Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

A REVIEW OF STATIC AND DYNAMIC OPTIMIZATION FOR RANKING AND SELECTION

Yijie Peng

Dept. of Industrial Engineering and Management
Peking University
Beijing, 100871, CHINA

Chun-Hung Chen

Dept. of Systems Engineering and Operations Research
George Mason University
Fairfax, VA 22030, USA

Edwin K. P. Chong

Dept. of Electrical and Computer Engineering
Colorado State University
Fort Collins, CO, 80523-1373, USA

Michael C. Fu

Robert H. Smith School of Business
University of Maryland
College Park, MD 20742, USA

ABSTRACT

We review static and dynamic optimization formulations for simulation allocation and selection procedures and revisit several sampling approaches under a single umbrella. We conduct some new simulation experiments to illustrate where the static optimization approach may be inadequate to capture the dynamic sampling decisions and show how many existing sampling procedures ignore certain important considerations.

1 INTRODUCTION

Ranking and selection (R&S) is an actively studied field in simulation (see Bechhofer et al. 1995; Chen and Lee 2011; Powell and Ryzhov 2012). The objective is to find the alternative with the largest mean from finite alternatives with unknown means:

$$\langle 1 \rangle \doteq \arg\max_{i=1,..,k} \mu_i \ .$$

The unknown mean of each alternative i, denoted by μ_i, can be estimated by sampling independently and identically distributed (i.i.d.) replications $X_{i,\ell}$, $\ell \in \mathbb{Z}^+$, $i = 1,\ldots,k$. Suppose $(X_{1,\ell},\ldots,X_{k,\ell})$ follows a joint distribution $Q(\cdot;\theta)$, where θ contains all unknown parameters in the sampling distribution, including the unknown means. The most common assumption is that the replications of different alternatives are independent, i.e., $Q(\cdot;\theta) = \prod_{i=1}^{k} Q_i(\cdot;\theta_i)$, where θ_i contains all unknown parameters in the (marginal) sampling distribution Q_i of the ith alternative. Normal sampling distribution is often assumed, i.e., $X_{i,\ell} \sim N(\mu_i,\sigma_i^2)$, $i = 1,\ldots,k$.

Applications of R&S include optimizing complex discrete event dynamic systems (DEDS) that are computationally intensive to simulate (Chen and Lee 2011), and finding the most effective drugs, where the economic cost of each sample for testing the effectiveness of the drug is expensive (Powell and Ryzhov 2012). In the case where the normal assumption is not satisfied, batching is often used to make the batched sample means approximately normal.

The problem of interest is how to allocate the simulation replications among the k alternatives to efficiently find the best alternative. A well-researched paradigm to address this problem is in an indifference zone (IZ) framework that dates back to Bechhofer (1954) and Rinott (1978). Recent developments can be found in Kim and Nelson (2001), Frazier (2014), Luo et al. (2015), and Ni et al. (2017). The sampling allocation procedures in the IZ framework guarantee a probability of correction selection (PCS) up to a certain level. In the IZ framework, the guarantee of PCS is primary, whereas the (statistical) efficiency for finding the best is secondary. Due to the need for guaranteeing a PCS level in the worst-case configuration, the sampling procedures in the IZ framework tend to allocate more replications than necessary in practice.

As a result, much of the recent work has focused on sampling allocation procedures whose primary goal is to enhance the efficiency for finding the best alternative, including optimal computing budget allocation (OCBA) (Chen et al. 2000; Chen et al. 2006), expected value of information (EVI) (Chick and Inoue 2001; Chick et al. 2010), knowledge gradient (KG) (Gupta and Miescke 1996; Frazier et al. 2008), and expected improvement (EI) (Jones et al. 1998; Ryzhov 2016), where the sampling-allocation decision is either formulated as a static optimization or a one-step optimization problem. Recently, Peng et al. (2016) and Peng et al. (2018b) formulate the decision in R&S as a dynamic optimization problem, where the optimal solution satisfies the Bellman equation of a stochastic control problem. In this work, we review static optimization and dynamic optimization approaches in R&S and revisit several sampling allocation procedures under a single umbrella. We conduct some new simulation experiments to illustrate where the static optimization approach may be inadequate to capture the dynamic sampling decisions and show how many existing sampling procedures ignore certain important considerations.

2 STATIC OPTIMIZATION

The early optimal sampling allocation procedures attempt to solve the problem formulated as a static optimization problem under both frequentist and Bayesian frameworks.

2.1 Frequentist Framework

One approach to the optimal sampling allocation is to formulate it as the following static optimization problem, which maximizes the PCS under fixed simulation budget:

$$\max_{n_1,..,n_k} \mathbb{P}\left(\bar{X}_{\langle 1\rangle}^{(n)} > \bar{X}_j^{(n)},\ j \neq \langle 1\rangle | \theta\right), \quad s.t.\ \sum_{i=1}^{k} n_i = n, \tag{1}$$

where n is the total number of simulation replications, n_i is the number of replications allocated to alternative i, and

$$\bar{X}_i^{(n)} \doteq \frac{\sum_{\ell=1}^{n_i} X_{i,\ell}}{n_i} .$$

Solving optimization problem (1) is difficult. To reduce the complexity of solving it, a surrogate problem is to maximize the large derivations rate of the probability of false selection (PFS) (Glynn and Juneja 2004):

$$\max_{w_1,..,w_k} \lim_{n\to\infty} -\frac{1}{n}\log\left(1-\mathbb{P}\left(\bar{X}_{\langle 1\rangle}^{(n)} > \bar{X}_j^{(n)},\ j \neq \langle 1\rangle | \theta\right)\right), \quad s.t.\ \sum_{i=1}^{k} w_i = 1, \tag{2}$$

where $w_i \doteq n_i/n$, $i = 1,\ldots,k$. In the case of normal sampling distributions, optimization problem (2) is equivalent to

$$\max_{w_1,..,w_k} \min_{j\neq\langle 1\rangle} G_j(w_{\langle 1\rangle}, w_j; \theta), \tag{3}$$

where

$$G_j(w_{\langle 1\rangle}, w_j; \theta) = \frac{(\mu_{\langle 1\rangle} - \mu_j)^2}{2(\sigma_{\langle 1\rangle}^2/w_{\langle 1\rangle} + \sigma_j^2/w_j)}, \quad j \neq \langle 1\rangle .$$

An approximate solution for the optimization above yields the following OCBA formula (Chen et al. 2000):

$$\frac{w_i}{w_j} = \left(\frac{\sigma_i \delta_j}{\sigma_j \delta_i}\right)^2, \quad i,j \neq \langle 1\rangle, \tag{4}$$

$$w_{\langle 1\rangle} = \sigma_{\langle 1\rangle}\sqrt{\sum_{i\neq\langle 1\rangle} \frac{w_i^2}{\sigma_i^2}}, \tag{5}$$

where $\delta_i \doteq \mu_{\langle 1\rangle} - \mu_i$, $i \neq \langle 1\rangle$. Notice that there is unknown parameter θ in optimization problems (1) and (2). A two-stage procedure can be used to implement the sampling allocation procedures derived in a static optimization problem under the frequentist framework (see Section 4).

2.2 Bayesian Framework

Neither optimization problem (1) nor (2) quantifies the uncertainty of parameter θ. To address this problem, a Bayesian framework is introduced. Suppose θ follows a prior distribution $F(\cdot;\zeta_0)$ that reflects our prior knowledge on the unknown parameter, where ζ_0 in the prior distribution is called a *hyperparameter*. Let $X_i^{(t)} \doteq (X_{i,1},\ldots,X_{i,t_i})$ and $\mathscr{E}_t \doteq \{\zeta_0, X_1^{(t)},\ldots,X_k^{(t)}\}$, where t is the number of allocated replications. The posterior and predictive distributions can be calculated using Bayes rule. In the case where the prior distribution is a conjugate prior of the sampling distribution, the posterior distribution lies in the same parametric family of the prior distribution, i.e., $F(\cdot;\zeta_t)$ where ζ_t is the posterior hyperparameter. Under the conjugate prior, the information set $\mathscr{E}_t$ can be completely determined by the posterior hyper-parameters, i.e., $\mathscr{E}_t = \zeta_t$. The conjugate prior for the normal distribution $N(\mu_i,\sigma_i^2)$ with unknown mean and known variance is a normal distribution $N(\mu_i^{(0)},(\sigma_i^{(0)})^2)$. The posterior distribution of μ_i is $N(\mu_i^{(t)},(\sigma_i^{(t)})^2)$, where

$$\mu_i^{(t)} = (\sigma_i^{(t)})^2\left(\frac{\mu_i^{(0)}}{(\sigma_i^{(0)})^2} + \frac{t_i\bar{X}_i^{(t)}}{\sigma_i^2}\right), \qquad (\sigma_i^{(t)})^2 = \left(\frac{1}{(\sigma_i^{(0)})^2} + \frac{t_i}{\sigma_i^2}\right)^{-1},$$

and the predictive distribution of X_{i,t_i+1} is $N(\mu_i^{(t)},\sigma_i^2+(\sigma_i^{(t)})^2)$. If $\sigma_i^{(0)}\to\infty$, $\mu_i^{(t)} = \bar{X}_i^{(t)}$, and the prior is the uninformative prior in this case. For a normal distribution with unknown variance, there is a normal-gamma conjugate prior (see DeGroot 2005).

Under the Bayesian framework, the optimal two-stage sampling allocation can be formulated as the following static optimization problem:

$$\max_{n_1,\ldots,n_k} \mathbb{E}\left[\mathbb{P}\left(\bar{X}_{\langle 1\rangle}^{(n)} > \bar{X}_j^{(n)},\ j\neq\langle 1\rangle \middle| \mathscr{E}_n\right)\middle|\mathscr{E}_{n_0}\right], \quad s.t.\ \sum_{i=1}^k n_i = n. \tag{6}$$

or its Lagrangian relaxation (Chick and Inoue 2001). In (6), $\langle 1\rangle$ is a measurable function of $(\mu_1,\ldots,\mu_k)$ following a posterior distribution conditional on $\mathscr{E}_n$, and the simulation replications allocated at the second stage follow a predictive distribution given information set $\mathscr{E}_{n_0}$. Solving optimization (6) is complicated, and requires approximations to obtain a closed-form solution (Chick and Inoue 2001). Optimization (6) lays out a theoretical foundation for optimal two-stage sampling allocation. However, it has been widely observed that sequential sampling procedures that incrementally increase simulation replications at the second stage typically perform better than the two-stage procedure (see Chen and Lee 2011).

3 DYNAMIC OPTIMIZATION

The dynamic decisions in R&S problem can be formulated as an allocation and selection (A&S) policy (Peng et al. 2016). The allocation policy is a sequence of mappings $\mathscr{A}_t(\cdot) = (A_1(\cdot),\ldots,A_t(\cdot))$ that sequentially allocates each sample to an alternative based on collected information, and the selection policy $\mathscr{S}(\cdot)$ makes a final decision to select the best alternative after exhausting all simulation replications.

3.1 Optimal A&S Policy

Define $\mathscr{E}_t^a \doteq \left\{\mathscr{A}_t(\mathscr{E}_{t-1}^a);\mathscr{E}_t\right\}$, and $A_{i,t}(\mathscr{E}_{t-1}^a) \doteq \mathbf{1}\left\{A_t(\mathscr{E}_{t-1}^a) = i\right\}$. The information collection procedure following a sampling allocation policy in the R&S problem is illustrated by (7) from Peng et al. (2018b) for allocating four samples among three alternatives. Given prior information $\mathscr{E}_0$, collected information set $\mathscr{E}_4^a$ is determined by the two tables in the figure. The allocation decision represented by the table on the left determines the (bold) observable elements in the table on the right.

$$\begin{array}{ccc} X_{1,1} & \mathbf{X_{2,1}} & X_{3,1} \\ \mathbf{X_{1,2}} & X_{2,2} & X_{3,2} \\ \mathbf{X_{1,3}} & X_{2,3} & X_{3,3} \\ X_{1,4} & X_{2,4} & \mathbf{X_{3,4}} \end{array} \qquad \begin{array}{ccc} A_{1,1}(\zeta_0)=0 & A_{2,1}(\zeta_0)=1 & A_{3,1}(\zeta_0)=0 \\ A_{1,2}(\mathscr{E}_1^a)=1 & A_{2,2}(\mathscr{E}_1^a)=0 & A_{3,2}(\mathscr{E}_1^a)=0 \\ A_{1,3}(\mathscr{E}_2^a)=1 & A_{2,3}(\mathscr{E}_2^a)=0 & A_{3,3}(\mathscr{E}_2^a)=0 \\ A_{1,4}(\mathscr{E}_3^a)=0 & A_{2,4}(\mathscr{E}_3^a)=0 & A_{3,4}(\mathscr{E}_3^a)=1 \end{array} \tag{7}$$

The sampling decision and information flow have an interactive relationship shown by (8) from Peng et al. (2018b).

$$\zeta_0 \twoheadrightarrow \mathscr{E}_1^a = \{A_1(\zeta_0) = 2; \mathscr{E}_1\} \qquad \twoheadrightarrow \cdots \twoheadrightarrow \mathscr{E}_4^a = \left\{A_1(\zeta_0) = 2, \ldots, A_4(\mathscr{E}_3^a) = 3; \mathscr{E}_4\right\} \tag{8}$$

The sampling decision and the information set are nested in each other as t evolves. However, Theorem 1 in Peng et al. (2018b) shows that the sampling allocation policy would not affect the Bayesian structure under a canonical assumption in R&S, i.e., the replications $(X_{1,\ell}, \ldots, X_{k,\ell})$, $\ell \in \mathbb{Z}^+$, are independent, while dependence between different alternatives in sampling distribution Q is allowed. From Peng et al. (2018b), the optimal A&S policy $(\mathscr{A}_n^*, \mathscr{S}^*)$ satisfies the following Bellman equation:

$$V_n(\mathscr{E}_n) \doteq V_n(\mathscr{E}_n; i)|_{i=\mathscr{S}^*(\mathscr{E}_n)}, \tag{9}$$

where $V_n(\mathscr{E}_n; i) \doteq \mathbb{E}[\mathbf{1}\{i = \langle 1 \rangle\}|\mathscr{E}_n]$, and

$$\mathscr{S}^*(\mathscr{E}_n) = \arg\max_{i=1,\ldots,k} V_n(\mathscr{E}_n; i),$$

and for $0 \le t < n$,

$$V_t(\mathscr{E}_t) \doteq V_t(\mathscr{E}_t; i)|_{i=A_{t+1}^*(\mathscr{E}_t)}, \tag{10}$$

where $V_t(\mathscr{E}_t; i) \doteq \mathbb{E}[V_{t+1}(\mathscr{E}_t, X_{i,t+1})|\mathscr{E}_t]$, and

$$A_{t+1}^*(\mathscr{E}_t) = \arg\max_{i=1,\ldots,k} V_t(\mathscr{E}_t; i) \ .$$

Although R&S shares many similarities with the multi-armed bandit (MAB) problem (Auer 2003), they also have following differences:

(i) In standard MAB problems, the rewards are collected at all steps, whereas in R&S, the reward is collected at the end.

(ii) In standard MAB problems, the reward for pulling one arm only depends on the state of that arm, whereas in R&S, the reward for selecting one alternative depends on the states of k alternatives.

The optimal selection policy can be rewritten as

$$\mathscr{S}^*(\mathscr{E}_n) = \arg\max_{i=1,\ldots,k} \mathbb{P}(\mu_i - \mu_j > 0, \ j \neq i|\mathscr{E}_n) \ .$$

The commonly assumed selection policy that picks the alternative with the largest posterior mean would not be necessarily the optimal selection policy, which is illustrated by a simple example from Peng et al. (2016). In Figure 1, assume μ_1, μ_2, and μ_3 are three independently distributed Bernoulli random variables. where the probabilities given are the posterior probabilities. Since $\mathbb{E}[\mu_1] = -1/4$, $\mathbb{E}[\mu_2] = \mathbb{E}[\mu_3] = 0$, and $Var(\mu_1) = 7\frac{9}{16}$, $Var(\mu_2) = 4$,

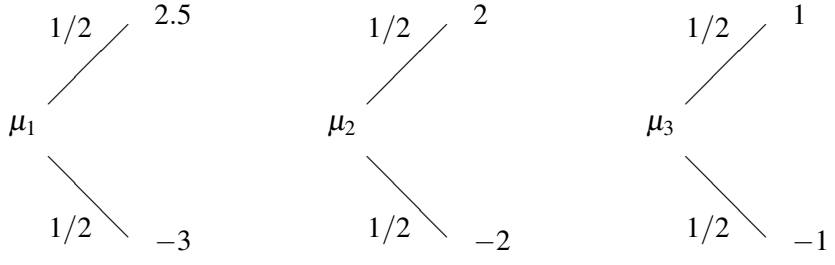

Figure 1: Illustration of the selection policy with three independent Bernoulli distributed random variables.

$Var(\mu_3) = 1$, although μ_1 has the smallest mean, it has the largest variance. By simple calculation, we have the following:

$$\mathbb{P}(\mu_1 > \mu_2, \mu_1 > \mu_3) = \frac{1}{2},\ \mathbb{P}(\mu_2 > \mu_1, \mu_2 > \mu_3) = \mathbb{P}(\mu_3 > \mu_1, \mu_3 > \mu_2) = \frac{1}{4}\ .$$

Therefore, the first alternative has the largest probability of being the best. The intuition is that if the first alternative takes the larger value in its value space, then it is the largest regardless of the realizations of the other two. A certain induced correlation affects the optimal selection policy (Peng et al. 2016).

3.2 Dynamic Allocation Policies

We revisit two well-known sampling allocation procedures, KG and EI, that solve one-step optimizations. The sampling allocation policy in KG is

$$\bar{A}_{t+1}(\mathscr{E}_t) \doteq \arg\max_{i=1,\ldots,k} \mathbb{E}\left[\mathbb{E}\left[\max\left(\mu_i^{(t+1)}, \max_{j\neq i}\mu_j^{(t)}\right)\middle|\mathscr{E}_t, X_{i,t_i+1}\right]\middle|\mathscr{E}_t\right],$$

and the sampling allocation policy in EI is

$$\bar{\bar{A}}_{t+1}(\mathscr{E}_t) \doteq \arg\max_{i=1,\ldots,k} \mathbb{E}\left[\max\left(\mu_i - \max_{j\neq i}\mu_j^{(t)}, 0\right)\middle|\mathscr{E}_t\right]\ .$$

Both KG and EI are consistent, because every alternative will be sampled infinitely often as the simulation budget grows to infinity. However, the asymptotic sampling ratios of KG and EI do not achieve the asymptotically optimal sampling ratio defined by (3) or the OCBA sampling ratio that is an approximate solution of (3). From Ryzhov (2016), under the normal sampling distribution with known variance, the asymptotic sampling ratio of KG is

$$\lim_{n\to\infty}\frac{n_i}{n_j} = \frac{\sigma_i\delta_j'}{\sigma_j\delta_i'},\quad i,j = 1,\ldots,k,\quad a.s., \tag{11}$$

where $\delta_i' \doteq |\mu_i - \max_{j\neq i}\mu_j|$, $i = 1,\ldots,k$, and the asymptotic sampling ratio of EI is

$$\lim_{n\to\infty}\frac{n_i}{n_j} = \frac{\sigma_i^2\delta_j^2}{\sigma_j^2\delta_i^2},\quad i,j \neq \langle 1\rangle,\quad a.s. \tag{12}$$

$$\lim_{n\to\infty} n_i/n_{\langle 1\rangle} = 0,\quad i \neq \langle 1\rangle,\quad a.s. \tag{13}$$

In (11), the asymptotic sampling ratios of the best alternative and the second-best alternative have symmetric roles, whereas either the asymptotically optimal sampling ratio or OCBA typically allocates more replications to the best alternative. Thus, KG "shortchanges" the best alternative asymptotically. For EI, the ratio (12) is identical to (4) in the OCBA formulas but (13), which prescribes that the proportion of replications allocated to the non-optimal alternatives vanish asymptotically, differs from (5) in the OCBA formula. In terms of PCS, the asymptotic sampling ratio of EI is even worse than equal allocation (EA). Specifically, the PFS of EA converges at a rate e^{-cn}, $c > 0$, whereas the PFS of EI converges at $e^{-o(n)}$, where $o(n)$ denotes lower order than n. This point is dramatically illustrated by numerical examples in Peng and Fu (2017). An asymptotically optimal myopic allocation policy (AOMAP) in Peng and Fu (2017), a variant of EI, has been proved to sequentially achieve the sampling ratio of OCBA. Another variant of EI in Chen and Ryzhov (2017) sequentially achieves the asymptotically optimal sampling ratio in (3).

Recently, Peng et al. (2018b) propose a dynamic sampling procedure derived in an approximate dynamic programming (ADP) paradigm. The idea of ADP is to find a suitable value function approximation (VFA) for Bellman equations (9) and (10). A simple treatment is to fix the selection policy to select the alternative with the largest posterior mean, and look one step ahead for allocating the next replication. Then, VFA becomes an integral of the standard normal density over a region encompassed by $k-1$ hyperplanes. In the case with three alternatives, the region is the shaded area in Figure 2 from Peng et al. (2018b). Since the standard normal density decays at an

exponential rate with respect to the distance from the origin, the VFA can be further simplified as the integral over the area of the largest inscribed ball encompassed by the hyperplanes, i.e., the circle in Figure 2 for the case of three alternatives. Owing to symmetry, maximizing the integral over the inscribed ball is equivalent to maximizing the size of the circle, which yields an analytical form for its solution.

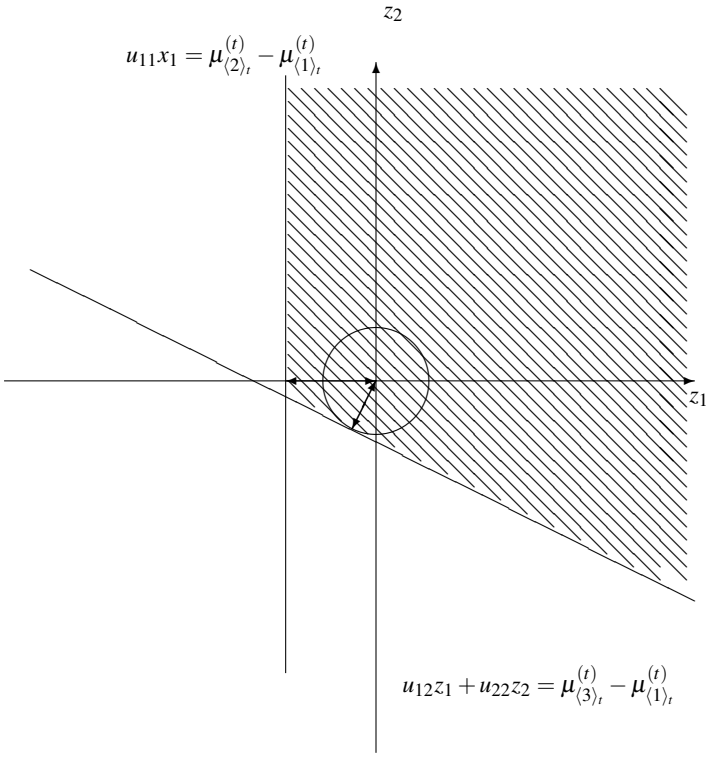

Figure 2: Area of integration for approximation is the circle, which captures the dominant values of integrand $\exp(-(z_1^2+z_2^2)/2)$.

With the approximations described above, an approximately optimal allocation policy (AOAP) in Peng et al. (2018b) is given by

$$\widehat{A}_{t+1}(\mathscr{E}_t) = \arg\max_{i=1,\ldots,k} \widehat{V}_t(\mathscr{E}_t; i), \tag{14}$$

where

$$\widehat{V}_t(\mathscr{E}_t; \langle 1\rangle_t) \doteq \min_{j\neq\langle 1\rangle_t} \frac{\left(\mu^{(t)}_{\langle 1\rangle_t} - \mu^{(t)}_j\right)^2}{\left(\sigma^{(t+1)}_{\langle 1\rangle_t}\right)^2 + \left(\sigma^{(t)}_j\right)^2},$$

$$\widehat{V}_t(\mathscr{E}_t; j) \doteq \min\left\{ \frac{\left(\mu^{(t)}_{\langle 1\rangle_t} - \mu^{(t)}_j\right)^2}{\left(\sigma^{(t)}_{\langle 1\rangle_t}\right)^2 + \left(\sigma^{(t+1)}_j\right)^2}, \ \min_{\ell\neq 1,j} \frac{\left(\mu^{(t)}_{\langle 1\rangle_t} - \mu^{(t)}_\ell\right)^2}{\left(\sigma^{(t)}_{\langle 1\rangle_t}\right)^2 + \left(\sigma^{(t)}_\ell\right)^2} \right\}, \quad j \neq \langle 1\rangle_t,$$

where $\langle 1\rangle_t \doteq \max_{i=1,\ldots,k} \mu^{(t)}_i$. Notice that AOAP has an analytical form that reflects a similar (posterior) mean-variance tradeoff as the OCBA formula. Not only is AOAP consistent but it also sequentially achieves the asymptotically optimal sampling ratio in (3). AOAP is based on a single-feature VFA. More features can be introduced to better approximate the true value function, e.g., Peng et al. (2018b) provide a two-feature VFA that includes both the information of (posterior) mean-variance and the information of induced correlations to avoid non-monotonicity of PCS in a certain low-confidence scenario (Peng et al. 2015), which usually occurs when the computing budget is low.

3.3 Rollout Policy

The A&S policy can also be viewed as a special partially observable Markov decision process (POMDP) with the state variable following a fixed distribution rather than a Markov process. We introduce a rollout policy that can be well integrated with the existing sampling procedures (Bertsekas and Castanon 1999). Given a base A&S policy $(\mathscr{A}_n^\pi, \mathscr{S}^\pi)$, e.g., AOAP as the sampling allocation policy and selecting the largest posterior mean as the selection policy, we can obtain a value function:

$$V_t^\pi(\mathscr{E}_t; i) \doteq \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{\mathscr{S}^\pi(\mathscr{E}_n^\pi) = \langle 1\rangle\} | \mathscr{E}_t, X_{i,t_i+1}\right] | \mathscr{E}_t\right],$$

where $\mathscr{E}_n^\pi$ is the information flow generated by $\mathscr{A}_n^\pi$ from step $t+1$ to step n. A rollout policy is

$$\widetilde{A}_{t+1}(\mathscr{E}_t) \doteq \arg\max_{i=1,\ldots,k} V_t^\pi(\mathscr{E}_t; i),$$

which is guaranteed to be at least as good as the base policy π. In general, it is difficult to have an analytical form of $V_t^\pi(\mathscr{E}_t; i)$. From Peng et al. (2018b), for discrete sampling distribution, the size of the state space grows exponentially with the numbers of alternatives and possible outcomes, and grows polynomially with the number of allocated replications. Thus, numerical calculation would be computationally infeasible for a problem with a moderate sizes of alternatives, outcomes, and replications.

Monte Carlo simulation could be a computationally feasible choice. We can generate a large number of sample paths (particles), and use particle filtering to iteratively update the posterior measure (Doucet 2001). R&S only requires a special case of particle filtering where the particles do not mutate to update the posterior distribution as follows:

$$\widehat{F}_i(\cdot|\mathscr{E}_t) = \frac{1}{M}\sum_{j=1}^{M} \mathbf{1}_{\theta_{i,j}^{(t)}}(\cdot) \to \widehat{F}_i(\cdot|\mathscr{E}_{t+1}) = \frac{1}{M}\sum_{j=1}^{M} \mathbf{1}_{\theta_{i,j}^{(t+1)}}(\cdot),$$

where M is the number of particles, $\theta_{i,j}^{(t)}$ is the jth particle for θ_i at step t, and $\mathbf{1}_x(\cdot)$ is the delta-measure with mass on x. The particles $\theta_{i,j}^{(t+1)}$, $j = 1,\ldots,M$, are resampled from $\theta_{i,j}^{(t)}$, $j = 1,\ldots,M$, with weights

$$w_{i,j} \doteq \frac{q_i\left(X_{i,t}; \theta_{i,j}^{(t)}\right)}{\sum_{\ell=1}^{M} q_i\left(X_{i,t}; \theta_{i,\ell}^{(t)}\right)}, \qquad j = 1,\ldots,M,$$

where $q_i(\cdot)$ is the density for the sampling distribution of the ith alternative, $i = 1,\ldots,k$.

We can also use the parallel rollout policy (Chang et al. 2004). Suppose $\Pi = \{\pi_1,\ldots,\pi_d\}$ is a set of base A&S policies, e.g., KG, EI, OCBA, and AOAP as base sampling allocation policies. Then, we can obtain the following value function:

$$V_t^\Pi(\mathscr{E}_t; i) \doteq \max_{j=1,\ldots,d} V_t^{\pi_j}(\mathscr{E}_t; i),$$

which in turn leads to the following parallel rollout policy:

$$\widetilde{\widetilde{A}}_{t+1}(\mathscr{E}_t) \doteq \arg\max_{i=1,\ldots,k} V_t^\Pi(\mathscr{E}_t; i)\ .$$

The parallel rollout policy is guaranteed to be at least as good as the best base policy in Π.

4 SIMULATION EXPERIMENTS

We provide numerical evidence that the static optimization formulation is inadequate to capture the dynamic decisions in R&S by showing that even with perfect information on the value of the parameters, neither optimal large deviations (OLD) that uses the sampling ratio given by (3) necessarily achieves a good performance nor does the static optimal sampling allocation (SOP) given by (1). Then, we show that many existing sampling allocation procedures do not

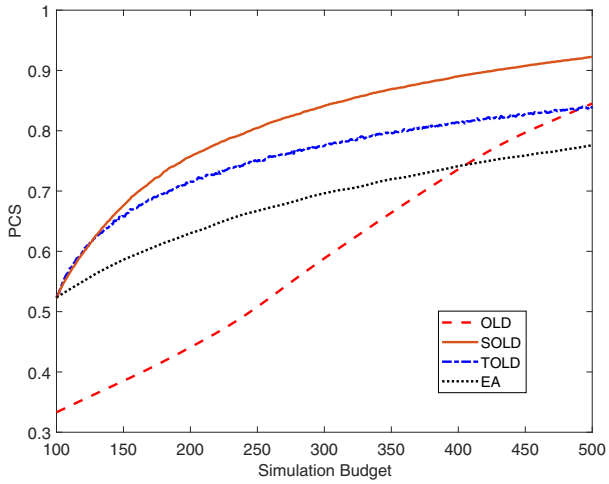

Figure 3: 10 alternatives with $\mu_i = 10 - i, \sigma_i = 6$, $i = 1, \ldots, 10$; PCS estimated by 10^5 macro-experiments.

capture well induced correlations.

Example 1. 10 alternatives following independent normal sampling distributions: $X_{i,\ell} \sim N(10-i, 6^2)$, $\ell \in \mathbb{Z}^+$, $i = 1, \ldots, k$. The total simulation budget is 500. We compare four different sampling allocation polices: OLD, which allocates replications according to sampling ratio in (3) given the true parameters; sequential OLD (SOLD), which uses 100 initial replications equally allocated to 10 alternatives to estimate means and variances and then allocates the rest of the replications one by one using a "most starving" rule that minimizes the gap between the number of allocated replication to each alternative and the number of replications prescribed by (3) with unknown parameters sequentially updated (Chen and Lee 2011); two-stage OLD (TOLD), which has the same first stage with SOLD and then uses the sampling ratio determined by (3) to allocate the rest of the replications in one shot; and, finally, EA. The optimization in (3) can be solved efficiently by a one-dimensional nonlinear convex optimization procedure (Peng et al. 2013).

From Figure 3, we can see that OLD is even worse than EA before the simulation budget reaches 400, and it only catches up with TOLD when the simulation budget reaches 500. SOLD has the best performance throughout. The observations above indicate that the asymptotic optimal sampling ratio per se does not achieve a good performance when the simulation budget is not large enough, whereas sequentially implementing it in a proper way could lead to a superior performance.

Example 2. 3 alternatives, where the first alternative is deterministic with $\mu_1 = 0$, and $X_{i,\ell} \sim N(-0.4, 3^2)$, $\ell \in \mathbb{Z}^+$, $i = 2, 3$. Then, PCS in (1) becomes

$$\mathbb{P}\left(\bar{X}_2^{(n)} < 0, \bar{X}_3^{(n)} < 0 \middle| \theta\right) = \Phi\left(\frac{0.4}{3n_2}\right)\Phi\left(\frac{0.4}{3n_3}\right),$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution. By symmetry, we know that the SOP in (1) is $n_1 = 0$ and $n_2 = n_3 = n/2$. From Figure 4, we can see that SOP is better than EA, which wastes $1/3$ of replications on the deterministic alternative, whereas it is worse than TOLD before the simulation budget reaches about 350; SOLD performs best throughout. Similar numerical observations that SOP and OCBA with perfect information do not perform as well as a sequential OCBA can be found in Chen et al. (2006).

Example 3. SOLD achieves a good performance in many classic R&S problems, because the (posterior) mean-variance tradeoff happens to lead to a desirable A&S policy in certain scenarios. However, this is not always

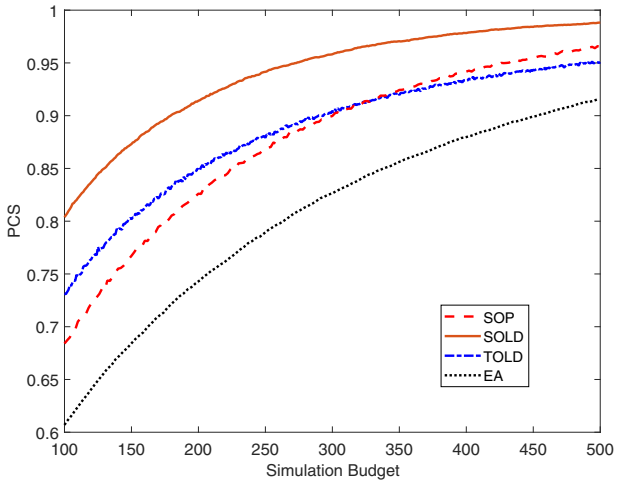

Figure 4: Three alternatives with $\mu_1 = 0$, $\mu_2 = \mu_3 = -0.4$, and $\sigma_1 = 0$, $\sigma_2 = \sigma_3 = 3$; PCS estimated by 10^5 macro-experiments.

the case. Recently, Shin et al. (2017) report the poor finite-time performance of sequentially implementing the asymptotically optimal sampling ratio for selecting the best quantile under some particular distributions. In addition, the (posterior) mean-variance tradeoff could lead to misleading results in a certain low-confidence scenario that is qualitatively described by three characteristics: the differences between the means of competing designs are small, the variances are large, and the simulation budget is small.

For this example, there are three alternatives with $\mu_1 = 0.001$, $\mu_2 = \mu_3 = 0$, and $\sigma_1 = \sqrt{2}$, $\sigma_2 = \sigma_3 = 1$, which falls into the low-confidence scenario with a total simulation budget of 60. We test seven sampling allocation policies: OCBA implemented sequentially by the "most starving" rule (Chen and Lee 2011); sequential EVI with 0-1 loss function (Chick et al. 2010); KG with uninformative prior (Frazier et al. 2008); PTV, for which the number of allocated replications to each alternative is proportional to its sample variance, implemented sequentially by the "most starving" rule; EA sequentially from the first to the last alternative in a cyclical manner; SOLD; EI with uninformative prior (Jones et al. 1998); AOAP in (14) with uninformative prior. In the first-stage, 30 initial replications are equally allocated to three alternatives to estimate unknown parameters or construct the uninformative prior, and the remaining 30 replications are allocated according to different sampling allocation procedures.

From Figure 5, we can see that for all but EA, PCS decreases with the number of allocated replications up through a simulation budget of 45. SOLD has the worst performance, and the trajectories of OCBA and EVI are indistinguishable. From the trajectory of EA, we can see that each time a replication is allocated to the first (best) alternative, PCS decreases, and when a replication is allocated to the second or third alternative, PCS increases. The decreasing of PCS is caused by ignoring the induced correlations (Peng et al. 2015). This problem has been addressed by a gradient-based myopic allocation policy (G-MAP) that takes account of the induced correlations in Peng et al. (2018a) under a Bayesian framework, as well as an offline learning scheme under the ADP paradigm in Peng et al. (2018b). Under the A&S umbrella, the final value function increases with respect to the number of simulation replications if the optimal selection is used, by noticing that

$$
\begin{aligned}
&\mathbb{E}\left[\mathbf{1}\{\mathscr{S}^*(\mathscr{E}_{n+1}) = \langle 1\rangle\}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{\mathscr{S}^*(\mathscr{E}_{n+1}) = \langle 1\rangle\}|\mathscr{E}_{n+1}\right]\right] \\
&= \mathbb{E}\left[\max_{i=1,\ldots,k} \mathbb{E}\left[\mathbf{1}\{i = \langle 1\rangle\}|\mathscr{E}_{n+1}\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\max_{i=1,\ldots,k} \mathbb{E}\left[\mathbf{1}\{i = \langle 1\rangle\}|\mathscr{E}_{n+1}\right]|\mathscr{E}_n\right]\right] \\
&\geq \mathbb{E}\left[\max_{i=1,\ldots,k} \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{i = \langle 1\rangle\}|\mathscr{E}_{n+1}\right]|\mathscr{E}_n\right]\right] = \mathbb{E}\left[\max_{i=1,\ldots,k} \mathbb{E}\left[\mathbf{1}\{i = \langle 1\rangle\}|\mathscr{E}_n\right]\right] = \mathbb{E}\left[\mathbf{1}\{\mathscr{S}^*(\mathscr{E}_n) = \langle 1\rangle\}\right],
\end{aligned}
$$

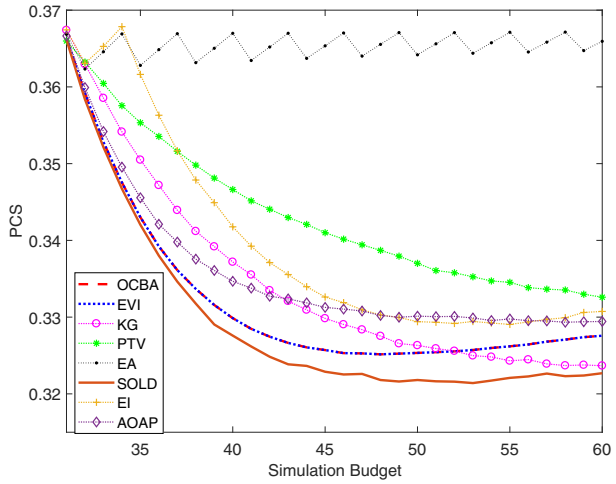

Figure 5: Three alternatives with $\mu_1 = 0.001$, $\mu_2 = \mu_3 = 0$, and $\sigma_1 = \sqrt{2}$, $\sigma_2 = \sigma_3 = 1$; PCS estimated by 10^6 macro-experiments.

where the inequality is due to the Jensen's inequality, and the equalities can be obtained by the definition of the optimal selection policy and the property of law of total expectation.

5 CONCLUSION

We reviewed several static optimization and dynamic optimization formulations for R&S. Under the umbrella of an A&S policy, we revisit several sampling allocation procedures. KG and EI do not have desirable asymptotic sampling ratios, whereas AOAP derived in an ADP paradigm (Peng et al. 2018b) achieves the asymptotically optimal sampling ratio. Simulation results demonstrate that the static optimization formulation is inadequate to capture dynamic sampling allocation decisions, and many existing sampling allocation procedures do not capture well certain induced correlations.

Under the A&S umbrella, we introduce a rollout policy that can be well integrated with existing sampling procedures with a performance improvement guarantee. Future research may leverage some artificial intelligence tools, e.g., neural networks, to better approximate the value function in the Bellman equations (9) and (10) using offline training. Many R&S related problems such as subset selection (Chen et al. 2008), feasibility determination (Xie and Frazier 2013), minimizing expected opportunity cost (Gao et al. 2017), and targeting and selection (Ryzhov 2018), where the only difference lies in the final reward in the Bellman equation (9), could be treated under a single umbrella.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grants 71720107003, 71571048, 71690232, and 61603321, and by the U.S. National Science Foundation under Awards ECCS-1462409, ECCS-1462409, CMMI-1462787, CMMI-1434419, CCF-1422658, and CCF-1712788.

REFERENCES

Auer, P. 2003. "Using Confidence Bounds for Exploitation-Exploration Trade-offs". *The Journal of Machine Learning Research* 3:397–422.

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., and D. A. Castanon. 1999. "Rollout Algorithms for Stochastic Scheduling Problems". *Journal of Heuristics* 5(1):89–108.

Chang, H. S., R. Givan, and E. K. P. Chong. 2004. "Parallel Rollout for Online Solution of Partially Observable Markov Decision Processes". *Discrete Event Dynamic Systems: Theory and Applications* 14(3):309–341.

Chen, C.-H., 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.-H., D. He, M. C. 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*. 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: Theory and Applications* 10(3):251–270.

Chen, Y., and I. O. Ryzhov. 2017. "Rate-Optimality of the Complete Expected Improvement Criterion". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, et al., 2173–2182. Piscataway, New Jersey: IEEE.

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.

Chick, S. E., and K. Inoue. 2001. "New Two-Stage and Sequential Procedures for Selecting the Best Simulated System". *Operations Research* 49(5):732–743.

DeGroot, M. H. 2005. *Optimal Statistical Decisions*. Wiley-Interscience.

Doucet, A. 2001. *Sequential Monte Carlo Methods*. Wiley Online Library.

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.

Frazier, P. I., W. B. Powell, and S. Dayanik. 2008. "A Knowledge-Gradient Policy for Sequential Information Collection". *SIAM Journal on Control and Optimization* 47(5):2410–2439.

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 R. G. Ingalls et al., 577–585. Piscataway, New Jersey: IEEE.

Gupta, S. S., and K. J. Miescke. 1996. "Bayesian Look Ahead One-Stage Sampling Allocations for Selection of the Best Population". *Journal of Statistical Planning and Inference* 54(2):229–244.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization* 13(4):455–492.

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.

Luo, J., L. J. Hong, B. L. Nelson, and Y. Wu. 2015. "Fully Sequential Procedures for Large-Scale Ranking-and-Selection Problems in Parallel Computing Environments". *Operations Research* 63(5):1177–1194.

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.-H. Chen, M. C. Fu, and J.-Q. Hu. 2013. "Efficient Simulation Resource Sharing and Allocation for Selecting the Best". *IEEE Transactions on Automatic Control* 58(4):1017–1023.

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 L. Yilmaz et al., 3678–3689. Piscataway, New Jersey: IEEE.

Peng, Y., C.-H. Chen, M. C. Fu, and J.-Q. 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. 2018a. "Gradient-Based Myopic Allocation Policy: An Efficient Sampling Procedure in a Low-Confidence Scenario". *IEEE Transaction Automatic Control (Early Access) https://ieeexplore.ieee.org/document/8118094/*.

Peng, Y., E. K. P. Chong, C.-H. Chen, and M. C. Fu. 2018b. "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., 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.

Ryzhov, I. O. 2016. "On the Convergence Rates of Expected Improvement Methods". *Operations Research* 64(6):1515–1528.

Ryzhov, I. O. 2018. "The local time method for targeting and selection". *forthcoming in Operations Research*.

Shin, D., M. Broadie, and A. Zeevi. 2017. "Tractable Sampling Strategies for Quantile-Based Ordinal Optimization". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, et al., 847–858. Piscataway, New Jersey: IEEE.

Xie, J., and P. I. Frazier. 2013. "Sequential Bayes-Optimal Policies for Multiple Comparisons with a Known Standard". *Operations Research* 61(5):1174–1189.

AUTHOR BIOGRAPHIES

YIJIE PENG is an Assistant Professor in the Department of Industrial Engineering and Management at Peking University. He received his the Ph.D. degree in management science from Fudan University in 2014, and worked at Fudan University, University of Maryland, and George Mason University as a research fellow before joining Peking University. He is a member of of INFORMS and IEEE. His research interests are in ranking and selection and sensitivity analysis in the field simulation optimization, with applications in data analytics, healthcare, and financial engineering. His email address is pengyijie@pku.edu.cn.

CHUN-HUNG CHEN is a Professor of Systems Engineering and Operations Research at George Mason University. He received his Ph.D. degree from Harvard University in 1994. He served as Co-Editor of the *Proceedings of the 2002 Winter Simulation Conference* and Program Co-Chair for 2007 INFORMS Simulation Society Workshop. He has served on the editorial boards of *IEEE Transactions on Automatic Control*, *IEEE Transactions on Automation Science and Engineering*, *IIE Transactions*, *Journal of Simulation Modeling Practice and Theory*, and *International Journal of Simulation and Process Modeling*. He is a Fellow of IEEE. His email address is cchen9@gmu.edu.

EDWIN K. P. CHONG is a Professor of Electrical and Computer Engineering and Professor of Mathematics at Colorado State University. He received his Ph.D. degree from Princeton University in 1991. He was the General Chair for the 2011 Joint 50th IEEE Conference on Decision and Control and European Control Conference. He has served on the editorial boards of *IEEE Transactions on Automatic Control*, *Computer Networks*, *Journal of Control Science and Engineering*, and *IEEE Expert Now*. He has served as President of the IEEE Control Systems Society and is a Fellow of IEEE. His e-mail addresses is edwin.chong@colostate.edu.

MICHAEL C. FU holds the Smith Chair of Management Science in the Robert H. Smith School of Business, with a joint appointment in the Institute for Systems Research, A. James Clark School of Engineering, University of Maryland, College Park. He has a Ph.D. in applied math from Harvard and degrees in math and EECS from MIT. He served as WSC2011 Program Chair, NSF Operations Research Program Director, *Management Science* Stochastic Models and Simulation Department Editor, and *Operations Research* Simulation Area Editor. He is a Fellow of INFORMS and IEEE. His email address is mfu@umd.edu.