy: An Efficient Sampling Procedure in a Low-confidence Scenario". *IEEE Transactions on Automatic Control* 63(9):3091–3097.

Peng, Y., E. K. Chong, C.-H. Chen, and M. C. Fu. 2018. "Ranking and Selection as Stochastic Control". *IEEE Transactions on Automatic Control* 63(8):2359–2373.

Powell, W. B., and I. O. Ryzhov. 2012. *"Ranking and Selection"*. In Chapter 4 in Optimal Learning, 71-88: New York: John Wiley and Sons.

Rinott, Y. 1978. "On Two-stage Selection Procedures and Related Probability-inequalities". *Communications in Statistics-Theory and Methods* 7(8):799–811.

Shi, Z., Y. Peng, L. Shi, C.-H. Chen, and M. C. Fu. 2021. "Dynamic Sampling Allocation under Finite Simulation Budget for Feasibility Determination". *INFORMS Journal on Computing, forthcoming*.

Shin, D., M. Broadie, and A. Zeevi. 2021. "Practical Nonparametric Sampling Strategies for Quantile-based Ordinal Optimization". In *INFORMS Journal on Computing, forthcoming*. INFORMS.

Wu, D., and E. Zhou. 2018. "Provably Improving the Optimal Computing Budget Allocation Algorithm". In *Proceedings of the 2018 Winter Simulation Conference*, edited by N. M. A. S. S. J. M. Rabe, A.A. Juan and B. Johansson, 1921–1932. Gothenburg, Sweden: Institute of Electrical and Electronics Engineers, Inc.

Zhang, G., H. Li, and Y. Peng. 2020. "Sequential Sampling for a Ranking and Selection Problem with Exponential Sampling Distributions". In *Proceedings of the 2020 Winter Simulation Conference*, edited by S. K. S. L.-M. Z. Z. T. R. K.-H. Bae, B. Feng and R. Thiesing, 2984–2996. Orlando, FL, USA: Institute of Electrical and Electronics Engineers, Inc.

Zhang, J., Y. Liu, Y. Zhao, and T. Deng. 2018. "Emergency Evacuation Problem for a Multi-source and Multi-destination Transportation Network: Mathematical Model and Case Study". *Annals of Operations Research*:1–29.

Zhang, S., L. H. Lee, E. P. Chew, J. Xu, and C.-H. Chen. 2015. "A Simulation Budget Allocation Procedure for Enhancing the Efficiency of Optimal Subset Selection". *IEEE Transactions on Automatic Control* 61(1):62–75.

AUTHOR BIOGRAPHIES

GONGBO ZHANG is a Ph.D. candidate in the Department of Management Science and Information Systems in Guanghua School of Management at Peking University, Beijing, China. He received the B.S. degree in mathematics and applied mathematics from College of Sciences, Northeastern University, Shenyang, China, in 2018. His research interests include stochastic modeling and analysis, simulation optimization. His email address is gongbozhang@pku.edu.cn.

YIJIE PENG is an Assistant Professor in Guanghua School of Management, with affiliate faculty appointments in the Institute of Artificial Intelligence and National Institute of Health Data Science, all at Peking University. His research interests include stochastic modeling and analysis, simulation optimization, machine learning, data analytics, and healthcare. He received his Ph.D. from the School of Management at Fudan University and his bachelors degree from the School of Mathematics and Statistics at Wuhan University. He was awarded the INFORMS Simulation Societys 2019 Outstanding Publication Award. His email address is pengyijie@pku.edu.cn, and his Web page is http://en.gsm.pku.edu.cn/faculty/pengyijie.

JIANGHUA ZHANG received the B.S. degree in mathematics and applied mathematics in 2000 and the M.S. degree in Operational Research and Cybernetics in 2003 from the School of Mathematics, Shandong University, Jinan, China. He received the Ph.D. degree in management science and engineering in 2008 from the School of Management, Fudan University, Shanghai, China. He is now a Professor at Shandong University. His research interests include decision theory, healthcare operation management, and humanitarian logistics. His email address is zhangjianghua@sdu.edu.cn

ENLU ZHOU received the B.S. degree with highest honors in Electrical Engineering from Zhejiang University, China, in 2004, and a PhD degree in Electrical Engineering from the University of Maryland, College Park, in 2009. She is currently an Associate Professor in the School of Industrial & Systems Engineering at Georgia Institute of Technology. Her research interests include simulation, stochastic control, and stochastic optimization. Her e-mail address is enlu.zhou@isye.gatech.edu.