GUESSING PREFERENCES:
A NEW APPROACH TO MULTI-ATTRIBUTE RANKING AND SELECTION

Peter I. Frazier
Aleksandr M. Kazachkov

School of Operations Research & Information Engineering
Cornell University
Ithaca, NY 14853, USA

ABSTRACT

We consider an analyst tasked with using simulation to help a decision-maker choose among several decision alternatives. Each alternative has several competing attributes, e.g., cost and quality, that are unknown but can be estimated through simulation. We model this problem in a Bayesian context, where the decision-maker's preferences are described by a utility function, but this utility function is unknown to the analyst. The analyst must choose how to allocate his simulation budget among the alternatives in the face of uncertainty about both the alternatives' attributes, and the decision-maker's preferences. Only after simulation is complete are the decision-maker's preferences revealed. In this context, we calculate the value of the information in simulation samples, and propose a new multi-attribute ranking and selection procedure based on this value. This procedure is able to incorporate prior information about the decision-maker's preferences to improve sampling efficiency.

1 INTRODUCTION

We consider the multi-attribute ranking and selection (R&S) problem, in which we have a number of alternative systems, each of which must be simulated to determine its underlying performance. In the single-attribute R&S problem, there is a single attribute of interest (for example, revenue), and the goal is to find the system for which the expected value of this attribute is largest. In the problem that we consider, there are multiple criteria of interest, and the simulation analyst has no clear way of combining them into a single criterion. Thus, our goal as simulation analysts is not to find the single best alternative, but to allocate the finite simulation budget available to us among the alternatives, in a way that best supports the multi-attribute selection decision that is to be made based on our simulation's results.

Although previous work has considered the multi-attribute R&S selection problem, we formulate the problem in a novel way. In our formulation, we consider two parties: (1) the simulation analyst and his R&S algorithm; and (2) a decision-maker who will later make a decision about some action to be performed in the real world based on the results of the simulation. We suppose that the decision-maker's decisions are driven by some underlying utility function that *does* combine the multiple attributes into a single objective function, but her utility function is *unknown* to the simulation analyst. The task that has been given to the simulation analyst is to use simulation to estimate the attributes' values for each alternative, and then present the results to the analyst, who will then make a decision according to her previously undisclosed preferences.

Although the simulation analyst does not know the decision-maker's preferences, he may have some prior belief about what they might be, based on previous interaction with the decision-maker or intuition about what is reasonable in the problem, and he can use this prior belief to focus simulation effort in a way that will best support her ultimate selection decision. This prior belief on the decision-maker's preferences allows us to derive a loss function for estimates of the alternatives' attribute values — the loss associated

with an estimate is the expected loss in utility suffered by the decision-maker when using the estimated attribute values rather than their true values.

The problem we study is which simulations the analyst should perform, taking into account his prior probability distribution on the values of the attributes, and on the utility function of the decision-maker. This choice of simulations should best support the decision-maker's final decision by maximizing the utility she can obtain by making a decision based on his results.

This formulation is only appropriate when simulation is inexpensive compared to the cost of accurately eliciting the decision-maker's utility function. Although researchers have developed methods for eliciting utility functions over multiple attributes (see, e.g., Borcherding, Eppel, and Von Winterfeldt (1991)), the successful application of these elicitation methods requires sensitivity and experience on the part of the experimentalist, as well as an investment of time and energy by the decision-maker who is the subject of the experiment (and who may be a high-level executive with a busy schedule). Complicating matters, in some situations the decision-maker is simply unavailable for interview before the decision must be made. Furthermore, even under excellent experimental conditions, the possibility for inaccuracy still exists (Froberg and Kane 1989), and residual uncertainty about the decision-maker's preferences may remain. (Indeed, if this residual uncertainty is large enough, we would still require methods like those discussed in this article.) In contrast, if the simulation experiment only requires a few extra hours or days to run without knowing the decision-maker's preferences beforehand, then running the simulation experiment in this way is preferable to requiring her input beforehand.

Our supposition that the decision-maker has an underlying (but unknown) utility function that combines multiple attributes into one utility holds in only a subset of situations for which multi-attribute R&S is required. For example, if an estimate of the Pareto frontier over multiple attributes is needed to support a multi-party negotiation whose parties have conflicting interests, then other models are more appropriate. We believe, however, our supposition is reasonable when there is only a single decision-maker — previous research on preferences over multiple attributes shows that preferences satisfying certain consistency requirements are necessarily consistent with a multi-attribute utility function having a simple functional form (see, e.g., Keeney and Raiffa (1993)).

We use a Bayesian framework to place a prior distribution on the alternatives' attribute values, which we subsequently learn through simulation. Using this Bayesian framework, we calculate the value of sampling information (Howard 1966) corresponding to the simulation samples that we can take. The simulation decision that we make is then to sample the alternative with the largest value of sampling information. This approach of valuing single samples according to the expected (myopic) value of the information they contain, and then taking the sample with the largest value, has been used by a number of researchers in other versions of the Bayesian R&S problem (Gupta and Miescke 1996, Frazier, Powell, and Dayanik 2008, Frazier, Powell, and Dayanik 2009, Chick, Branke, and Schmidt 2007, Chick, Branke, and Schmidt 2010), where it has sometimes been called the knowledge-gradient method (Frazier 2009). The value of information has also been used more broadly in R&S to choose sets of samples to take next (Chick and Inoue 2001b, Chick and Inoue 2001a, Chick and Gans 2009). This approach has also been used in many problems outside of simulation (see Frazier (2010) for a review).

A number of other researchers have considered the multi-attribute R&S problem. One related line of research concerns the multiobjective optimal computing budget allocation (MOCBA), studied in Lee, Teng, Chew, Lye, Lendermann, Karimi, Chen, and Koh (2005), Lee, Chew, and Teng (2007), Chen and Lee (2009), Lee, Chew, Teng, and Goldsman (2010). Unlike our work, the measure of the quality of an estimate of the Pareto frontier used by MOCBA procedures is not based on the interaction between the simulation analyst and the decision-maker, and does not allow the simulation analyst to incorporate prior information about the decision-maker's preference to improve sampling efficiency. This line of papers resides in a larger body of work on the optimal computing budget allocation (Chen, Lin, Yücesan, and Chick 2000, Chen and Lee 2010).

Another related line of research on multi-attribute R&S considers the problem of maximizing one attribute subject to a constraint on one or more other attributes. This work includes Andradóttir, Goldsman, and Kim (2005), Andradóttir and Kim (2010). This use of constraints has been considered by Morrice and Butler (2006), Merrick (2009), which argues that utility functions are a more appropriate way to handle multiple attributes in many applications.

Within computer science and decision analysis, a number of papers have used prior distribution on a decision-maker's unknown utility to construct efficient experimental designs for eliciting utility functions. This line of research begins with Chajewska, Koller, and Parr (2000), and includes Boutilier (2002), Abbas (2004). This line of research shares the modeling technique of placing a prior on a decision-maker's unknown preferences, but differs in the ultimate goal toward which this modeling technique is used.

2 PROBLEM FORMULATION

We begin by giving, in Section 2.1, a statistical model that describes sampling, as well as our Bayesian prior and posterior beliefs on the unknown sampling distributions for each alternative. Then, in Section 2.2, we give a model for the preferences of the decision-maker. In Section 2.3 we discuss how the prior on decision-maker preferences used by this model may be selected.

The sampling model, which assumes independent normal samples with an independent normal prior on the sampling means and an independent gamma prior on the sampling precisions, is standard, and similar models have been used in other research on Bayesian R&S (Chick and Inoue 2001b, Frazier and Powell 2008, Chick, Branke, and Schmidt 2010) as well as more broadly in Bayesian statistics (DeGroot 1970, Gelman, Carlin, Stern, and Rubin 2004). The model of the decision-maker's preferences, however, is novel.

2.1 Sampling Model

We suppose that there are a discrete set of alternatives $1, \ldots, k$, each with m attributes. Each alternative x and attribute j has a corresponding sampling distribution that is normal with mean θ_{xj} and variance λ_{xj}, with samples that are independent across alternatives, objectives, and time. We let $\theta = \{\theta_{xj} : x = 1, \ldots, k, j = 1, \ldots, m\}$ and $\lambda = \{\lambda_{xj} : x = 1 \ldots, k, j = 1, \ldots, m\}$ be the collections of all sampling means and variances respectively. At each time $n = 1, 2, \ldots$ we observe a sample y_{nj} from each attribute j for an alternative x_n of our choice, and thus the distribution of y_{nj} conditioned on x_n, θ, and λ is

$$y_{nj} | x_n, \theta, \lambda \quad \sim \quad \text{Normal}(\theta_{x_n j}, \lambda_{x_n j}), \ j = 1, \ldots, m,$$

We emphasize that the y_{nj} are conditionally independent of each other across j, because the sampling distributions are assumed independent across attribute. We let $y_n = (y_{n1}, \ldots, y_{nm})$.

Neither the sampling means nor the sampling variances are known. We adopt a Bayesian approach, in which we place a prior distribution on their values, and update this prior using observed data to obtain posterior distributions. For ease of analysis, we assume a conjugate prior distribution (see, e.g., (DeGroot 1970)), which takes the following form. Under the prior,

$$\theta_{xj} \mid \lambda_{xj} \quad \sim \quad \text{Normal}(\mu_{0xj}, \lambda_{xj}/\rho_{0xj}),$$
$$1/\lambda_{xj} \quad \sim \quad \text{Gamma}(a_{0xj}, b_{0xj}),$$

with independence across x and j, where a_{0xj}, b_{0xj}, μ_{0xj} and ρ_{0xj} are given parameters. Here, the gamma distribution is parameterized in terms of the inverse scale b_{0xj}, and not the scale, so that the mean of the gamma distribution is a_{0xj}/b_{0xj} and the mode is $(a_{0xj} - 1)/b_{0xj}$ when $a_{0xj} \geq 1$ and $b_{0xj} > 0$. In practice, one common choice is to set these parameters to $a_{0xj} = -\frac{1}{2}$, $b_{0xj} = 0$, $\rho_{0xj} = 0$, $\mu_{0xj} = 0$, which provides a non-informative prior that can be interpreted as imposing as little subjective prior belief about θ_{xj} and λ_{xj} as possible.

The posterior distribution on θ and λ at time n, which is the conditional distribution of θ and λ given observed data $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, has the same form,

$$\theta_{xj} \mid x_1, \ldots, x_n, y_1, \ldots, y_n, \lambda_{xj} \quad \sim \quad \text{Normal}(\mu_{nxj}, \lambda_{xj}/\rho_{nxj}),$$
$$1/\lambda_{xj} \mid x_1, \ldots, x_n, y_1, \ldots, y_n \quad \sim \quad \text{Gamma}(a_{nxj}, b_{nxj}),$$

where the posterior distribution remains independent across x and j, and the parameters a_{nxj}, b_{nxj}, μ_{nxj} and ρ_{nxj} can be calculated recursively. In this recursive calculation, the parameters remain unchanged from n to $n+1$ for those $x \neq x_{n+1}$, and are given by the following expressions for $x = x_{n+1}$:

$$a_{n+1,x,j} = a_{nxj} + 1/2, \qquad b_{n+1,x,j} = b_{nxj} + (1/2)(y_{n+1,j} - \mu_{nxj})^2 \rho_{nxj}/(\rho_{nxj} + 1),$$
$$\rho_{n+1,x,j} = \rho_{nxj} + 1, \qquad \mu_{n+1,x,j} = (\rho_{nxj}\mu_{nxj} + y_{n+1,j})/(\rho_{nxj} + 1). \tag{1}$$

In practice, rather than tracking both a_{nxj} and ρ_{nxj}, we need only track the number of samples taken from each alternative x. This is because a_{nxj} is a_{0xj} plus half the number of samples taken from x, and ρ_{nxj} is ρ_{0xj} plus the number of samples taken from x.

Given this posterior on θ and λ at time n, the marginal posterior distribution on θ_{xj} is independent across x and j and has the Student-t distribution with $2a_{nxj}$ degrees of freedom, location parameter μ_{nxj}, and precision $a_{nxj}\rho_{nxj}/b_{nxj}$.

In Section 3, we use the maximum a posterior estimate $\widehat{\lambda}_{nxj}$ of λ_{xj}, which is

$$\widehat{\lambda}_{nxj} = b_{nxj}/(a_{nxj} - 1). \tag{2}$$

To ensure that this estimator is well-defined and reasonably accurate, we require $a_{nxj} \geq 2$ before using its value. We let $\widehat{\lambda}_n = \{\widehat{\lambda}_{nxj} : x = 1 \ldots, k, j = 1, \ldots, m\}$ be our time-n estimate of all of the sampling variances.

Also in Section 3, we use the conditional distribution of $\mu_{n+1,x,j}$ given the information available at time n, $x_{n+1} = x$, and λ. We calculate this distribution here. Given this information, the conditional distribution of $y_{n+1,j}$ is normal with mean μ_{nxj} and variance $\lambda_{xj}/\rho_{nxj} + \lambda xj = \lambda_{xj}(\rho_{nxj} + 1)/\rho_{xnj}$. This is because $y_{n+1,j}$ is the sum of θ_{xj}, which is conditionally Normal$(\mu_{nxj}, \lambda_{xj}/\rho_{nxj})$, with independent unbiased normally distributed noise with variance λ_{xj}. Then, $\mu_{n+1,x,j} = (\rho_{nxj}\mu_{nxj} + y_{n+1,j})/(\rho_{nxj} + 1)$ is a linear transformation of a conditionally normal random variable, and is itself conditionally normal. Its mean and variance is given by

$$\mathbb{E}_n[\mu_{n+1,x,j} \mid x_{n+1} = x, \lambda] = \frac{\rho_{nxj}\mu_{nxj} + \mu_{nxj}}{\rho_{nxj} + 1} = \mu_{nxj},$$

$$\text{Var}_n[\mu_{n+1,x,j} \mid x_{n+1} = x, \lambda] = \frac{\text{Var}_n[y_{n+1,j} \mid x_{n+1} = x, \lambda]}{(\rho_{nxj} + 1)^2} = \frac{\lambda_{xj}}{\rho_{nxj}(\rho_{nxj} + 1)}.$$

where $\mathbb{E}_n$ and Var_n indicate the conditional expectation and conditional variance (respectively) with respect to the information available at time n. In the computation of the variance, we have used that $\text{Var}_n[y_{n+1,j} \mid x_{n+1} = x, \lambda] = \text{Var}_n[\theta_{xj} \mid \lambda] + \lambda_{xj} = (\lambda_{xj}/\rho_{nxj}) + \lambda_{xj} = \lambda_{xj}(\rho_{nxj} + 1)/\rho_{nxj}$.

2.2 The Decision-Maker's Preferences

We suppose that there is a decision-maker to whom we will deliver the simulation results, and that she has a utility function $U : \mathbb{R}^m \mapsto \mathbb{R}$. We suppose that this utility function is linear, and so is given by $U(\theta) = c \cdot \theta = \sum_{j=1}^m c_j \theta_j$, for some vector of preferences $c \in \mathbb{R}^m$ and any vector of attribute values $\theta \in \mathbb{R}^m$. Here $\cdot$ indicates the dot product. If the preference vector c were known, then we could use a single-objective ranking & selection strategy, of which several have been developed in the literature (a partial list includes Chen, Lin, Yücesan, and Chick (2000), Chick and Inoue (2001b), Gupta and Miescke (1996), Frazier, Powell, and Dayanik (2009)).

Instead, we suppose that the preference vector c is unknown and we have a prior probability distribution on its value. When the simulation analyst has little information about the preferences of the decision-maker, it may be prudent to specify a non-informative prior on c, as described below in Section 2.3. In other cases, he may have more specific information, and this framework allows incorporating that information into his simulation sample selection.

Having a prior probability distribution on c can be interpreted as being uncertain about the true preferences of the decision-maker, but having some reasonable guesses based on conversations with the decision-maker, experience with the decision-maker's decisions in previous situations, or intuition about the current problem. For example, in a staffing problem for a small hospital, an employee charged with running simulations might guess that a hospital administrator would feel that saving an average of 15 minutes of waiting time for each patient visiting the emergency room might be worth the salary of 1 additional nurse, but is unlikely to be worth the salary of 10 additional nurses. If the employee had a great deal of previous experience with solving problems for this administrator, or a great deal of experience working with similar problems, that employee might be able to make more refined guesses. In making these guesses, the employee would not feel certain that these guesses were correct, but would feel that they were reasonable.

We will see that these prior distributions allow one to make better use of one's ability to simulate, in estimating which alternatives will be closest to optimizing the decision-maker's true preferences.

After taking a number of samples N (this number of samples can be fixed a priori, or can be chosen adaptively by the policy), we suppose that the simulation analyst provides to the decision maker a collection of alternatives and their estimated characteristics. The decision-maker then chooses the best alternative according to the estimated attribute values and her underlying and previously unobserved preferences encoded in the preference vector c. The choice that she makes is

$$\arg\max_x \mathbb{E}_N\left[U(\theta_x) \mid U\right] = \arg\max_x c \cdot \mathbb{E}_N\left[\theta_x\right] = \arg\max_x c \cdot \mu_{Nx} \tag{3}$$

where $\theta_x = (\theta_{x1}, \ldots, \theta_{xm})$ is the vector of true attributes of alternative x, and $\mu_{Nx} = \mathbb{E}_N\left[\theta_x\right]$ is the conditional expectation of θ_x under our posterior belief at time N.

The linearity of the utility function implies that the variance at the final time N of our posterior belief about θ plays no role in the choice made by the decision-maker. This corresponds to the decision-maker's lack of aversion to the risk in not knowing the true distribution of the chosen alternative. This is risk that could be addressed by more simulation. Additionally, the decision-maker does not consider the sampling variances λ when making her decision, which corresponds to a lack of aversion to the risk from the randomness that will appear when implementing her chosen alternative in the real world. In future work we plan to consider the role of risk aversion. The posterior and sampling variances at earlier times play a role in the choice of which alternative to sample.

With the choice (3) by the decision-maker, the value she obtains at the final time is $\max_x c \cdot \mu_{Nx}$, and the expected value obtained by a measurement policy π is

$$\mathbb{E}^\pi\left[\max_x c \cdot \mu_{Nx}\right]. \tag{4}$$

The simulation analyst's goal is then to use a measurement policy that maximizes the quantity (4).

2.3 The Choice of Prior Distribution

In this section we discuss how the simulation analyst might choose the prior distribution on c.

We first observe that the simulation analyst may often know the sign of the decision-maker's preference for each attribute, $\text{sign}(c_j)$, even if he does not know the preference vector itself. For example, in the hospital staffing problem, we know that the sign of the attribute corresponding to cost is negative (less cost is preferred), as is the sign of the attribute corresponding to patient waiting time (less waiting time is

preferred). In this case, the prior on c can be restricted to that part of $\mathbb{R}^m$ that has the correct set of signs. In the rest of this section, we suppose without loss of generality that the sign of each c_j is positive (if not, we simply flip the sign of samples of attribute j).

Second, we observe that the preference order over alternatives induced by a preference vector c is identical to that induced by c/a for any scalar a. Thus, for specifying a prior on the order of the decision-maker's preference over alternatives, it is enough to specify a prior on the vector with unit magnitude, $c/||c||$. One can then make the simplifying assumption that $||c|| = 1$. In the case where there are $m = 2$ attributes, all that is then required is to specify a prior on an angle between 0 and $\pi/2$. More generally, with $m \geq 2$, one must specify a prior on the intersection of the unit sphere and the positive orthant.

This assumption that $||c|| = 1$ is not without loss of generality. Even though dividing c by a scalar does not change the decision-maker's preference order, choosing a prior that places weight on several different magnitudes $||c||$ affects the value of sampling information discussed in Section 3.1. In $m = 2$, such a prior might correspond to the belief that the overall magnitude of the decision-maker's preferences are large when the angle of the preferences $c/||c||$ takes one value, while the overall magnitude is lower when it takes other values. While this might be occasionally useful, and is allowed by our framework, we feel that in most situations this additional flexibility is not needed, and we recommend specifying priors with $||c|| = 1$.

When the simulation analyst has little information about the preferences of the decision-maker, it may be prudent to specify a non-informative prior on c. One convenient non-informative prior is the one that is uniform on the intersection of the unit sphere and the positive orthant. When $m = 2$, this prior is the one that is uniform on the angles $[0, \pi/2]$.

In Section 3 we provide exact expressions for the value of information and the knowledge-gradient policy when the prior on c is discrete. Thus, for computational convenience, we recommend approximating continuous priors through discretization. When $m = 2$ and one has a non-informative prior on c, a convenient discrete prior is $c_\ell = (\cos(\alpha_\ell), \sin(\alpha_\ell))$, $\alpha_\ell = \frac{\pi}{2}(\ell-1)/(L-1)$, for $\ell = 1, \ldots, L$, and $\mathbb{P}\{c = c_\ell\} = 1/L$.

3 KNOWLEDGE-GRADIENT POLICY

The knowledge-gradient (KG) policy is a one-step lookahead method based on a value of information calculation. It supposes that the current opportunity to sample is the last that will be allowed, and that immediately afterward the simulation analyst must provide results to the decision-maker. In the KG policy, we make the decision that would be optimal in this case.

3.1 Derivation of the Knowledge-Gradient Policy

To calculate this KG decision, we calculate the expected incremental value that would be obtained by showing results to the decision-maker *after* sampling alternative x, as opposed to showing results to the decision-maker *before* sampling x. We call this quantity the *KG factor* for alternative x. The KG policy tells us to measure the alternative with the largest KG factor.

To make calculation of the KG factor analytically tractable, we suppose, only for the purposes of this calculation, that the sampling variances λ are given to us. When using the KG policy in problems with unknown sampling variance, we take the current maximum a posteriori estimate $\widehat{\lambda}_{nxj}$ of λ_{nxj} from (2), and substitute this for the true sampling variance when calculating KG factors. After each sample we update these estimates of the sampling variances.

With the assumption that the sampling variances λ are known, the KG factor for sampling alternative x at time n is denoted $v_{nx}(\lambda)$ and is defined to be,

$$v_{nx}(\lambda) = \mathbb{E}_n\left[\left(\max_{x'} c \cdot \mu_{n+1,x'}\right) - \left(\max_{x'} c \cdot \mu_{n,x'}\right) | x_{n+1} = x, \lambda\right].$$

We now calculate this more explicitly. We condition on c and use the tower property of conditional expectation to obtain

$$v_{nx}(\lambda) = \mathbb{E}_n \left[\mathbb{E}_n \left[\left(\max_{x'} c \cdot \mu_{n+1,x'} \right) - \left(\max_{x'} c \cdot \mu_{n,x'} \right) \mid c, x_{n+1} = x \right] \mid x_{n+1} = x, \lambda \right]$$
$$= \mathbb{E}_n \left[v_{nx}(c, \lambda) \mid x_{n+1} = x, \lambda \right] = \mathbb{E}_n \left[v_{nx}(c, \lambda) \mid \lambda \right],$$

where $v_{nx}(c, \lambda)$ is defined to be $v_{nx}(c, \lambda) = \mathbb{E}_n \left[\left(\max_{x'} c \cdot \mu_{n+1,x'} \right) - \left(\max_{x'} c \cdot \mu_{n,x'} \right) \mid c, x_{n+1} = x, \lambda \right]$, and the last equality follows because $v_{nx}(c, \lambda)$ does not depend upon x_{n+1}. The last expectation is over our uncertainty about c. The quantity $v_{nx}(c, \lambda)$ is the expected incremental improvement from sampling x that we would obtain if we also knew the decision-makers preferences c. We do not know these preferences, and so we must take an expectation over this uncertainty.

The quantity $v_{nx}(\lambda, c)$ is computed more explicitly in the following proposition. Its proof may be found in the appendix.

Proposition 1 $v_{nx}(c, \lambda) = \widetilde{\sigma}_{nx}(c, \lambda) f \left(\frac{-\Delta_{nx}(c)}{\widetilde{\sigma}_{nx}(c, \lambda)} \right)$, where

$$(\widetilde{\sigma}_{nx}(c, \lambda))^2 = \sum_{j=1}^{m} c_j^2 \frac{\lambda_{xj}}{\rho_{nxj}(\rho_{nxj} + 1)}, \quad f(d) = d\Phi(d) + \varphi(d), \quad \Delta_{nx}(c) = \left| c \cdot \mu_{nx} - \max_{x' \neq x} c \cdot \mu_{nx'} \right|,$$

Φ is the normal cumulative distribution function, and φ is the normal probability density function.

Proposition 1 implies that the KG factor for sampling alternative x is

$$v_{nx}(\lambda) = \mathbb{E}_n \left[\widetilde{\sigma}_{nx}(c, \lambda) f \left(\frac{-\Delta_{nx}(c)}{\widetilde{\sigma}_{nx}(c, \lambda)} \right) \mid \lambda \right],$$

where the expectation is over c.

The posterior on c is the same as the prior because we obtain no information about the decision-maker's preferences until the final time. If the prior on c is discrete over a set $c_1, \ldots, c_L$, with probabilities $p_\ell = \mathbb{P}\{c = c_\ell\}$ that c is equal to each of these values, then this expectation may be computed exactly as:

$$v_{nx}(\lambda) = \sum_\ell p_\ell \widetilde{\sigma}_{nx}(c_\ell, \lambda) f \left(\frac{-\Delta_{nx}(c_\ell)}{\widetilde{\sigma}_{nx}(c_\ell, \lambda)} \right). \tag{5}$$

When the sampling variances are known, the KG policy tells us to measure the alternative x with the largest such value, $x_{n+1} \in \arg\max_x v_{nx}(\lambda)$. When the sampling variances are unknown, we substitute the maximum a posteriori estimate $\widehat{\lambda}_n$ for the sampling variances λ, and the KG policy is

$$x_{n+1} \in \arg\max_x v_{nx}(\widehat{\lambda}_n). \tag{6}$$

Together, (5) and (6) define the KG policy.

3.2 Summary of the Knowledge-Gradient Policy

We now summarize the way in which samples are collected and inference is performed in the KG policy:

1. **Initialization**: Choose the parameters of the prior distribution, a number of initial samples n_0, and an overall number of samples $N > kn_0$. If we have no strong prior beliefs about the sampling means and variances, use the non-informative prior: $a_{0xj} = -1/2$, $b_{0xj} = 0$, $\rho_{0xj} = 0$, $\mu_{0xj} = 0$. Choose a discrete prior on c, $p_\ell = \mathbb{P}\{c = c_\ell\}$ for $\ell = 1, \ldots, L$.

2. **Initial Sampling**: Choose the number of initial samples n_0 satisfying $a_{0xj} + n_0 \geq 2$ for each x and j. Collect n_0 samples from each alternative and let $n = kn_0$. Calculate the parameters a_{nxj}, b_{nxj}, ρ_{nxj}, μ_{nxj} for all x,j from the data collected using the recursive expression (1).

3. **Calculate KG Factors:** For each alternative x,

   (a) Calculate $\widehat{\lambda}_{nxj} = b_{nxj}/(a_{nxj} - 1)$ for $j = 1, \ldots, m$.

   (b) Calculate $\widetilde{\sigma}(c_\ell, \widehat{\lambda}_n) = \left(\sum_{j=1}^{m} c_{\ell j}^2 \widehat{\lambda}_{nxj} / (\rho_{nxj}(\rho_{nxj} + 1)) \right)^{1/2}$ for $\ell = 1, \ldots, L$.

   (c) Calculate $\Delta_{nx}(c_\ell) = \left| c_\ell \cdot \mu_{nx} - \max_{x' \neq x} c_\ell \cdot \mu_{nx'} \right|$ for $\ell = 1, \ldots, L$.

   (d) Calculate the KG factor using $f(d) = d\Phi(d) + \varphi(d)$ and

$$v_{nx} = \sum_{\ell=1}^{L} p_\ell \widetilde{\sigma}_{nx}(c_\ell, \lambda) f\left(\frac{-\Delta_{nx}(c_\ell)}{\widetilde{\sigma}_{nx}(c_\ell, \lambda)} \right),$$

4. **Sample:** Sample alternative $x_{n+1} \in \arg\max_x v_{nx}(\widehat{\lambda}_n)$ and observe $y_{n+1,j}$ for each $j = 1, \ldots, m$.

5. **Update Posterior:** Update the parameters $a_{n+1,x,j}$, $b_{n+1,x,j}$, $\rho_{n+1,x,j}$, $\mu_{n+1,x,j}$ of the posterior over all attributes j and all alternatives x. For $x = x_{n+1}$, the updated parameters are computed from $y_{n+1,j}$ via (1), and for $x \neq x_{n+1}$ the parameters are unchanged from time n.

6. **Loop:** Let $n = n + 1$. If $n = N$, stop sampling. Otherwise go to Step 3

4 EMPIRICAL RESULTS

We performed computational experiments to assess the effectiveness of the KG policy and to illustrate how the prior on decision maker preferences influences its behavior.

In Figure 1, we compared the expected performance of the KG policy against that of the equal allocation policy, which allocates the same number of samples to each point in round-robin fashion. The problem considered has $k = 20$ alternatives, each with $m = 2$ attributes, with a sampling budget of $N = 600$, including the $n_0 = 5$ initial samples per alternative. We assumed a known sampling variance of $\lambda_{xj} = 1$ across all alternatives and attributes. The KG policy began with a non-informative prior on the sampling means. There were $L = 5$ possible values c_ℓ for c, equidistant from each other on the unit quarter circle in the first quadrant. That is, $c_\ell = (\cos(\alpha_\ell), \sin(\alpha_\ell))$ where $\alpha_\ell = \frac{\pi}{2}(\ell - 1)/4$ with $\ell = 1, \ldots, 5$. Our prior on c placed equal probability on each c_ℓ, so $p_\ell = 1/5$.

Performance was averaged over 1000 replications. In each replication, we sampled a new set of sampling means θ from an independent standard normal distribution. At each iteration $n \geq 100$ after the initial stage of sampling, we calculated the expected value the decision-maker would receive if N were equal to n, based on the true sampling means θ and the prior on c. This value is $\sum_{\ell=1}^{L} p_\ell(c \cdot \theta_{x^*(n,c)})$, where $x^*(n,c) \in \arg\max_x c \cdot \mu_{nx}$ is the alternative that the decision-maker would choose at time n as given in (3). The expected value of this performance measure is identical to the expected value of a performance measure calculated by also randomly sampling one hidden true value of c on each replication.

Figure 1 plots this average performance as a function of n for $n > kn_0$, i.e., after KG completes its initial stage of sampling. During this initial stage, KG uses the equal allocation policy, and the performance of the two policies is identical. The maximum standard error of the average performance (which is as an estimate of expected performance), under both KG and equal allocation and over all n, was 0.011.

Figure 1 shows that the KG policy performs well against equal allocation in this experiment. This is not surprising, since the equal allocation policy is known to underperform more sophisticated adaptive sampling policies in a number of other R&S problems. It demonstrates, however, that the behavior of the KG policy is reasonable. In future work, we plan to compare the performance of the KG policy against other more sophisticated multi-attribute R&S procedures, both using the new measure of quality of a multi-attribute sampling procedure, and more traditional measures of quality.
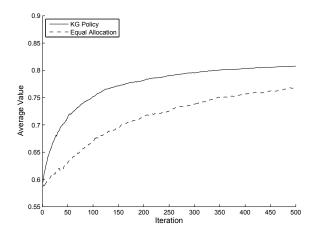
Figure 1: Bayes performance of the KG and equal allocation policies as a function of the number of samples taken after the initial stage of experiments completes (the iteration, $n - kn_0$). For each value of n and each policy, an estimate of the expected Bayes reward (equation (4)) is plotted. Here, $k = 20$, $\lambda_{xj} = 1$, $\mu_{0xj} = 0$, $n_0 = 5$. There were five possible values of c, and they were all equiprobable. The maximum standard error was 0.011.

In Figure 2, we consider $k = 50$ alternatives with $m = 2$ attributes each, whose sampling means were again generated from a standard normal distribution. The KG policy began with a non-informative prior, with $n_0 = 5$ samples taken from each alternative, and a known sampling variance of $\lambda_{xj} = 1$. Rather than averaging over many replications, we display the results from one replication after $N = 600$ samples, where each alternative's true sampling mean $\theta_x = (\theta_{x1}, \theta_{x2})$ is plotted as a filled black square, and the posterior mean is plotted as an open diamond. The sampling mean and posterior mean are connected by a line, and the length of this line illustrates the posterior variance and the number of times that the alternative was sampled – shorter lines usually correspond to alternatives that were sampled more often. We use the same $L = 5$ possible values c_ℓ as in Figure 1, but with two different priors on these values. In the left panel, the prior places more weight on those c_ℓ preferring large values of θ_1 at the expense of θ_2, while in the right panel the prior places more weight on those c_ℓ preferring large values of θ_2 at the expense of θ_1. This prior was

$$p_\ell = \frac{2^{L-\ell+1}}{\sum_{i=1}^{L} 2^{L-i+1}} \text{ (left panel),} \qquad p_\ell = \frac{2^\ell}{\sum_{i=1}^{L} 2^i} \text{ (right panel),} \qquad \ell = 1, \ldots, L.$$

In both panels, the KG policy has sampled more often from those alternatives on the Pareto front (those alternatives up or to the right), and less often from those alternatives away from the Pareto front. Additionally, the KG policy has sampled more from the alternatives that performed better in the attribute that was considered more important. The alternatives with a larger value on the horizontal axis show a smaller variance in Figure 2(a) than in Figure 2(b), while those with a larger value on the vertical axis show a smaller variance in Figure 2(b) than in Figure 2(a). Although the KG policy spends more effort on the attribute believed to be more important, it does also spend some effort on the attributes believed to be less important, as evidenced in Figure 2(a), for example, by the greater number of samples allocated to those alternatives with larger values on the vertical axis than to those alternatives away from the Pareto frontier. Thus, by altering the prior, one can encourage the KG policy to spend more time focusing on one attribute at the expense of the others. At the same time, this encouragement does not preclude the KG policy from spending effort along the entire Pareto frontier.

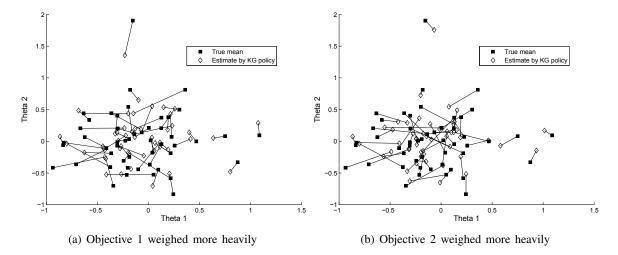(a) Objective 1 weighed more heavily   (b) Objective 2 weighed more heavily

Figure 2: Illustration of the true values θ and posterior estimates μ_{Nx} under the KG policy after exhausting the sampling budget, with five different values of c. In the left panel, the horizontal axis is likely to be more important. In the right panel, the vertical axis is likely to be more important. In each panel, for each alternative x, lines connect a square at the true sampling mean $\theta_x = (\theta_{x1}, \theta_{x2})$ with its posterior mean $\mu_{Nx} = (\mu_{Nx1}, \mu_{Nx2})$. In these figures, $k = 50$.

5   CONCLUSION

We have considered the multi-attribute Bayesian R&S problem from a new viewpoint, in which the decision-maker has a single utility function, but this utility function is unknown to the simulation analyst. Within this framework, we have developed a new procedure for choosing which alternative to sample next. This procedure myopically maximizes the value of the information obtained. By using the prior on the preferences of the decision-maker, we can influence the algorithm to spend more time estimating the attributes of those alternatives that are more important for the decision-maker's final selection.

A number of possible extensions present themselves for future work. One might extend the model to allow for correlated samples across criteria for a single alternative, or across alternatives through the use of common random numbers. One might also explore non-myopic or batch algorithms, which have the potential for providing better performance within this framework. One might also study the role that the decision-maker's risk aversion plays in the decisions that the simulation analyst should make. One might also consider correlated prior distributions, allowing us to learn about unsampled alternatives from the sampling results of related alternatives, and enabling us to solve large-scale problems in which the number of alternatives far exceeds the number of available samples.

Another interesting direction for future work is to enhance the model of interaction between the simulation analyst and the decision-maker to allow some feedback on preferences during simulation. In this extension, the simulation analyst could ask the decision-maker for feedback about interim or hypothetical results in order to reduce the amount of simulation. In this enhanced model, the analyst could show the decision-maker interim or hypothetical results and ask her which of them she would prefer. This would allow the decision-maker to obtain some partial information about the decision-maker's preferences, which the algorithm could then use to focus its sampling effort. This interaction could be considered a form of prior elicitation. This interaction would be modeled as bearing some cost (because it occupies the decision-maker's time), and so the analyst would be led to to use this interaction only when it would significantly decrease the amount of simulation effort required to accurately support the decision-maker's ultimate decision.

The overarching goal of this work has been to more accurately model the way in which simulation is used as an aid to human decision-making. We hope that the model described in this paper will contribute

to the progress being made in the larger literature toward this goal, and that better models will lead to even more productive use of simulation by society at large.

A PROOF OF PROPOSITION 1

First $c \cdot \mu_{n+1,x'} = c \cdot \mu_{n,x'}$ for $x' \neq x = x_{n+1}$. Now consider $c \cdot \mu_{n+1,x}$. We observed in Section 2.1 that, conditional on the information available at time n, $x_{n+1} = x$, and λ, $\mu_{n+1,x,j} \sim \text{Normal}\left(\mu_{nxj}, \lambda_{xj}/(\rho_{nxj}(\rho_{nxj}+1))\right)$ with independence across j. This shows that the conditional distribution of $c \cdot \mu_{nx}$ is

$$\text{Normal}(c \cdot \mu_{nx}, \sum_{j=1}^{m} c_j^2 \lambda_{xj}/(\rho_{nxj}(\rho_{nxj}+1))).$$

Observe that $(\widetilde{\sigma}_{nx}(c,\lambda))^2 = \sum_{j=1}^{m} c_j^2 \lambda_{xj}/(\rho_{nxj}(\rho_{nxj}+1))$ is the variance of the distribution. Let Z be an independent standard normal random variable, so $\max_{x'} c \cdot \mu_{n+1,x'}$ is equal in conditional distribution to $\max\{\max_{x' \neq x} c \cdot \mu_{nx'}, c \cdot \mu_{nx} + \widetilde{\sigma}_{nx}(c,\lambda)Z\}$.

In general we have for $a_1, a_2, b \in \mathbb{R}$ that,

$$\mathbb{E}\left[\max\{a_1, a_2 + bZ\}\right] = \max(a_1, a_2) + \mathbb{E}\left[\max\{0, -|a_1 - a_2| + bZ\}\right]$$

$$= \max(a_1, a_2) + b\mathbb{E}\left[\max\{0, -\frac{|a_1 - a_2|}{b} + Z\}\right] = \max(a_1, a_2) + bf\left(\frac{-|a_1 - a_2|}{b}\right)$$

where $f(d) = \mathbb{E}[\max\{0, d+Z\}] = \int_{-\infty}^{\infty} \varphi(z) \max\{0, d+z\} \, dz$

$$= \int_{-d}^{\infty} \varphi(z)(d+z) \, dz = d\int_{-d}^{\infty} \varphi(z) \, dz + \int_{-d}^{\infty} \varphi(z)z \, dz = d\Phi(d) + \varphi(d).$$

Thus, with $a_1 = \max_{x' \neq x} c \cdot \mu_{nx'}$, $a_2 = c \cdot \mu_{nx}$, and $b = \widetilde{\sigma}_{nx}(c,\lambda)$, we have $\max(a_1, a_2) = \max_{x'} c \cdot \mu_{nx'}$, and

$$\mathbb{E}_n\left[\max_{x'} c \cdot \mu_{n+1,x'}\right] = \max_{x'} c \cdot \mu_{nx} + \widetilde{\sigma}_{nx}(c,\lambda)f\left(\frac{-\Delta_{nx}(c)}{\widetilde{\sigma}_{nx}(c,\lambda)}\right),$$

where $\Delta_{nx}(c) = |a_1 - a_2| = \left|c \cdot \mu_{nx} - \max_{x' \neq x} c \cdot \mu_{nx'}\right|$ and $v_{nx}(c,\lambda) = \widetilde{\sigma}_{nx}(c,\lambda)f\left(\frac{-\Delta_{nx}(c)}{\widetilde{\sigma}_{nx}(c,\lambda)}\right)$.

AUTHOR BIOGRAPHIES

PETER I. FRAZIER is an assistant professor in the School of Operations Research and Information Engineering at Cornell University. He received a Ph.D. in Operations Research and Financial Engineering from Princeton University in 2009. He is the recipient of an AFOSR Young Investigator Award in 2010. His research interest is in dynamic programming and Bayesian statistics, focusing on the optimal acquisition of information. He works on applications in simulation, global optimization, operations management, and medicine. His web page can be found via http://www.orie.cornell.edu, and his email address is pf98@cornell.edu.

ALEKSANDR M. KAZACHKOV received his B.S. in Operations Research and Information Engineering from Cornell University in May 2011. He will be starting the Ph.D. program in Algorithms, Combinatorics, and Optimization at Carnegie Mellon University in the Fall of 2011. His email address is akazachk@gmail.com.

REFERENCES

Abbas, A. 2004. "Entropy methods for adaptive utility elicitation". *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 34 (2): 169–178.

Andradóttir, S., D. Goldsman, and S. Kim. 2005, December. "Finding the best in the presence of a stochastic constraint". In *Proceedings of the 2005 Winter Simulation Conference*, edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 732–738. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Andradóttir, S., and S. Kim. 2010. "Fully sequential procedures for comparing constrained systems via simulation". *Naval Research Logistics (NRL)* 57 (5): 403–421.

Borcherding, K., T. Eppel, and D. Von Winterfeldt. 1991. "Comparison of weighting judgements in multiattribute utility measurement". *Management Science* 37 (12): 1603–1619.

Boutilier, C. 2002. "A POMDP formulation of preference elicitation problems". In *Proceedings of the National Conference on Artificial Intelligence*, edited by K. Ford, 239–246. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Chajewska, U., D. Koller, and R. Parr. 2000. "Making rational decisions using adaptive utility elicitation". In *Proceedings of the National Conference on Artificial Intelligence*, edited by K. Ford, 363–369. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Chen, C., and L. Lee. 2010. *Stochastic simulation optimization: an optimal computing budget allocation*. World Scientific Pub Co Inc.

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

Chen, E., and L. Lee. 2009. "A multi-objective selection procedure of determining a Pareto set". *Computers & Operations Research* 36 (6): 1872–1879.

Chick, S., J. Branke, and C. Schmidt. 2007, December. "New greedy myopic and existing asymptotic sequential selection procedures: preliminary empirical results". In *Proceedings of the 2007 Winter Simulation Conference*, edited by S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 289–296. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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

Chick, S., and N. Gans. 2009. "Economic Analysis of Simulation Selection Problems". *Management Science* 55 (3): 421–437.

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

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

DeGroot, M. H. 1970. *Optimal Statistical Decisions*. New York: McGraw Hill.

Frazier, P. 2009. *Knowledge-Gradient Methods for Statistical Learning*. Ph. D. thesis, Princeton University.

Frazier, P. 2010. *Wiley Encyclopedia of Operations Research and Management Science*, Chapter Learning with Dynamic Programming. Wiley.

Frazier, P., and W. Powell. 2008, December. "The Knowledge-Gradient Stopping Rule for Ranking and Selection". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Moench, O. Rose, T. Jefferson, and J. W. Fowler, 305–312. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Frazier, P., 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.

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

Froberg, D., and R. Kane. 1989. "Methodology for measuring health-state preferences–II: Scaling methods". *Journal of Clinical Epidemiology* 42 (5): 459–471.

Gelman, A., J. Carlin, H. Stern, and D. Rubin. 2004. *Bayesian data analysis*. second ed. Boca Raton, FL: CRC Press.

Gupta, S., and K. 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.

Howard, R. 1966. "Information Value Theory". *Systems Science and Cybernetics, IEEE Transactions on* 2 (1): 22–26.

Keeney, R., and H. Raiffa. 1993. *Decisions with multiple objectives: Preferences and value tradeoffs*. Cambridge Univ Pr.

Lee, L., E. Chew, and S. Teng. 2007, December. "Finding the pareto set for multi-objective simulation models by minimization of expected opportunity cost". In *Proceedings of the 2007 Winter Simulation Conference*, edited by S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 513–521. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lee, L., E. Chew, S. Teng, and D. Goldsman. 2010. "Finding the non-dominated pareto set for multi-objective simulation models". *IIE Transactions* 42 (9): 656–674.

Lee, L., S. Teng, E. Chew, K. Lye, P. Lendermann, I. Karimi, Y. Chen, and C. Koh. 2005, December. "Application of multi-objective simulation-optimization techniques to inventory management problems". In *Proceedings of the 2005 Winter Simulation Conference*, edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 1684–1691. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Merrick, J. 2009. "Bayesian Simulation and Decision Analysis: An Expository Survey". *Decision Analysis* 6 (4): 222–238.

Morrice, D., and J. Butler. 2006, December. "Ranking and Selection with Multiple Targets". In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 222–230. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.