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

SEQUENTIAL SAMPLING FOR A RANKING AND SELECTION PROBLEM WITH EXPONENTIAL SAMPLING DISTRIBUTIONS

Gongbo Zhang, Haidong Li

College of Engineering
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

ABSTRACT

We study a ranking and selection problem with exponential sampling distributions. Under a Bayesian framework, we derive the posterior distribution of the performance parameter, and provide a normal approximation for the posterior distribution based on a central limit theorem to efficiently learn about the performance parameter. We formulate dynamic sampling decision as a stochastic control problem, and propose a sequential sampling procedure, which maximizes a value function approximation one-step ahead and is proved to be consistent. Numerical results demonstrate the efficiency of the proposed method.

1 INTRODUCTION

Ranking and selection (R&S) has been widely studied in simulation optimization field (see Bechhofer et al. 1995; Powell and Ryzhov 2012). The objective is to select the alternative with the largest mean performance from finite k alternatives:

$$\langle 1\rangle \stackrel{\Delta}{=} \arg\max_{i=1,2,\cdots,k} \mu_i \ ,$$

where μ_i, $i=1,2,\cdots,k$ is the unknown mean performance of each alternative i, which can be estimated by Monte Carlo simulation under a finite sampling budget constraint T.

The unknown performance of each alternative μ_i is estimated by independent and identically distributed (i.i.d.) simulation replications $X_{i,\ell}$, $\ell\in\mathbb{Z}^+$, $i=1,2,\cdots,k$. Simulation replications of different alternatives are often assumed to be independent, i.e., $f(\cdot;\theta)=\prod_{i=1}^k f_i(\cdot;\theta_i)$, where $f(\cdot;\theta)$ is the joint density of $(X_{1,\ell},X_{2,\ell},\cdots,X_{k,\ell})$ with $\theta\in\Theta$ containing all unknown parameters in the joint distribution, and $f_i(\cdot;\theta_i)$ is th density of $X_{i,\ell}$ with $\theta_i\in\theta$ containing all unknown parameters in the marginal sampling distribution of alternative i. In the classic R&S problem, the normal sampling distribution is predominantly assumed, i.e., $X_{i,\ell}\sim N\left(\mu_i,\sigma_i^2\right)$, $\ell\in\mathbb{Z}^+$, $i=1,2,\cdots,k$. In the case where the sampling distribution is not normal, the normal assumption can be justified by batching (Chen and Lee 2011).

How to allocate the simulation replications among k alternatives for efficiently selecting the best alternative is the key problem in the R&S problem. The indifference zone (IZ) paradigm is well-researched, which guarantees a probability of correct selection (PCS) up to a certain probability level (see Bechhofer 1954; Rinott 1978; Kim and Nelson 2001; Ni et al. 2017). The optimal computing budget allocation (OCBA) (Chen et al. 2006; Chen et al. 2000; Peng et al. 2016) is a sampling allocation procedure whose goal is to enhance the efficiency for finding the best alternative by intelligently allocating simulation replications based on the (posterior) mean and variance information. Recently, Peng et al. (2018) formulate the sequential allocation problem as a stochastic control problem, and a dynamic allocation procedure that maximizes a value function approximation (VFA) one-step ahead is proved to achieve an asymptotically optimal sampling ratio of PCS for selecting the best under the normal assumption.

Many sequential sampling procedures (Gao et al. 2017; Peng and Fu 2017; Shin et al. 2018; Chen and Ryzhov 2019b) are derived to achieve a static sampling ratio with certain desirable asymptotic properties such as optimizing the large deviations rate of PCS in Glynn and Juneja (2004), which offer no theoretical support for the finite sample performance. In classic R&S problems where the objective is to select the alternative with the optimal mean from alternatives following normal sampling distributions, the sequential sampling procedure derived in such a way to achieve desirable asymptotic properties could be an efficient sampling allocation policy, because the mean-variance tradeoff happens to lead to a good sampling allocation policy in certain scenarios. However, this is not always the case in general settings. Recently, Shin et al. (2016) report the poor finite-time performance of sequentially implementing the asymptotically optimal sampling ratio for selecting the best quantile under some particular distributions. In Peng et al. (2015) and Peng et al. (2018), there are examples where the mean-variance tradeoff could lead to misleading results in a certain low-confidence scenario (i.e., differences between competing alternatives are small, variances are large, and the simulation budget is relatively small), which highlights a significant difference between the asymptotic property and finite-sample property for a sampling procedure.

In this work, we assume simulation replications for each alternative follow an exponential distribution, i.e., $X_{i,\ell} \sim E(\lambda_i),\ i = 1,2,\cdots,k$. The sojourn time in a stationary M/M/1 queueing model follows an exponential distribution (Shortle et al. 2018), and other application background related to the exponential sampling distribution contains product life in reliability engineering (Romeu 2003). Compared with the classic R&S problem with the normal assumption, the literature on the R&S problem with exponential sampling distributions is relatively scarce, with a few exceptions including Gao and Gao (2016), which proposed an sampling procedure called OCBA-exp under an exponential assumption, and Chen and Ryzhov (2019a), which proposed a sampling procedure called BOLD under more general sampling distributions. Both methods are derived to achieve the optimal large deviations rate of PCS without a theoretical support on the finite-sample performance.

Our work focuses on developing an efficient sequential sampling procedure for finding the best alternative among alternatives following exponential sampling distributions. Under a Bayesian framework, we derive the posterior distribution of a performance parameter, i.e., the reciprocal of the mean, and provide a normal approximation for the posterior distribution based on a central limit theorem to efficiently learn about the performance parameter. We formulate the dynamic sampling decision as a stochastic control problem, and propose a sampling procedure called dynamic allocation for exponential distribution (DAED), which maximizes a VFA one-step ahead. The proposed DAED has a theoretical support on the finite-sample performance in derivation, and it is proved to be consistent as the simulation budget goes to infinity. In numerical experiments, our method has an edge over OCBA-exp and BOLD when the number of alternatives is relatively large and the simulation budget is relatively small. Moreover, the numerical results demonstrate that the sampling procedure with asymptotic optimality may perform poorly under finite simulation budget, and the desirable asymptotic properties of OCBA-exp and BOLD are inadequate to explain their good finite-sample performance.

The rest of the paper is organised as follows. Section 2 formulates the proposed problem. The Bayesian and stochastic control framework are introduced in Section 3. The value function approximation and dynamic allocation scheme are proposed in Section 4, and Section 5 provides numerical results. The last section concludes the paper.

2 PROBLEM FORMULATION

Among k alternatives with unknown mean μ_i, $i = 1,2,\cdots,k$, our objective is to find the best alternative:

$$\langle 1 \rangle \overset{\Delta}{=} \arg\max_{i=1,2,\cdots,k} \mu_i \ ,$$

where $\langle i\rangle$, $i=1,2,\cdots,k$ are indices ranked by their performances such that $\mu_{\langle 1\rangle}>\mu_{\langle 2\rangle}>\cdots>\mu_{\langle k\rangle}$, and each μ_i is estimated by Monte Carlo simulation. Let T be the total number of simulation budget and $X_{i,\ell}$ be the ℓ-th replication for alternative i, which follows an independent exponential distribution, i.e., $X_{i,\ell}\sim E(\lambda_i)$, where λ_i is the unknown rate parameter and $\mu_i=1/\lambda_i$, $i=1,2,\cdots,k$. Then the best alternative with the highest unknown mean can also be expressed by

$$\langle 1\rangle \stackrel{\Delta}{=} \arg\min_{i=1,2,\cdots,k} \lambda_i \ .$$

At the t-th step, the sampling policy $A_t(\mathscr{E}_{t-1})\in\{1,2,\cdots,k\}$ allocates the t-th replication to an alternative based on posterior information $\mathscr{E}_{t-1}$ collected throughout $(t-1)$ allocated samples. The information set $\mathscr{E}_t$ contains all sample observations and prior information ζ_0. Under a Bayesian framework, the selection decision at the t-th step $\langle 1\rangle_t$ can be the alternative with the smallest posterior estimate of λ_i, $i=1,2,\ldots,k$. Correct selection means the selected alternative is the best alternative, i.e., $\langle 1\rangle_t=\langle 1\rangle$, and the posterior PCS for selecting the best alternative can be expressed as

$$\begin{aligned}\mathrm{PCS_t} &= \Pr\{\langle 1\rangle_t=\langle 1\rangle \,|\, \mathscr{E}_t\}\\ &= \Pr\{\bigcap_{j=2,3,\cdots,k}\left(\lambda_{\langle 1\rangle_t}<\lambda_{\langle j\rangle_t}\right)|\mathscr{E}_t\}.\end{aligned}$$

Similar to that in Peng et al. (2018), the dynamic sampling decision can be captured by a stochastic control problem with state $\mathscr{E}_t$, action A_{t+1} for $0\le t<T$, transition $\mathscr{E}_t\to\{\mathscr{E}_t\cup X_{i,t+1}\}$, and given a sampling policy $\mathscr{A}_T(\cdot)=(A_1(\cdot),\cdots,A_T(\cdot))$, the expected reward/cost can be defined recursively by

$$\begin{aligned}V_T(\mathscr{E}_T;\mathscr{A}_T(\cdot)) &\triangleq \mathbb{E}[\mathbb{1}(\langle 1\rangle_T=\langle 1\rangle)|\mathscr{E}_T]\\ &= \Pr\left(\lambda_{\langle 1\rangle_T}<\lambda_{\langle j\rangle_T},\ j\in\{2,3,\cdots,k\}\,\middle|\,\mathscr{E}_T\right)\ ,\end{aligned}$$

which is a posterior integrated PCS (IPCS) and $\mathbb{1}\{\cdot\}$ is an indicator function that equals to 1 if the event in the bracket is true, and for $0\le t<T$,

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

where t_i denotes the number of simulation replications allocated to the i-th alternative after allocating t simulation replications, i.e., $t_i=\sum_{j=1}^{t}A_{i,j}(\mathscr{E}_{j-1})$, $i=1,2,\cdots,k$. We have the information set $\mathscr{E}_t=\{\zeta_0,X_1^{(t_1)},X_2^{(t_2)},\cdots,X_k^{(t_k)}\}$, where $X_i^{(t_i)}=(X_{i,1},X_{i,2},\cdots,X_{i,t_i})$. Then, the optimal sampling policy is well defined by

$$\mathscr{A}_T{}^*(\cdot) \stackrel{\Delta}{=} \arg\max_{\mathscr{A}_T(\cdot)} V_0(\zeta_0;\mathscr{A}_T(\cdot))\ ,$$

where ζ_0 contains all hyper-parameters for the parametric family of the prior distribution. In principle, the optimal sampling policy in the stochastic control problem can be solved by backward induction, but it suffers from curse-of-dimensionality. To avoid this issue, we adopt an approximate dynamic programming (ADP) scheme, which makes dynamic decisions based on a VFA and keeps learning it with decisions moving forward (see Powell 2007).

3 POSTERIOR ESTIMATES

We use a Bayesian rule to obtain posterior estimates of the unknown parameters λ_i, $i=1,2,\cdots,k$. We introduce the conjugate prior for the exponential sampling distribution, which is a Gamma distribution $Gamma(\alpha_i^{(0)},\beta_i^{(0)})$ with density $f_i(\lambda_i;\alpha_i^{(0)},\beta_i^{(0)})$, where

$$f_i(\lambda_i;\alpha_i,\beta_i)=\frac{\beta_i{}^{\alpha_i}}{\Gamma(\alpha_i)}\lambda_i^{\alpha_i-1}e^{-\beta_i\lambda_i},\ \ \lambda_i\in(0,\infty),\ \ \alpha_i,\beta_i>0\ ,$$

and $\Gamma(\alpha_i)$ is a gamma function defined by

$$\Gamma(\alpha) \stackrel{\Delta}{=} \int_0^\infty z^{\alpha-1} e^{-z} dz.$$

By conjugacy, the posterior distribution of λ_i is $Gamma(\alpha_i^{(t)}, \beta_i^{(t)})$, where

$$\alpha_i^{(t)} = \alpha_i^{(0)} + t_i \ ,$$

and

$$\beta_i^{(t)} = \beta_i^{(0)} + \sum_{j=1}^{t_i} X_{i,j} = \beta_i^{(0)} + t_i \bar{X}_i^{(t)}.$$

In addition, $\alpha_i^{(t)} > 1$ implies that $f_i(\lambda_i; \alpha_i, \beta_i)$ has a unimodal but skewed shape and the skewness reduces as the value of $\alpha_i^{(t)}$ increases. The posterior mean is $\alpha_i^{(t)} \Big/ \beta_i^{(t)}$ and the posterior variance is $\alpha_i^{(t)} \Big/ (\beta_i^{(t)})^2$.

Denote the prior information set as $\zeta_0 = \{\alpha_i^{(0)}, \beta_i^{(0)},\ i = 1, 2, \cdots, k\}$ and the posterior hyper-parameters set at each step as $\zeta_t = \{\alpha_i^{(t)}, \beta_i^{(t)},\ i = 1, 2, \cdots, k\}$. Notice that the dimension of the state space $\mathscr{E}_t$ grows as the step grows. However, by conjugacy, the sample information set $\mathscr{E}_t$ can be completely determined by the posterior hyper-parameters $\alpha_i^{(t)}$ and $\beta_i^{(t)}$. Then the dimension of the state space is the dimension of the hyper-parameters, which is fixed at any step.

Remark 1 In most existing allocation procedures, the estimated performance for alternative i at step t can be expressed as $\bar{X}_i^{(t)} = \sum_{j=1}^{t_i} X_{i,j} \big/ t_i$, which is the sample mean. Although $\bar{X}_i$ is an unbiased estimator of μ_i, $1 \big/ \bar{X}_i$ is not an unbiased estimator of λ_i, while the bias λ_i / t_i converges to 0 as $t_i \to \infty$ (see Johnson et al. 2002). The estimated performance $\bar{X}_i^{(t)}$ follows an Erlang distribution $Erlang(t_i, t_i\lambda_i)$, which is a special Gamma distribution $Gamma(t_i, t_i\lambda_i)$ with a probability density $f_i(\bar{X}_i^{(t)}; t_i, t_i\lambda_i)$.

Since the Gamma distribution is not closed under linear operations, which would cause a difficulty in calculating the posterior PCS, we use normal distribution $N(\tau_i^{(t)}, (\sigma_i^{(t)})^2)$, $i = 1, 2, \cdots, k$ to approximate the posterior distribution of λ_i, $i = 1, 2, \cdots, k$ where

$$\tau_i^{(t)} = \alpha_i^{(t)} \Big/ \beta_i^{(t)}, \text{ and } (\sigma_\mathrm{i}^{(\mathrm{t})})^2 = \alpha_\mathrm{i}^{(\mathrm{t})} \Big/ (\beta_\mathrm{i}^{(\mathrm{t})})^2.$$

The $Gamma(\alpha_t, \beta_t)$ where $\alpha_t = \alpha_0 t$ and $\beta_t = \beta_0 t$ with $t, \alpha_0 \in \mathbb{Z}^+$ and $\beta_0 > 0$ is proved to converge in distribution to a normal distribution as $t \to \infty$ in the following theorem.

Theorem 1 As $t \to \infty$,

$$\sqrt{t}\left(Y_t - \frac{\alpha_0}{\beta_0}\right) \xrightarrow{d} N\left(0, \frac{\alpha_0}{\beta_0^2}\right) \ ,$$

where $Y_t \sim Gamma(\alpha_t, \beta_t)$, $\alpha_t \in \mathbb{Z}^+$, and $\xrightarrow{d}$ denotes convergence in distribution.

Proof. Note that $Gamma(n, \lambda)$ can be represented as the sum of n independent and identically distributed exponential random variates with rate parameter λ. Then let $Y_t = \sum_{i=1}^{\alpha_t} \chi_i$, where $\chi_i \sim E(\beta_t)$. By a central limit theorem, we have

$$\frac{\sqrt{\alpha_t}\left(\frac{1}{\alpha_t} Y_t - \frac{1}{\beta_t}\right)}{\sqrt{\frac{1}{\beta_t^2}}} \xrightarrow{d} N(0, 1) \ \text{ as } \ \mathrm{t} \to \infty.$$

With $\alpha_t = \alpha_0 t$ and $\beta_t = \beta_0 t$, we have

$$\sqrt{t}\left(\frac{\beta_0}{\sqrt{\alpha_0}}Y_t - \sqrt{\alpha_0}\right) \xrightarrow{d} N(0,1) \ \text{ as } \ t \to \infty \ ,$$

which implies that $\sqrt{t}\left(Y_t - \frac{\alpha_0}{\beta_0}\right) \xrightarrow{d} N\left(0, \frac{\alpha_0}{\beta_0^2}\right)$ as $t \to \infty$. $\square$

The asymptotic result in Theorem 1 justifies the asymptotic normality of $Gamma(\alpha_i^{(t)}, \beta_i^{(t)})$. In practice, normal approximation to gamma distribution performs well when $\alpha_i^{(t)} > 30$ (see Tsokos 2012). Therefore, we use normal distribution $N(\tau_i^{(t)}, (\sigma_i^{(t)})^2)$ to approximate the posterior distribution of λ_i. Then the posterior hyper-parameters set at each step becomes $\zeta_t = \{\tau_i^{(t)}, (\sigma_i^{(t)})^2, i = 1, 2, \cdots, k\}$, and the information set $\mathscr{E}_t$ can be completely determined by the posterior hyper-parameters $\tau_i^{(t)}$ and $(\sigma_i^{(t)})^2$.

4 DYNAMIC SAMPLING POLICY

With the normal approximation for the posterior distribution of the performance parameters, the joint distribution of vector

$$\left(\lambda_{\langle 2\rangle_t} - \lambda_{\langle 1\rangle_t}, \lambda_{\langle 3\rangle_t} - \lambda_{\langle 1\rangle_t}, \cdots, \lambda_{\langle k\rangle_t} - \lambda_{\langle 1\rangle_t}\right) \ ,$$

follows a joint normal distribution with mean vector

$$\left(\tau^{(t)}_{\langle 2\rangle_t} - \tau^{(t)}_{\langle 1\rangle_t}, \tau^{(t)}_{\langle 3\rangle_t} - \tau^{(t)}_{\langle 1\rangle_t}, \cdots, \tau^{(t)}_{\langle k\rangle_t} - \tau^{(t)}_{\langle 1\rangle_t}\right) \ ,$$

and covariance matrix $\Gamma'\Lambda\Gamma$, where $\Lambda = \mathrm{diag}((\sigma^{(\mathrm{t})}_{\langle 1\rangle_\mathrm{t}})^2, (\sigma^{(\mathrm{t})}_{\langle 2\rangle_\mathrm{t}})^2, \cdots, (\sigma^{(\mathrm{t})}_{\langle \mathrm{k}\rangle_\mathrm{t}})^2)$ is a diagonal matrix, and

$$\Gamma = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{pmatrix}_{k\times(k-1)} .$$

In order to derive a sequential sampling procedure with an analytical form, we adopt the VFA technique developed in Peng et al. (2018). At any step t, we treat the $(t+1)$-th step as the last step and try to maximize the expected reward/cost given that the $(t+1)$-th replication is allocated to alternative i:

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

where

$$\widetilde{V}_{t+1}\left(\mathscr{E}_{t+1}\right) \stackrel{\Delta}{=} \Pr(\lambda_{\langle 1\rangle_{t+1}} < \lambda_{\langle j\rangle_{t+1}}, \ j = 2, 3, \cdots, k | \mathscr{E}_{t+1}).$$

The posterior probability above is the integral of a $(k-1)$ dimensional standard normal distribution over a region encompassed by some hyperplanes $\sum_{\ell=1}^{j-1} u_{\ell,(j-1)} z_\ell = \tau^{(t)}_{\langle 1\rangle_t} - \tau^{(t)}_{\langle j\rangle_t}$, where z_i, $i = 1, 2, \cdots, k-1$ are independent standard normal random variables, and $U = [u_{ij}]_{(k-1)\times(k-1)}$ is an upper triangular matrix of Cholesky decomposition for $\Gamma'\Lambda\Gamma = U'U$. Due to exponential decay of the normal density, the posterior probability is approximated by an integral over a maximum tangent inner ball in the integral region, which is visualized in Figure 1. More details about this approximation can be found in Peng et al. (2018). By symmetry of the normal density, maximizing the integral is equivalent to maximizing the volume of the ball, which has the following analytical formula:

$$\widetilde{V}_{t+1}\left(\mathscr{E}_{t+1}\right) = \min_{j \neq 1} \frac{\left(\tau^{(t)}_{\langle j \rangle_t} - \tau^{(t)}_{\langle 1 \rangle_t}\right)^2}{\left(\sigma^{(t)}_{\langle j \rangle_t}\right)^2 + \left(\sigma^{(t)}_{\langle 1 \rangle_t}\right)^2} .$$

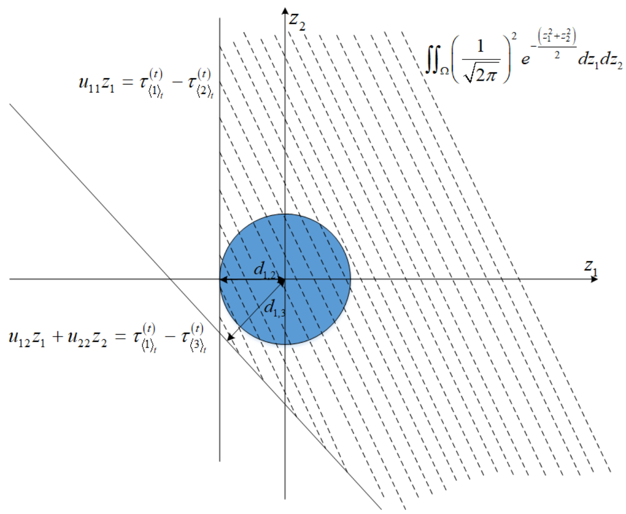

Figure 1: The posterior probability is captured by the size of the largest internally tangent circle.

By certainty equivalent approximation (see Bertsekas 2005),

$$\widetilde{V}_{t+1}\left(\mathscr{E}_t \cup \mathbb{E}\left\{X_{i,t+1} \middle| \mathscr{E}_t\right\}\right) \approx \mathbb{E}\left(\tilde{V}_{t+1}\left(\mathscr{E}_t \cup \left\{X_{i,t+1}\right\}\right) \middle| \mathscr{E}_t\right) ,$$

we have the following VFA. For $\ell, j = 2, 3, \cdots, k$, $\ell \neq j$,

$$\widehat{V}_t\left(\mathscr{E}_t; 1\right) \stackrel{\Delta}{=} \widetilde{V}_{t+1}\left(\mathscr{E}_t \cup E\left\{X_{1,t+1} \middle| \mathscr{E}_t\right\}\right) = \min_{j \neq 1} \frac{\left(\tau^{(t)}_{\langle j \rangle_t} - \tau^{(t)}_{\langle 1 \rangle_t}\right)^2}{\left(\sigma^{(t)}_{\langle j \rangle_t}\right)^2 + \left(\sigma^{(t+1)}_{\langle 1 \rangle_t}\right)^2} ,$$

and

$$\widehat{V}_t\left(\mathscr{E}_t; j\right) \stackrel{\Delta}{=} \widetilde{V}_{t+1}\left(\mathscr{E}_t \cup E\left\{X_{j,t+1} \middle| \mathscr{E}_t\right\}\right) = \min\left\{ \frac{\left(\tau^{(t)}_{\langle j \rangle_t} - \tau^{(t)}_{\langle 1 \rangle_t}\right)^2}{\left(\sigma^{(t+1)}_{\langle j \rangle_t}\right)^2 + \left(\sigma^{(t)}_{\langle 1 \rangle_t}\right)^2}, \min_{\ell \neq j,1} \frac{\left(\tau^{(t)}_{\langle \ell \rangle_t} - \tau^{(t)}_{\langle 1 \rangle_t}\right)^2}{\left(\sigma^{(t)}_{\langle \ell \rangle_t}\right)^2 + \left(\sigma^{(t)}_{\langle 1 \rangle_t}\right)^2} \right\} .$$

A dynamic allocation for exponential distribution (DAED) that optimizes the VFA is given by

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

The proposed DAED utilizes the information on the posterior means and variances of the posterior distributions, and it is proved to be consistent in the following theorem.

Theorem 2 DAED defined by (1) is consistent, i.e.,

$$\lim_{t \to \infty} \langle 1 \rangle_t = \langle 1 \rangle \quad a.s.$$

Proof. We only need to prove that every alternative i, $i=1,2,\cdots,k$ will be sampled infinitely often almost surely, and then the conclusion of theorem holds by the consistency of the Bayesian estimate. If alternative i is sampled finitely often and alternative j is sampled infinitely often, then by the law of large numbers, we have

$$\lim_{t\to\infty}\left[\widehat{V}_t\left(\mathscr{E}_t;i\right)-\widetilde{V}_t\left(\mathscr{E}_t\right)\right]>0,\ a.s.,\quad \lim_{t\to\infty}\left[\widehat{V}_t\left(\mathscr{E}_t;j\right)-\widetilde{V}_t\left(\mathscr{E}_t\right)\right]=0,\ a.s.\ ,$$

which contradicts with the sampling scheme that the alternative with the largest VFA is sampled in DAED. Therefore, DAED must be consistent.

□

5 NUMERICAL EXPERIMENTS

In the numerical experiments, we test the performance of different sampling procedures for selecting the best alternative from alternatives following the exponential sampling distribution. We use the following approaches for comparison.

- Equal allocation (EA), which equally allocates the total simulation budget to each alternative, i.e., $t_i = T\big/k$, $i=1,2,\cdots,k$.
- Optimal computing budget allocation (OCBA) in Chen C-H and Lee (2011), which is an efficient sampling procedure derived for selecting the largest mean under the normal assumption. At each step, it allocates Δ simulation replications to alternatives according to the sampling ratios:

$$\frac{w_{\langle i\rangle}}{w_{\langle j\rangle}}=\frac{\sigma^2_{\langle i\rangle}\left(\mu_{\langle 1\rangle}-\mu_{\langle j\rangle}\right)^2}{\sigma^2_{\langle j\rangle}\left(\mu_{\langle 1\rangle}-\mu_{\langle i\rangle}\right)^2},\ \ i,j\neq\langle 1\rangle\,,\ \ w_{\langle 1\rangle}=\sigma_{\langle 1\rangle}\sqrt{\sum\nolimits_{i\neq\langle 1\rangle}\frac{w^2_{\langle i\rangle}}{\sigma^2_{\langle i\rangle}}}\ , \tag{2}$$

  where $w_{\langle i\rangle}$ is the sampling ratio for alternative $i=1,2,\cdots,k$. To implement OCBA sequentially, we use the "most starving" rule that minimizes the gap between the number of allocated replications to each alternative and the number of replications prescribed by (2) (see Chen and Lee 2011).
- Optimal computing budget allocation with exponential sampling distributions (OCBA-exp) in Gao and Gao (2016), which has an analytical sampling ratios derived to approximately optimize the large deviations rate of PCS under the exponential sampling distributions. At each step, it allocates Δ simulation replications to alternatives according to the sampling ratios:

$$\frac{w_{\langle i\rangle}}{w_{\langle j\rangle}}=\frac{\sigma_{\langle i\rangle}\left(\mu_{\langle 1\rangle}-\mu_{\langle j\rangle}\right)}{\sigma_{\langle j\rangle}\left(\mu_{\langle 1\rangle}-\mu_{\langle i\rangle}\right)},\ \ i,j\neq\langle 1\rangle\,,\ \ w_{\langle 1\rangle}=\sqrt{\sum\nolimits_{i\neq\langle 1\rangle}w^2_{\langle i\rangle}},$$

  where $\mu_i = 1/\lambda_i$ and $\sigma_i^2 = 1/\lambda_i^2$, $i=1,2,\ldots,k$. We use the "most starving" rule to implement a sequential OCBA-exp algorithm in numerical experiments.
- Balancing optimal large deviations (BOLD) in Chen and Ryzhov (2019a), which is derived to sequentially achieve the optimal large deviations rate of PCS for general sampling distribution. For the exponential sampling distributions, at each step, BOLD allocates Δ simulation replications to alternative $\langle 1\rangle$ when

$$\sum_{j\neq\langle 1\rangle}\frac{\frac{\lambda_{\langle 1\rangle}\left(w_{\langle 1\rangle}+w_{\langle j\rangle}\right)}{w_{\langle 1\rangle}\lambda_{\langle 1\rangle}+w_{\langle j\rangle}\lambda_{\langle j\rangle}}-1-\log\frac{\lambda_{\langle 1\rangle}\left(w_{\langle 1\rangle}+w_{\langle j\rangle}\right)}{w_{\langle 1\rangle}\lambda_{\langle 1\rangle}+w_{\langle j\rangle}\lambda_{\langle j\rangle}}}{\frac{\lambda_{\langle j\rangle}\left(w_{\langle 1\rangle}+w_{\langle j\rangle}\right)}{w_{\langle 1\rangle}\lambda_{\langle 1\rangle}+w_{\langle j\rangle}\lambda_{\langle j\rangle}}-1-\log\frac{\lambda_{\langle j\rangle}\left(w_{\langle 1\rangle}+w_{\langle j\rangle}\right)}{w_{\langle 1\rangle}\lambda_{\langle 1\rangle}+w_{\langle j\rangle}\lambda_{\langle j\rangle}}}>1,$$

  and if the above inequality does not hold, BOLD allocates Δ simulation replications to alternative $\langle j\rangle$, $j\neq 1$, where

$$\langle j\rangle=\arg\min_{j=2,\cdots k}w_{\langle 1\rangle}\log\frac{\lambda_{\langle 1\rangle}\left(w_{\langle 1\rangle}+w_{\langle j\rangle}\right)}{w_{\langle 1\rangle}\lambda_{\langle 1\rangle}+w_{\langle j\rangle}\lambda_{\langle j\rangle}}+w_{\langle j\rangle}\log\frac{\lambda_{\langle j\rangle}\left(w_{\langle 1\rangle}+w_{\langle j\rangle}\right)}{w_{\langle 1\rangle}\lambda_{\langle 1\rangle}+w_{\langle j\rangle}\lambda_{\langle j\rangle}}.$$

In numerical experiments, we set $\Delta = 1$ for OCBA, OCBA-exp, and BOLD. We compare the proposed DAED with seven different sampling policies: EA; OCBA, which allocates simulation replications according to the OCBA rule with parameters estimated sequentially by available sample information; OCBAt, which allocates simulation replications according to the OCBA rule given the true parameters; OCBA-exp, which allocates simulation replications according to the OCBA-exp rule with parameters estimated sequentially by available sample information; OCBA-expt, which allocates replications according to the OCBA-exp rule given the true parameters; BOLD, which allocates simulation replications according to the BOLD rule with parameters estimated sequentially by available sample information; BOLDt, which allocates simulation replications according to the BOLD rule given the true parameters. The initial $n_0 = 10$ simulation replications are allocated to each alternative for estimating the parameters. The efficiency of each sampling procedure is measured by IPCS estimated by 100,000 independent macro experiments. IPCSs of different sampling procedures are presented as functions of the sampling budget t, i.e., $\mathrm{IPCS_t} \stackrel{\Delta}{=} \mathbb{E}\left[1\left(\langle 1\rangle_{\mathrm{t}} = \langle 1\rangle\right)\right]$.

Experiment 1 : There are $k = 10$ competing alternatives following exponential sampling distributions. The hyper-parameters in the prior are $\alpha_i^{(0)} = 2$ and $\beta_i^{(0)} = 10$, $i = 1,2,\cdots,10$. The total simulation budget is $T = 500$ and the first 100 simulation replications are equally allocated to each alternative.

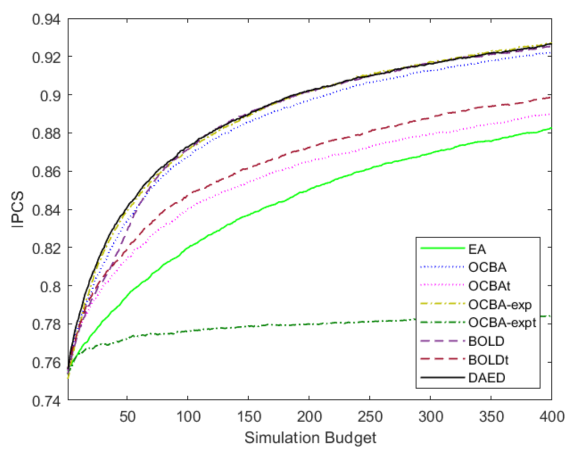

Figure 2: 10 alternatives following exponential sampling distributions with unknown parameters following gamma prior *Gamma* $(2,10)$. Initial simulation replications $n_0 = 10$ for each alternative. IPCSs are estimated by 10^5 independent experiments.

From Figure 2, we can see that OCBA-expt is the worst among all sampling procedures. BOLDt has an edge over OCBAt that is better than EA. The IPCS of BOLD increases in a slower rate than the IPCS of OCBA at the beginning and surpasses the latter when the simulation budget reaches 70. BOLD catches up with DAED and OCBE-exp as the simulation budget increases. DAED and OCBA-exp have a comparable performance, which is the best among all procedures. Note that the OCBA-exp and BOLD rules lead to much poorer performances when the true parameters are given as inputs, which implies that the desirable asymptotic property is inadequate to explain the good finite-sample performance of OCBA-exp and BOLD in the experiment.

Experiment 2 : There are $k = 10$ competing alternatives following exponential sampling distributions. The hyper-parameters in the prior shape parameters are $\alpha_i^{(0)} = 5$ and $\beta_i^{(0)} = 10$, $i = 1,2,\cdots,10$. The skewness of the prior distribution in Experiment 2 is smaller than the skewness of the prior distribution in Experiment 1. The total simulation budget is $T = 500$ and the first 100 simulation replications are equally allocated to each alternative.

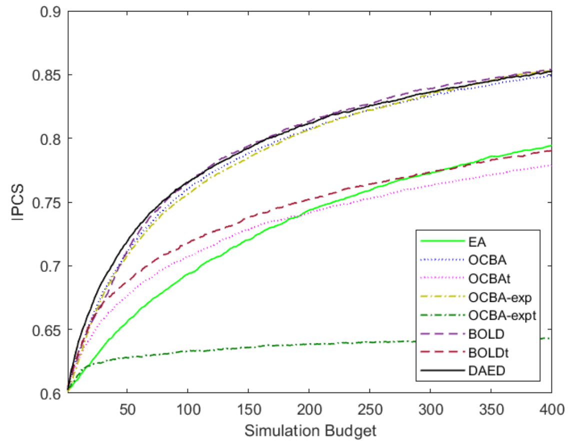

Figure 3: 10 alternatives following exponential sampling distributions with unknown parameters following gamma prior $Gamma(5,10)$. Initial simulation replications $n_0 = 10$ for each alternative. IPCSs are estimated by 10^5 independent experiments.

From Figure 3, we can see that the IPCS of OCBA-expt flattens and becomes the worst as the simulation budget grows. The IPCS of EA surpasses OCBAt when the simulation budget reaches 200, and it also surpasses BOLDt when the simulation budget reaches 300. The IPCS of BOLD increases with the simulation budget at a slow pace at the beginning, and it surpasses OCBA with the number of allocated simulation replications up to 50. BOLD catches up with DAED and OCBA-exp, which have a comparable performance and a slight edge over OCBA, as the simulation budget increases.

Experiment 3: There are $k = 30$ competing alternatives following the exponential sampling distributions. The hyper-parameters in the prior are $\alpha_i^{(0)} = 5$ and $\beta_i^{(0)} = 100$, $i = 1,2,\cdots,30$. The total simulation budget is $T = 900$ and the first 300 simulation replications are equally allocated to each alternative.

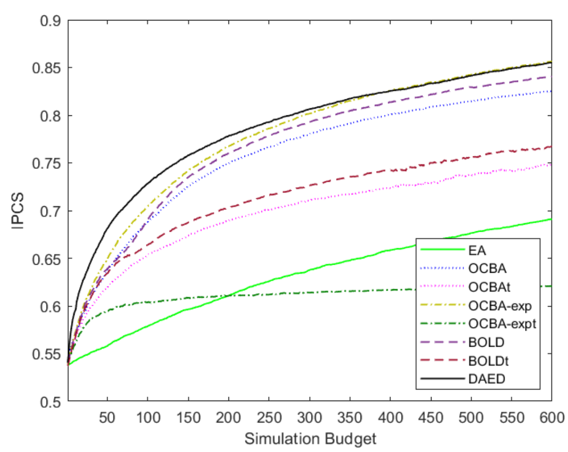

Figure 4: 30 alternatives following exponential sampling distributions with unknown parameters following gamma prior $Gamma(5,100)$. Initial simulation replications $n_0 = 10$ for each alternative. IPCSs are estimated by 10^5 independent experiments.

From Figure 4, we can see that the IPCS of EA increases at a slow pace at the beginning, and it surpasses the OCBA-expt which reaches a plateau when the simulation budget reaches 200. DAED stands out among all sampling procedures in terms of the performance. DAED has a substantial edge over OCBA-exp at the beginning, and the difference between DAED and OCBA-exp vanishes as the simulation budget increases. The advantage of DAED appears to be more significant when the number of competing alternatives is larger. DAED and OCBA-exp lead BOLD throughout the experiment, and the three sampling procedures developed under exponential sampling distributions dominate OCBA that is derived under normal sampling distributions. Although BOLDt and OCBAt are better than EA, their performances are much worse than the sequential sampling procedures with parameters updated sequentially, which again indicates the significant difference between the asymptotic property and finite-sample performance for sampling procedures.

Experiment 4 : There are $k = 5$ competing alternatives following exponential sampling distributions. The prior is a uniform distribution on $[0.3, 0.7]$. Although the prior is not the conjugate prior of the exponential distribution assumed in deriving DAED, the proposed sampling procedure can still apply as we set the hyper-parameters $\alpha_0 = \beta_0 = 0$ for the algorithm in the experiment. The total simulation budget is $T = 450$ and the first 50 simulation replications are equally allocated to each alternative.

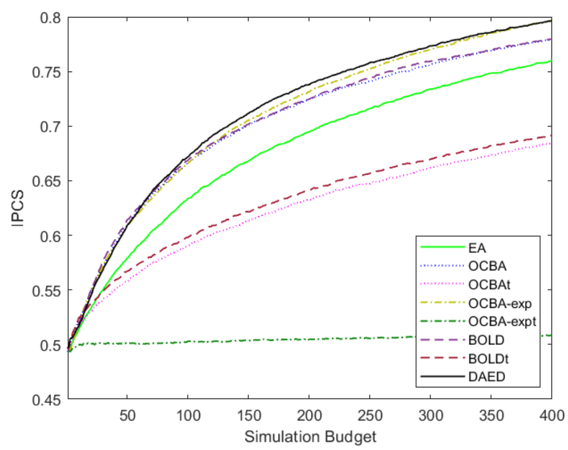

Figure 5: 5 alternatives following exponential sampling distributions with unknown parameters following uniform prior on $[0.3, 0.7]$. Initial replications $n_0 = 10$ for each alternative. IPCSs are estimated by 10^5 independent experiments.

From Figure 5, we can see that the IPCS of OCBA-expt is almost flat. EA is much better than OCBAt and BOLDt. The performance of BOLD is comparable to OCBA and the difference between BOLD and EA becomes smaller as the simulation budget increases. DAED has a slight lead over OCBA-exp at the beginning, and their performances become almost identical as simulation budget grows large.

Compared with 10 competing alternatives, numerical results show that the performance of DAED is better when the number of competing alternatives is 30. The advantage of DAED appears to be more significant as the number of competing alternatives increases.

6 CONCLUSION

This paper studies the dynamic sampling allocation problem for selecting the best alternative from alternatives following exponential sampling distributions. Under a Bayesian framework, we develop a normal approximation scheme based on a central limit theorem to efficiently learn a performance parameter. A stochastic control framework is proposed to capture the dynamic sampling decision, and an efficient sequential sampling procedure called DAED is derived by maximizing a VFA one-step ahead, which is proved to be consistent. Analysis on the convergence rate of the normal approximation and deriving the asymptotic sampling ratio of DAED deserve future work.

Numerical experiments show that when the number of competing alternatives is relatively large and the simulation budget is relatively small, DAED is superior to the existing methods developed for the problem setting in this work, which could be attributed to the finite-sample theoretical support in deriving DAED. Particularly, we find that the asymptotically optimal sampling ratio per se may not lead to a good finite-sample performance for a sampling procedure, whereas sequentially implementing the asymptotically optimal sampling ratio properly can result in a good sampling policy in some circumstances.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation of China (NSFC) under Grants 71901003, 91846301, 71790615, 71720107003, and 71690232.

REFERENCES

Bechhofer, R. E. 1954. "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances". *Annals of Mathematical Statistics* 25(1):16–39.

Bechhofer, R. E., T. J. Santner, and D. M. Goldsman. 1995. *Design and Analysis 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., D. He, and M. C. Fu. 2006. "Efficient Dynamic Simulation Allocation in Ordinal Optimization". *IEEE Transactions on Automatic Control* 51(12):2005–2009.

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

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

Chen, Y., and I. O. Ryzhov. 2019a. "Balancing Optimal Large Deviations in Ranking and Selection". In *Proceedings of the 2019 Winter Simulation Conference*, edited by S. L.-M. M. R. C. S. P. H. N. Mustafee, K.-H.G. Bae and Y.-J. Son, 3368–3379. National Harbor, MD, USA: Institute of Electrical and Electronics Engineers, Inc.

Chen, Y., and I. O. Ryzhov. 2019b. "Complete Expected Improvement Converges to an Optimal Budget Allocation". *Advances in Applied Probability* 51(1):209–235.

Gao, F., and S. Gao. 2016. "Optimal Computing Budget Allocation With Exponential Underlying Distribution". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 682–689. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Gao, S., W. Chen, and L. Shi. 2017. "A New Budget Allocation Framework for the Expected Opportunity Cost". *Operations Research* 65(3):787–803.

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

Johnson, R. A., D. W. Wichern et al. 2002. *Applied Multivariate Statistical Analysis*, Volume 5. Prentice hall Upper Saddle River, NJ: Duxbury.

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

Ni, E. C., D. F. Ciocan, S. G. Henderson, and S. R. Hunter. 2017. "Efficient Ranking and Selection in Parallel Computing Environments". *Operations Research* 65(3):821–836.

Peng, Y., C. Chen, M. C. Fu, and J. Hu. 2016. "Dynamic Sampling Allocation and Design Selection". *INFORMS Journal on Computing* 28(2):195–208.

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(8):3091–3097.

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

Peng, Y., and M. C. Fu. 2017. "Myopic Allocation Policy with Asymptotically Optimal Sampling Rate". *IEEE Transactions on Automatic Control* 62(4):2041–2047.

Powell, W. B. 2007. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. New Jersey: John Wiley & Sons.

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". *Communication in Statistics: Theory and Methods* 7(8):799–811.

Romeu, J. L. 2003. "Practical Reliability Engineering". *Technometrics* 45(2):173.

Shin, D., M. Broadie, and A. Zeevi. 2016. "Tractable Sampling Strategies for Quantile-Based Ordinal Optimization". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 847–858. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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

Shortle, J. F., J. M. Thompson, D. Gross, and C. M. Harris. 2018. *Fundamentals of Queueing Theory*. Hoboken, USA: John Wiley & Sons.

Tsokos, C. P. 2012. *Probability for Engineering, Mathematics, and Sciences*. Mason, OH, United States: Cengage Learning.

AUTHOR BIOGRAPHIES

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

HAIDONG LI is a Ph.D. candidate in the Department of Industrial Engineering and Management, Peking University, Beijing, China. He received his B.S. Degree from the Department of Engineering Mechanics at Peking University. His research interests include simulation optimization and network analysis. He is a visiting scholar under the supervision of Prof. Henry Lam during 2019-2020. His email address is haidong.li@pku.edu.cn.

YIJIE PENG is an Assistant Professor 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 from Wuhan University, Wuhan, China, in 2007, and the Ph.D. degree in management science from Fudan University, Shanghai, China, in 2014, respectively. His research interests include stochastic modeling and analysis, simulation optimization, machine learning, data analytics, and healthcare. His email address is pengyijie@pku.edu.cn, and his Web page is http://en.gsm.pku.edu.cn/faculty/pengyijie.