Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

SEQUENTIAL SAMPLING FOR BAYESIAN ROBUST RANKING AND SELECTION

Xiaowei Zhang
Liang Ding

Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong, CHINA

ABSTRACT

We consider a Bayesian ranking and selection problem in the presence of input distribution uncertainty. The distribution uncertainty is treated from a robust perspective. A naive extension of the knowledge gradient (KG) policy fails to converge in the new robust setting. We propose several stationary policies that extend KG in various aspects. Numerical experiments show that the proposed policies have excellent performance in terms of both probability of correction selection and normalized opportunity cost.

1 INTRODUCTION

Simulation is a general-purpose tool for facilitating decision-making related to complex stochastic systems. In particular, a decision-maker may need to choose one from a finite collection of alternatives. The alternatives could be investment strategies, surgery schedules, supply chain configurations, warehouse layouts, etc. However, the value of each alternative is typically unknown and must be measured via simulation. This is known as the ranking and selection (R&S) problem (Bechhofer, Santner, and Goldsman 1995). Simulating a complex system could be computationally expensive. Given a computational budget in terms of measurements or sampling opportunities, the goal of the R&S problem is to allocate the computational budget in an efficient way such that as much information as possible about the true value of each alternative can be obtained.

When constructing simulation models for the alternatives, a decision-maker often encounters the challenge of choosing proper input distributions, if multiple candidate distributions can fit the input data reasonably well. One approach to address the issue is Bayesian model averaging (BMA); see Hoeting et al. (1999) for a general introduction on the subject and Chick (2001) for its application in the context of stochastic simulation. With BMA one may specify one's prior belief on the probability that a candidate input distribution is the correct one. Then the value of the simulation model in the presence of input distribution uncertainty is essentially the average value over the candidate input distributions weighted by the prior belief. BMA may be appropriate for a decision-maker who is risk-neutral with respect to the input distribution.

Another approach, motivated by robust optimization (Ben-Tal and Nemirovski 2002, Ben-Tal, El Ghaoui, and Nemirovski 2009), is more appealing to a risk-averse decision-maker especially when implementing an alternative is highly costly. Rather than averaging over candidate input distributions, this approach uses the "worse-case" scenario among them to represent the value of an alternative. Fan, Hong, and Zhang (2013) adopts this robust perspective to address the R&S problem in the presence of input distribution uncertainty. In particular, they model the input distribution uncertainty by a finite set of possible distributions, and the best alternative in the R&S problem is then defined as the one having the best worse-case performance. We follow this modeling perspective in the present paper and call it a robust R&S problem.

In contrast to the frequentist approach in Fan, Hong, and Zhang (2013), we take a Bayesian view on the true value of a *system*, which denotes an alternative under a candidate input distribution. Beginning

with a prior belief about the systems, our belief is updated after one or several measurements. When the computational budget is exhausted, we use the final belief to rank the systems and select the optimal one. This Bayesian framework for R&S has been widely used to develop sequential sampling policies, including the optimal computing budget allocation (OCBA) policy (Chen, Dai, and Chen 1996, Chen et al. 2000, Chen, Chen, and Yücesan 2000, and He, Chick, and Chen 2007), the knowledge gradient (KG) policy (Gupta and Miescke 1996, Frazier, Powell, and Dayanik 2008, and Frazier, Powell, and Dayanik 2009), and the expected value of information (EVI) approach (Chick and Inoue 2001a, Chick and Inoue 2001b, and Chick, Branke, and Schmidt 2010).

Specifically, we model the unknown mean performance measures of all the systems using a multivariate normal random variable. Following the setup of the KG policy, we assume that the prior belief about the systems is multivariate normal and that sampling a system produces unbiased random output with *known* variance. This implies that the posterior belief is multivariate normal as well. Further, we assume independence among alternatives but explore correlations among the performance measures of an alternative under different input distributions. By doing so, sampling an alternative under one input distribution helps us learn about the same alternative under different input distributions, but provides no useful information about other alternatives. Such an independence assumption can be relaxed but would significantly increase computational expense due to lack of sparsity in the large covariance matrix of the systems.

We formulate a sequential sampling scheme for the robust R&S problem as a dynamic program that aims to minimize the terminal reward, which is of a "minimax" form due to the robust perspective and is collected after a given number of sampling opportunities are executed. This dynamic program is an extension of the one analyzed in Frazier, Powell, and Dayanik (2008) and Frazier, Powell, and Dayanik (2009) for the KG policy. We focus on sampling policies that are stationary in time because they can be computed much more easily. However, a naive extension of the KG policy to the new robust setting fails to converge, in the sense of sampling each each system infinitely often if an infinite computational budget is available. Hence, we propose stationary policies that extend the KG policy in nontrivial ways by revising the objective of the dynamic program to another related one. Numerical experiments demonstrate that the proposed policies have excellent performance with regard to identifying the best alternative in terms of both probability of correct selection and normalized opportunity cost.

The rest of the paper is organized as follows. Section 2 introduces our Bayesian framework for robust R&S and formulates the sequential sampling decisions as a dynamic program. Section 3 proposes several stationary sampling policies for solving the dynamic program. Section 4 presents numerical experiments and Section 5 concludes.

2 PROBLEM FORMULATION

In the setting of stochastic simulation, the performance measure of a simulation model is generally expressed as a function g of the decision variable s and the environmental variable ξ, where the former is controllable and deterministic whereas the latter is uncontrollable and random. The mean performance that we attempt to estimate via simulation is then

$$\mathbb{E}_P[g(s,\xi)],$$

where the expectation is taken with respect to ξ having probability distribution P. For example, in a queueing simulation s may be the number of servers, ξ may be the collection of interarrival times and service times, while g could be the steady-state waiting time.

Suppose that we have a set of M distinct possible decisions or alternatives $\mathscr{S} = \{s_1, \ldots, s_M\}$ and a set of K distinct possible distributions $\mathscr{P} = \{P_1, \ldots, P_K\}$. For a given distribution P, we define the optimal decision to be the one that delivers the smallest mean performance, i.e.,

$$\min_{s\in\mathscr{S}} \mathbb{E}_P[g(s,\xi)].$$

In light of the uncertainty about the distribution P, when assessing the decisions we adopt a robust perspective and base the comparison on the worst-case performance of a decision over the set $\mathscr{P}$. In particular, we

are interested in the following optimization problem,

$$\min_{s\in\mathscr{S}}\max_{P\in\mathscr{P}}\mathbb{E}_P[g(s,\xi)]. \tag{1}$$

2.1 Bayesian Formulation

To facilitate the presentation, we refer to the pair (s_i,P_j) as "system (i,j)" and let $\theta_{i,j}=\mathbb{E}_{P_j}[g(s_i,\xi)]$, $i=1,\ldots,M$, $j=1,\ldots,K$. We let θ denote the matrix formed by the $\theta_{i,j}$'s and $\theta_{i:}^\intercal$ denote its i^{th} row, i.e., $(\theta_{i,1},\ldots,\theta_{i,K})$. Suppose that samples from system (i,j) are independent and have a normal distribution with *unknown* mean $\theta_{i,j}$ and *known* variance $\delta_{i,j}^2$. (In general, $g(s_i,\xi)$ is not normally distributed. Nevertheless, the sample average of a sufficiently large number of its multiple independent replications has approximately a normal distribution by the law of large numbers. We can view such a sample average as "one sample".)

Applying a Bayesian approach, we assume that the prior belief about θ is a multivariate normal distribution with mean μ^0 and covariance Σ^0, i.e., $\theta\sim\mathcal{N}(\mu^0,\Sigma^0)$, where Σ^0 is indexed by $((i,j),(i',j'))$, $1\le i,i'\le M$, $1\le j,j'\le K$. Further, we assume that the prior belief about θ is such that $\theta_{1:},\ldots,\theta_{M:}$ are mutually independent.

Consider a sequence of N sampling decisions, $(x^0,y^0),(x^1,y^1),\ldots,(x^{N-1},y^{N-1})$. At each time $0\le n<N$, the sampling decision (x^n,y^n) selects a system from the set $\{(i,j):1\le i\le M,1\le j\le K\}$. Conditionally on the decision (x^n,y^n), the sample observation is $\hat{z}^{n+1}=\theta_{x^n,y^n}+\varepsilon^{n+1}$, where $\varepsilon^{n+1}\sim\mathcal{N}(0,\delta^2_{x^n,y^n})$ is the sampling error. We assume that the errors $\varepsilon^1,\ldots,\varepsilon^N$ are mutually independent and are independent of θ.

We define a filtration $\{\mathscr{F}^n:0\le n<N\}$, where $\mathscr{F}^n$ is the sigma-algebra generated by the samples observed and the decisions made by time n, namely, $(x^0,y^0),\hat{z}^1,\ldots,(x^{n-1},y^{n-1}),\hat{z}^n$. We use $\mathbb{E}_n[\cdot]$ to denote the conditional expectation $\mathbb{E}[\cdot|\mathscr{F}^n]$ and define $\mu^n:=\mathbb{E}_n[\theta]$ and $\Sigma^n:=\mathrm{Cov}[\theta|\mathscr{F}^n]$. By Bayes rule, the posterior distribution of θ conditionally on $\mathscr{F}^n$ is multivariate normal with mean μ^n and covariance Σ^n. Our uncertainty about θ decreases during the process of the sequential sampling. After all the N sampling decisions are executed, the decision-maker selects a system that attains $\min_i\max_j\mu^N_{i,j}$ in light of (1).

We now present the updating equations that stipulate how μ^{n+1} and Σ^{n+1} are expressed explicitly in terms of μ^n, Σ^n, (x^n,y^n), and $\hat{z}^{n+1}$. The independence assumption on $\theta_{x:}$ and $\theta_{x':}$ in the prior belief results in their independence in the posterior distribution, i.e.,

$$\Sigma^n_{x:,x':}=\mathbf{0},\quad\text{if } x\neq x',$$

for all $0\le n<N$, where $\Sigma^n_{x:,x':}$ denotes the covariance matrix of $\theta_{x:}$ and $\theta_{x':}$ conditionally on $\mathscr{F}^n$. Sampling system (x,y) provides no information about system (x',y') if $x'\neq x$. Adopting the calculation in Section 2.1 of Frazier, Powell, and Dayanik (2009) for each $\theta_{x:}$, $x=1,\ldots,M$, we find that

$$\mu^{n+1}_{x:}=\begin{cases}\mu^n_{x:}+\dfrac{\hat{z}^{n+1}-\mu^n_{x,y}}{\delta^2_{x,y}+\Sigma^n_{(x,y),(x,y)}}\Sigma^n_{x:,x:}e_y, & \text{if } x^n=x,\ y^n=y,\\ \mu^n_{x:}, & \text{if } x^n\neq x,\ y^n=y,\end{cases} \tag{2}$$

and

$$\Sigma^{n+1}_{x:,x:}=\begin{cases}\Sigma^n_{x:,x:}-\dfrac{\Sigma^n_{x:,x:}e_ye_y^\intercal\Sigma^n_{x:,x:}}{\delta^2_{x,y}+\Sigma^n_{(x,y),(x,y)}}, & \text{if } x^n=x,\ y^n=y,\\ \Sigma^n_{x:,x:}, & \text{if } x^n\neq x,\ y^n=y,\end{cases}$$

where e_y is a vector in $\mathbb{R}^K$ whose elements are all 0's except a single 1 at index y.

We now define a $\mathbb{R}^K$-valued function $\tilde{\sigma}$ as

$$\tilde{\sigma}(\Sigma,x,y):=\frac{\Sigma_{x:,x:}e_y}{\sqrt{\delta^2_{x,y}+\Sigma_{(x,y),(x,y)}}}, \tag{3}$$

and we define a random variable Z^{n+1} as

$$Z^{n+1} := \frac{\hat{z}^{n+1} - \mu^n_{x^n,y^n}}{\sqrt{\delta^2_{x^n,y^n} + \Sigma^n_{(x^n,y^n),(x^n,y^n)}}}.$$

Then Z^{n+1} is standard normal conditionally on $\mathscr{F}^n$, since

$$\mathrm{Var}[\hat{z}^{n+1} - \mu^n_{x^n,y^n} | \mathscr{F}^n] = \mathrm{Var}[\theta_{x^n,y^n} + \varepsilon^{n+1} | \mathscr{F}^n] = \delta^2_{x^n,y^n} + \Sigma^n_{x^n:,x^n:}.$$

It follows from (2) and (3) that

$$\mu^{n+1}_{x:} = \begin{cases} \mu^n_{x:} + \tilde{\sigma}(\Sigma^n, x^n, y^n) Z^{n+1}, & \text{if } x^n = x,\ y^n = y, \\ \mu^n_{x:}, & \text{if } x^n \neq x,\ y^n = y. \end{cases} \tag{4}$$

2.2 Dynamic Program

We suppose that the decision-maker makes the sampling decisions sequentially. In particular, the decision (x^n, y^n) is $\mathscr{F}^n$-measurable so that a sampling decision depends only on samples observed and decisions made in the past. Let Π denote the set of sampling policies that satisfy the above sequential requirement,

$$\Pi := \Big\{ \big((x^0,y^0),\ldots,(x^{N-1},y^{N-1})\big) : 1 \le x^n \le M, 1 \le y^n \le K, (x^n,y^n) \text{ is } \mathscr{F}^n\text{-measurable}, 0 \le n < N \Big\}.$$

We will use π to denote a generic element in Π and write $\mathbb{E}^\pi[\cdot]$ to indicate the expectation taken when the sampling policy is fixed to be π.

Our goal is to solve

$$\min_{\pi\in\Pi} \mathbb{E}^\pi \Big[\min_{1\le i\le M} \max_{1\le j\le K} \mu^N_{i,j} \Big] \tag{5}$$

with a dynamic programming approach. Clearly, μ^n takes its values in $\mathbb{R}^{M\times K}$ while Σ^n is in the space of positive semidefinite matrices of size $(MK)\times(MK)$. We define $\mathbb{S}$, the state space of $S^n := (\mu^n, \Sigma^n)$, to be the cross-product of these two spaces. Define the value function $V^n : \mathbb{S} \mapsto \mathbb{R}$

$$V^n(s) := \min_{\pi\in\Pi} \mathbb{E}^\pi \Big[\min_{1\le i\le M} \max_{1\le j\le K} \mu^N_{i,j} \Big| S^n = s \Big], \quad s \in \mathbb{S}.$$

Then, the terminal value function is given by

$$V^N(s) = \min_{1\le i\le M} \max_{1\le j\le K} \mu_{i,j}, \quad s = (\mu,\Sigma) \in \mathbb{S},$$

and our goal in (5) is to compute $V^0(s)$ for any $s \in \mathbb{S}$. The dynamic programming principle dictates that the value function $V^n(s)$, for any $0 \le n < N$, can be computed by recursively solving

$$V^n(s) = \min_{1\le x\le M, 1\le y\le K} \mathbb{E}\Big[V^{n+1}(S^{n+1}) \Big| S^n = s, (x^n,y^n) = (x,y) \Big].$$

Unfortunately, the above recursive equation has no analytical solution and a numerical solution is computationally infeasible due to the curse of dimensionality caused by the continuous nature of the state space. Consequently, we will focus on policies that are stationary in time. This is a treatment widely used in the literature including Frazier, Powell, and Dayanik (2009), Chick, Branke, and Schmidt (2010), and so forth.

3 STATIONARY POLICIES FOR SEQUENTIAL SAMPLING

For a policy π, we denote by $\mathcal{A}^{\pi,n}:\mathbb{S}\mapsto\{1,\ldots,M\}\times\{1,\ldots,K\}$ the decision function associated with π, i.e. $\mathcal{A}^{\pi,n}(S^n)=(x^n,y^n)$ almost surly under $\mathbb{P}^\pi$, the probability measure induced by π. We call a policy π *stationary* if $\mathcal{A}^{\pi,n}$ is independent of n, i.e. $\mathcal{A}^{\pi,0}=\mathcal{A}^{\pi,1}=\cdots=\mathcal{A}^{N-1}$ almost surly under $\mathbb{P}^\pi$. Moreover, we simply write $\mathcal{A}^\pi$ if π is stationary.

Note that we may write the terminal value $V^N(S^N)$ as a telescoping sequence,

$$\min_{1\le i\le M}\max_{1\le j\le K}\mu^N_{i,j}=V^N(S^N)=\left[V^N(S^N)-V^N(S^{N-1})\right]+\cdots+\left[V^N(S^{n+1})-V^N(S^n)\right]+V^N(S^n),\qquad(6)$$

which decompose the terminal value into the sum of a single-period reward $V^N(S^n)$ at time n and subsequent single-period rewards $V^N(S^k)-V^N(S^{k-1})$ at time $k=n+1,\ldots,N$. Now consider the stationary policy that aims to maximize the expected single-period reward and name it the *naive knowledge gradient* (NKG) policy. Its decision function is given by

$$\mathcal{A}^{\pi^{\mathrm{NKG}}}(s)=\mathop{\arg\min}_{1\le x\le M,1\le y\le K}\left\{\mathbb{E}_n\Big[\min_{1\le i\le M}\max_{1\le j\le K}\mu^{n+1}_{i,j}\big|S^n=s,(x^n,y^n)=(x,y)\Big]-\min_{1\le i\le M}\max_{1\le j\le K}\mu^n_{i,j}\right\},\quad s\in\mathbb{S}.$$

To compute the above decision function, the key step is compute the expectation inside the curly braces. Note that by (4), $\mu^{n+1}_{i,j}$ is a linear transform of the same standard normal random variable Z^{n+1} for all (i,j)'s. This expectation can be expressed in the form of $\sum_k\mathbb{E}([a+bZ)\mathbb{I}_{c_k\le Z<c_{k+1}}]$, for some constants a, b, and c_k's. The sequence of c_k's is in fact the change points of a piecewise linear function, formed by the minimum of the K maxima of linear functions that transform $\mu^n_{i,j}$ to $\mu^{n+1}_{i,j}$. These change points can be computed by a sweep line algorithm combined with a divide-and-conquer strategy; see Section 6.2.1 of Sharir and Agarwal (1995) for details of such an algorithm.

Note that if $K=1$, then NKG is reduced to KG. The name KG stems from the following observation: $\min_i\max_j\mu^{n+1}_{i,j}-\min_i\max_j\mu^n_{x,j}$ may be thought of as a gradient in some sense since it represents the incremental random value of the sampling decision (x,y) at time n. It is shown in Frazier, Powell, and Dayanik (2009) that KG is *convergent*, in the sense that it samples each system infinitely often if given infinite computational budget. Hence, a convergent policy can eventually identify the system that is truly the optimal given sufficient computational budget. However, NKG, as a naive extension of KG to the setting of $K\ge 2$, is not convergent in general.

Convergence of a policy on its own indicates little about efficiency of the policy in the finite sample case. For instance, the equal allocation policy which allocates the computational budget in a round-robin fashion equally among the systems guarantees that every system is sampled infinitely often if infinite computational budget is available, and thus it is convergent. But its performance in the finite sample case is not particularly satisfying. Nevertheless, convergence should be a desired feature of a good sampling policy as it ensures that the policy does not "stick" in a proper subset of the systems, in which case the other systems would not be sampled infinitely often and thus would never be learned perfectly even given infinite computational budget.

We demonstrate the non-convergence of NKG via the following somewhat artificial example. More realistic numerical results are given in Section 4.

Example 1 Let $M=K=2$. Suppose that every element of Σ^0 is 0 except $\Sigma^0_{(1,1),(1,1)}>0$. In other words, the prior belief about θ is such that $\theta_{1,1}$ is unknown, whereas $\theta_{1,2},\theta_{2,1},\theta_{2,2}$ are all known with certainty. The updating equation (4) implies that if $(x^0,y^0)=(1,1)$, then

$$\mu^1_{1,1}=\mu^0_{1,1}+\sigma Z,\quad\text{and}\quad\mu^1_{i,j}=\mu^0_{i,j},\,(i,j)\ne(1,1),$$

for some $\sigma>0$, where Z is a standard normal random variable; otherwise, $\mu^1_{i,j}=\mu^0_{i,j}$ for any (i,j).

Clearly, the expected single-period reward associated with the sampling decision (i,j) is 0 if $(i,j) \neq (1,1)$. With $(x^0, y^0) = (1,1)$, the same quantity becomes

$$\mathbb{E}\left[\max\left(\mu_{1,1}^0 + \sigma Z^1, \mu_{1,2}^0\right) \wedge \max(\mu_{2,1}^0, \mu_{2,2}^0)\right] - \min_i \max_j \mu_{i,j}^0. \tag{7}$$

Without loss of generality, set $\mu_{1,1}^0 = 0$. Consider the special case where

$$\mu_{1,2}^0 < 0 < \max(\mu_{2,1}^0, \mu_{2,2}^0).$$

It follows that $\min_i \max_j \mu_{i,j}^0 = 0$ and that (7) equals

$$a\mathbb{P}(\sigma Z < a) + b\mathbb{P}(\sigma Z > b) + \mathbb{E}[\sigma Z \mathbb{I}_{a \leq \sigma Z \leq b}], \tag{8}$$

where $a = \mu_{1,2}^0$ and $b = \max(\mu_{2,1}^0, \mu_{2,2}^0)$, since a and b are both constants. It is easy to show that (8) is positive if $a + b > 0$. Hence, the optimal decision is not to sample the unknown $\theta_{1,1}$ but to sample any of the known systems, in which case the state of the systems remains the same in all the subsequent time epochs. Consequently, if NKG is adopted, system (1,1) will never be sampled.

By contrast, if $K = 1$, then it is equivalent to set $a = -\infty$ in (8) and the expected single-period reward is always negative if the decision is to sample system $(1,1)$. So the same policy would encourage exploration of uncertainty rather than discourage it, thereby being convergent.

This example highlights the relative importance of the prior in our robust setting, since NKG may be convergent for some priors whereas not for others. Technically, the reason why NKG fails to converge in general is as follows. In Frazier, Powell, and Dayanik (2009), a critical step towards establishing the convergence of the KG policy is to show the monotonicity of the value function $V^n(s)$ in n. In particular, with $K = 1$ it can be shown that $V^{n+1}(s) \geq V^n(s)$ for any $s \in \mathbb{S}$, which essentially stems from Jensen's inequality $\mathbb{E}[\min_i \mu_i^N] \leq \min_i \mathbb{E}[\mu_i^N]$. However, with $K \geq 2$ we lose such monotonicity of $V^n(s)$ in n since $\mathbb{E}[\min_i \max_j \mu_{ij}^N]$can be either greater or smaller than $\min_i \max_j \mathbb{E}[\mu_{ij}^N]$.

3.1 Modified Objective

We now consider a different optimization problem that is somewhat related to (5)

$$\max_{\pi \in \Pi} \mathbb{E}^\pi \Big[\sum_{i=1}^M \max_{1 \leq j \leq K} \mu_{i,j}^N \Big]. \tag{9}$$

Define the value function $U^n : \mathbb{S} \mapsto \mathbb{R}$

$$U^n(s) \coloneqq \max_{\pi \in \Pi} \mathbb{E}^\pi \Big[\sum_{i=1}^M \max_{1 \leq j \leq K} \mu_{i,j}^N \Big| S^n = s \Big], \quad s \in \mathbb{S}.$$

Applying the same telescoping decomposition as (6) to $U^N(S^N)$, we may construct a stationary myopically optimal policy for the problem (9), which we refer to as *maximum knowledge gradient* (MKG) policy. Its decision function is defined by

$$\mathcal{A}^{\pi^{\text{MKG}}}(s) = \underset{1 \leq x \leq M, 1 \leq y \leq K}{\arg\max} \mathbb{E}_n \Big[U^N(S^{n+1}) - U^N(S^n) \Big| S^n = s, (x^n, y^n) = (x, y) \Big], \quad s \in \mathbb{S}.$$

Since we assume that $\theta_{x:}$ and $\theta_{x':}$ are independent in the prior distribution of θ, a sampling decision (x,y) has no impact on system (x',y') if $x' \neq x$. In particular, by (4),

$$\begin{aligned}
& \mathbb{E}_n\Big[U^N(S^{n+1}) - U^N(S^n)\big|S^n = s,(x^n,y^n) = (x,y)\Big] \\
= \;& \mathbb{E}_n\Big[\sum_{1\le i\le M}\max_{1\le j\le K}\mu_{i,j}^{n+1}\big|S^n = s,(x^n,y^n) = (x,y)\Big] - \sum_{1\le i\le M}\max_{1\le j\le K}\mu_{i,j}^n \\
= \;& \mathbb{E}_n\Big[\max_{1\le j\le K}\mu_{x,j}^{n+1}\big|S^n = s,(x^n,y^n) = (x,y)\Big] - \max_{1\le j\le K}\mu_{x,j}^n \\
& + \sum_{i\neq x}\mathbb{E}_n\Big[\max_{1\le j\le K}\mu_{i,j}^{n+1}\big|S^n = s,(x^n,y^n) = (x,y)\Big] - \max_{1\le j\le K}\mu_{i,j}^n \\
= \;& \mathbb{E}_n\Big[\max_{1\le j\le K}\mu_{x,j}^{n+1}\big|S^n = s,(x^n,y^n) = (x,y)\Big] - \max_{1\le j\le K}\mu_{x,j}^n, \qquad (10)
\end{aligned}$$

where the last equality holds because $\mu_{i,j}^{n+1} = \mu_{i,j}^n$ if $i \neq x$. Hence, the MKG policy satisfies

$$\mathcal{A}^{\pi^{\mathrm{MKG}}}(s) = \operatorname*{arg\,max}_{1\le x\le M,1\le y\le K}\Big\{\mathbb{E}_n\Big[\max_{1\le j\le K}\mu_{x,j}^{n+1}\big|S^n = s,(x^n,y^n) = (x,y)\Big] - \max_{1\le j\le K}\mu_{x,j}^n\Big\}, \quad s\in\mathbb{S}. \qquad (11)$$

Note that the above maximization problem can be rewritten as

$$\max_{1\le x\le M}\Big\{\max_{1\le y\le K}\Big\{\mathbb{E}_n\Big[\max_{1\le j\le K}\mu_{x,j}^{n+1}\big|S^n = s,(x^n,y^n) = (x,y)\Big] - \max_{1\le j\le K}\mu_{x,j}^n\Big\}\Big\}.$$

For each $x = 1,\ldots,K$, the quantity inside the outer curly braces in the above display represents the knowledge gradient for alternative x. So the MKG policy chooses largest among the M knowledge gradients, thereby leading to its name.

Most of the effort for computing (11) resides in computing the expectation

$$\mathbb{E}_n\Big[\max_{1\le j\le K}\mu_{x,j}^{n+1}\big|S^n = s,(x^n,y^n) = (x,y)\Big] = \mathbb{E}_n\Big[\max_{1\le j\le K}\big(\mu_{x,j}^n + \tilde{\sigma}_j(\Sigma^n,x,y)Z^{n+1}\big)\big|S^n = s,(x^n,y^n) = (x,y)\Big]$$

thanks to (4), where Z^{n+1} is a standard normal random variable. This is the same as the computation of the KG policy, so we omit the details and refer to Section 3.1 of Frazier, Powell, and Dayanik (2009).

3.2 Maximum Weighted Knowledge Gradient

The MKG policy can be generalized to solve the following optimization problem

$$\max_{\pi\in\Pi}\mathbb{E}^\pi\Big[\sum_{i=1}^M w_i\max_{1\le j\le K}\mu_{i,j}^N\Big], \qquad (12)$$

where $w_1,\ldots,w_M$ are positive constants. In particular, we compute the stationary policy that maximizes

$$\mathbb{E}_n\Big[\sum_{1\le i\le M}w_i\max_{1\le j\le K}\mu_{i,j}^{n+1}\big|S^n = s,(x^n,y^n) = (x,y)\Big] - \sum_{1\le i\le M}w_i\max_{1\le j\le K}\mu_{i,j}^n,$$

and name it *maximum weighted knowledge gradient* (MWKG) policy. In the same vein as the derivation leading to (10), the decision function of the MWKG policy is

$$\mathcal{A}^{\pi^{\mathrm{MWKG}}}(s) = \operatorname*{arg\,max}_{1\le x\le M,1\le y\le K}w_x\Big\{\mathbb{E}_n\Big[\max_{1\le j\le K}\mu_{x,j}^{n+1}\big|S^n = s,(x^n,y^n) = (x,y)\Big] - \max_{1\le j\le K}\mu_{x,j}^n\Big\}, \quad s\in\mathbb{S}.$$

Its computation is similar as that of the MKG policy. We here solve

$$\max_{1\le x\le M} w_x\Big\{\max_{1\le y\le K}\Big\{\mathbb{E}_n\Big[\max_{1\le j\le K}\big(\mu^n_{x,j}+\tilde{\sigma}_j(\Sigma^n,x,y)Z^{n+1}\big)\,\big|S^n=s,(x^n,y^n)=(x,y)\Big]-\max_{1\le j\le K}\mu^n_{x,j}\Big\}\Big\}.$$

We now connect the problem (12) to our original problem (5). Suppose that we can find a set of positive weights $(w_i : 1\le i\le M)$ for which

$$\min_i \max_j \mu^N_{i,j}\approx -\sum_i w_i \max_j \mu^N_{i,j},$$

then we may solve problem (12) in lieu of problem (5). Since more information about θ is produced as more measurements are made, we do not keep the weights w_i's fixed. Instead, we update their values at each time n in order that the approximation adapt to the updated distribution of θ. Specifically, at each time n we generate L i.i.d. copies of θ according to its posterior distribution conditionally on $\mathscr{F}^n$, i.e. $\mathcal{N}(\mu^n,\Sigma^n)$. Given these i.i.d. copies, we fit $(w^n_i : i=1,\dots,M)$ using constrained least squares

$$\begin{array}{ll}\min\limits_{c^n,w^n_1,\dots,w^n_M} & \sum\limits_{l=1}^{L}\Big[c+\sum\limits_{i=1}^{M} w^n_i \max\limits_{1\le j\le K}\theta^{n,l}_{i,j}+\min\limits_{1\le i\le M}\max\limits_{1\le j\le K}\theta^{n,l}_{i,j}\Big]^2\\ \text{s.t.} & w^n_i>0,\quad i=1,\dots,M\end{array}\tag{13}$$

where $\theta^{n,l}_{i,j}$ denotes the l^{th} copy of $\theta_{i,j}|\mathscr{F}^n$. We then define the *maximum adaptively weighted knowledge gradient* (MAWKG) policy as follows. At each time $n=0,1,\dots,N-1$, first compute the weights $(w^n_i : i=1,\dots,M)$ via (13), and then compute the decision

$$\mathcal{A}^{\pi^{\text{MAWKG}}}(s)=\mathop{\arg\max}_{1\le x\le M,1\le y\le K} w^n_x\Big\{\mathbb{E}_n\Big[\max_{1\le j\le K}\mu^{n+1}_{x,j}\big|S^n=s,(x^n,y^n)=(x,y)\Big]-\max_{1\le j\le K}\mu^n_{x,j}\Big\},\quad s\in\mathbb{S}.$$

4 NUMERICAL EXPERIMENTS

We have proposed four stationary policies for sequential sampling of the Bayesian robust R&S problem, i.e. NKG, MKG, MWKG, and MAWKG. The difference between MWKG and MAWKG in terms of implementation is that the weights w_i's are fitted at time 0 and kept fixed afterwards for the former, whereas adaptively fitted at each time $n=0,1,\dots,N-1$ for the latter. When implementing both policies, we use $L=1000$ i.i.d. copies of $\theta|\mathscr{F}^n$ for the constrained least squares (13). We will also include two additional polices as follows into the numerical experiments.

- *Equal allocation* (EA). The sampling decisions are determined in a round-robin fashion: the sequence of decisions are $(1,1),(2,1),\dots,(M,1),(1,2),(2,2),\dots,(M,2),\dots,(1,K),(2,K),\dots,(M,K)$ and repeat the sequence if necessary.
- *Maximum variance* (MV). The sampling decision at each time n is to choose system (i,j) that has the maximum variance $\Sigma^n_{(i,j),(i,j)}$.

The comparison is based on 1000 randomly generated problems, each of which is parameterized by a number of sampling opportunities N, a number of systems $M\times K$, an initial mean $\mu^0\in\mathbb{R}^{M\times K}$, an initial covariance matrix $\Sigma^0\in\mathbb{R}^{MN\times MN}$, and sampling variance $\delta^2_{i,j}$, $i=1,\dots,M$, $j=1,\dots,K$. Specifically, we set $M=K=10$ and $\delta_{i,j}=1$ for each (i,j), and choose Σ^0 from the class of power exponential covariance functions, particularly

$$\Sigma^0_{(i,j),(i',j')}=\begin{cases}100e^{-|j-j'|^2}, & \text{if } i=i',\\ 0, & \text{if } i\ne i'.\end{cases}$$

Each $\mu^0_{i,j}$ is generated independently according to the uniform distribution on $[-1,1]$.

For each randomly generated problem, the true value θ is generated according to the prior belief of the problem, i.e. $\mathcal{N}(\mu^0, \Sigma^0)$. In the motivational robust R&S problem (1), we interpret M as the number of possible decisions or alternatives s_i of a simulation model, and K as the number of possible input distributions P_j. We argue that a decision-maker relying on the simulation model is more concerned of the decision s_i than of the distribution P_i. Suppose that we select system (x^N, y^N) at time N, i.e. $\mu^N_{x^N,y^N} = \min_i \max_j \mu^N_{i,j}$. Let system (i^*, j^*) be the true optimal system, i.e. $\theta_{i^*,j^*} = \min_i \max_j \theta_{i,j}$. Then, we consider it as a correct selection if $x^N = i^*$, regardless of the value of l. In other words, it really matters to select the correct alternative and not so much with the correct input distribution. In addition to the probability of correct selection in the above sense, we also compare policies based on normalized opportunity cost (NOC) of incorrect selection

$$\frac{\left|\theta_{i^*,j^*} - \max_j \theta_{x^N,j}\right|}{\sqrt{\frac{1}{MK}\sum_{i,j}|\theta_{i^*,j^*} - \theta_{i,j}|^2}}. \tag{14}$$

We apply the six completing policies to the 1000 randomly generated problem for different values of N to observe how each policy converges. For a fixed N, we record for each problem whether a policy selects the correct alternative after N sampling decisions as well as the realized NOC (14). By doing so, we estimate probability of selecting the correct alternative and NOC for each policy given N.

Table 1 presents various statistics of the realized NOC for representative values of N. Note that each problem consists of $MK = 100$ systems. Hence, $N = 100$ represents a scenario where one has sufficient computational budget, since each system is sampled once if the EA policy is applied, whereas $N = 50$ and $N = 20$ represent normal and low budgets, respectively.

Our numerical experiments show that the relative performance of the six competing policies changes considerably for different levels of computational budget. The best policy is different depending on if the budget is low, normal, or sufficient. First, if the computational budget is low, MKG outperforms the other policies by a clear margin despite of its simplicity; surprisingly, although it is the most sophisticated policy, MAWKG has the worst performance under the same circumstance, even worse than the naive EA policy. We believe that this is because at the beginning stage of sequential sampling, the state of θ are significantly different from the true value of θ, so the weights, despite being updated adaptively at each time, are too noisy and thus our effort of exploration is severely misled. In other words, MAWKG needs certain "warm-up" stage for its performance to improve. In particular, if the prior is very different from the truth, then the weights estimated in the first few rounds of sampling are so wrong that the sampling decisions may be terrible. This makes learning knowledge of the systems very inefficient in terms of reducing the overall uncertainty, which in return causes subsequent bad sampling decisions. Such a vicious cycle makes MAWKG perform poorly when the computational budget is small.

As more computational budget is available, the performance of MAWKG improves dramatically. In particular, with normal computational budget, MAWKG produces the smallest NOC and it is significantly better than the second best policy MKG; the worst performance is delivered by EA, which is not surprising since it utilizes no information of the systems at all.

At last, if the computational budget is sufficiently high, then all the policies except NKG produce small NOC, which implies that they are able to identify the optimal system, or at least the optimal row of θ, with sufficiently many sampling opportunities. The only exception, NKG, fails to do so and its NOC is at least one order of magnitude larger than the others.

Figure 1, on the other hand, illustrates the asymptotic behavior of the six competing policies in terms of the probability of correct selection (PCS) and the mean NOC as the computational budget increases. The conclusions we draw from Figure 1 are consistent with those from Table 1. First, all the policies except NKG are convergent. Second, the performance of the three proposed convergent policies, i.e. MKG, MWKG, and MAWKG, are comparable in most scenarios. They are all considerably better than the other three, EA, MV, and NKG. Only in the low budget scenario, the rate of selecting the correct alternative is lower with MAWKG than with MKG or MWKG.

Table 1: NOC based on 1000 randomly generated problems. The boxed numbers indicate the smallest means among all the policies, whereas the bold numbers the largest. Q_1 and Q_3 denote the first and third quartiles, respectively.

Budget	Stat.	Sampling Policy					
		EA	MV	NKG	MKG	MWKG	MAWKG
$N = 20$	Q_1	0.2145	0.1535	0.1778	0.0000	0.0000	0.1795
	Median	0.6276	0.5569	0.4137	0.3193	0.3502	0.6874
	Q_3	1.0273	0.9421	0.8306	0.7530	0.8157	1.1145
	Max	2.6338	2.6750	3.3601	2.6319	2.4955	2.2792
	Mean	0.6842	0.6020	0.5693	0.4544	0.4778	**0.7092**
$N = 50$	Q_1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Median	0.3384	0.0407	0.0088	0.0000	0.0000	0.0000
	Q_3	0.8074	0.4849	0.3788	0.0000	0.0000	0.0000
	Max	2.1869	2.1459	2.4811	1.7174	1.6749	1.7004
	Mean	**0.4755**	0.3022	0.2669	0.0607	0.0612	0.0316
$N = 100$	Q_1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Median	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Q_3	0.0000	0.0000	0.3476	0.0000	0.0000	0.0000
	Max	2.0913	0.4792	2.9566	0.6214	0.7567	0.4949
	Mean	0.0325	0.0149	**0.2598**	0.0128	0.0124	0.0114

To summarize, based on our numerical experiments we recommend MKG as the first choice due to its relative simplicity and good performance regardless of the computational budget. However, if the computational budget is not low and if correct selection is valued much more than simplicity of implementation, then MAWKG is the best choice.

5 CONCLUSIONS

We have proposed four extensions of the knowledge gradient policy to deal with Bayesian R&S in the presence of input distribution uncertainty. NKG, as a naive extension, is not convergent in general. The numerical experiments show that MKG, MWKG, and MAWKG are all convergent. Their performance are comparable in most scenarios except for the low budget case in which the performance of MAWKG is not satisfying. Theoretical analysis of the asymptotic behavior of these policies and their efficiency in the finite sample case are left to future research.

ACKNOWLEDGMENTS

The research is partially supported by the Hong Kong Research Grants Council under General Research Fund Project No. 624112.

REFERENCES

Bechhofer, R. E., T. J. Santner, and D. M. Goldsman. 1995. *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. New York: John Wiley & Sons, Inc.

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski. 2009. *Robust Optimization*. Princeton University Press.

Ben-Tal, A., and A. Nemirovski. 2002. "Robust Optimization – Methodology and Applications". *Mathematical Programming* 92 (3): 453–480.

Chen, C.-H., L. Dai, and H.-C. Chen. 1996. "A Gradient Approach for Smartly Allocating Computing Budget for Discrete Event Simulation". In *Proceedings of the 1996 Winter Simulation Conference*, edited by J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 398–405. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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, H.-C., C.-H. Chen, and E. Yücesan. 2000. "Computing Efforts Allocation for Ordinal Optimization and Discrete Event Simulation". *IEEE Transactions on Automatic Control* 45 (5): 960 – 964.

Chick, S. E. 2001. "Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty". *Operations Research* 49 (5): 744–758.

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. 2001a. "New Procedures to Select the Best Simulated System Using Common Random Numbers". *Management Science* 47 (8): 1133–1149.

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

Fan, W., L. J. Hong, and X. Zhang. 2013. "Robust Selection of the Best". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 868–876. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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

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

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.

He, D., S. E. Chick, and C.-H. Chen. 2007. "Opportunity Cost and OCBA Selection Procedures in Ordinal Optimization for a Fixed Number of Alternative Systems". *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37 (5): 951–961.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky. 1999. "Bayesian Model Averaging: A Tutorial". *Statistical Science* 14 (4): 382–417.

Sharir, M., and P. K. Agarwal. 1995. *Davenport-Schinzel Sequences and Their Geometric Applications*. Cambridge University Press.

AUTHOR BIOGRAPHIES

XIAOWEI ZHANG is an Assistant Professor in the Department of Industrial Engineering and Logistics Management at the Hong Kong University of Science and Technology. He received his Ph.D. in Operations Research from Stanford University in 2011. He is a member of INFORMS and his research interests include input uncertainty, simulation optimization, rare-event simulation, and stochastic modeling in service engineering. His email address is xiaoweiz@ust.hk.

LIANG DING is a PhD candidate in the Department of Industrial Engineering and Logistics Management at the Hong Kong University of Science and Technology. He graduated from the University of Toronto in 2015 with B.S. in Mathematics and B.S. in Computer Science. His email address is ldingaa@connect.ust.hk.

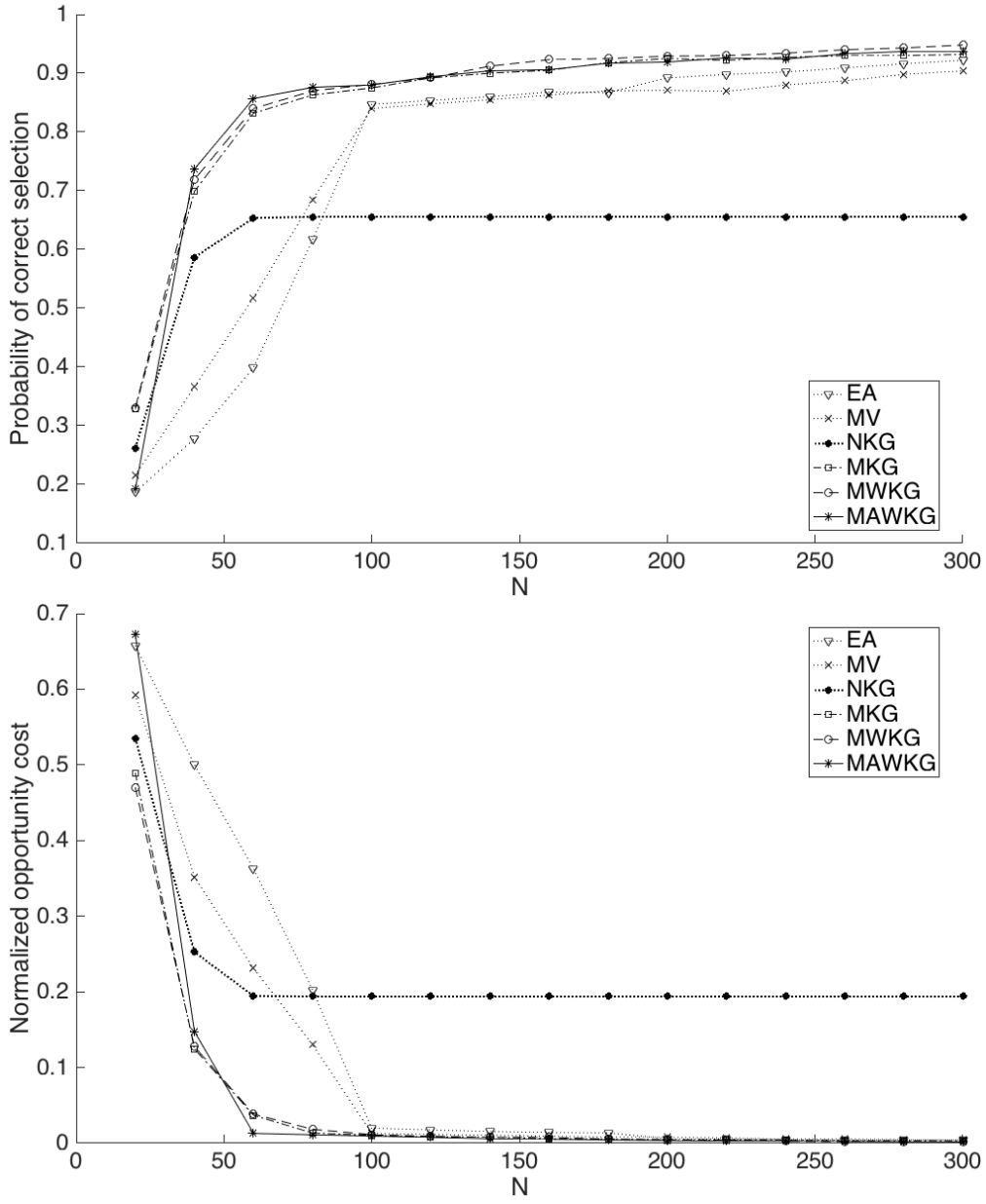

Figure 1: PCS (up) and NOC (down) estimated based on 1000 randomly generated problems.