Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, eds.

TUTORIAL: OPTIMIZATION VIA SIMULATION WITH BAYESIAN STATISTICS AND DYNAMIC PROGRAMMING

Peter Frazier

Cornell University
Ithaca, NY 14850, USA

ABSTRACT

Bayesian statistics comprises a powerful set of methods for analyzing simulated systems. Combined with dynamic programming and other methods for sequential decision making under uncertainty, Bayesian methods have been used to design algorithms for finding the best of several simulated systems. When the dynamic program can be solved exactly, these algorithms have optimal average-case performance. In other situations, this dynamic programming analysis supports the development of approximate methods with sub-optimal but nevertheless good average-case performance. These methods with good average-case performance are particularly useful when the cost of simulation prevents the use of procedures with worst-case statistical performance guarantees. We provide an overview of Bayesian methods used for selecting the best, providing an in-depth treatment of the simpler case of ranking and selection with independent priors appropriate for smaller-scale problems, and then discussing how these same ideas can be applied to correlated priors appropriate for large-scale problems.

1 INTRODUCTION

Optimization via simulation (OvS) is the act of solving an optimization problem whose objective function cannot be calculated analytically, and must instead be estimated with stochastic simulation. Such problems are often quite difficult to solve in a computationally tractable manner. This tutorial discusses how fast algorithms for OvS can be designed in a principled way using Bayesian statistics and dynamic programming.

In particular, this tutorial shows a conceptual approach to creating algorithms with optimal average-case performance. Here, average-case optimal means that when the performance of such an algorithm is averaged over a large number of typical problem instances, its average performance is the best possible among all algorithms. While the worst-case performance of such an algorithm might not be particularly good, and so there may be problem instances on which the algorithm is very slow or has poor accuracy, an algorithm that works very well most of the time can be of great value in many situations.

In practice, the conceptual method this tutorial describes for creating an average-case optimal algorithm is often not computationally feasible, because it requires solving a large dynamic program. However, approximations to the solution to the dynamic program often provide practical algorithms that work very well in the average-case, even though they are sub-optimal.

Several other tutorials and surveys cover the use of Bayesian methods within simulation more broadly (Chick 2000; Chick 2005; Chick 2006; Merrick 2009), and these have some treatment of OvS, although dynamic programming is not discussed. The survey article Frazier (2010) covers several of the topics here, including the use of dynamic programming, but moves at a faster pace. The tutorial chapter Powell and Frazier (2008) and the recent textbook Powell and Ryzhov (2012), both written at the advanced undergraduate level, cover the same problems within OvS, and include treatment of dynamic programming.

A number of textbooks cover dynamic programming or Bayesian statistics separately. Dynamic programming textbooks include Powell (2007), Dynkin and Yushkevich (1979), Bertsekas (2005), Bertsekas (2007) and Ross (1983). Bayesian statistics textbooks include Gelman et al. (2004) at the advanced

undergraduate or introductory graduate level, and Berger (1985) at the advanced graduate level. The combination of Bayesian statistics and dynamic programming that we consider was pioneered within Bayesian sequential experimental design (see the monographs Berry and Fristedt (1985), Wetherill and Brown (1991)) and the one-stage analysis that we perform in Section 5 is a value of information calculation, pioneered by Howard (1966).

In this tutorial, we first focus on the problem of OvS with a small number of alternatives and no special structure. This problem is also called ranking and selection (R&S). After providing overviews of OvS and R&S (Section 2) and Bayesian statistics (Section 3), we consider Bayesian inference for R&S in Section 4, and then decision-making for the one-stage version of R&S in Section 5 and the fully sequential version in Section 6. In Section 7, we consider large-scale OvS problems.

The specific algorithms that we discuss are the (R1,...,R1) algorithm (Gupta and Miescke 1996), analyzed later under the name knowledge-gradient (KG) policy for independent beliefs (Frazier, Powell, and Dayanik 2008); a known-variance version of the LL(B) policy (Chick and Inoue 2001); and the KG policy for correlated beliefs (Frazier, Powell, and Dayanik 2009).

2 OPTIMIZATION VIA SIMULATION AND RANKING AND SELECTION

In this tutorial, we apply Bayesian methods to two problems: ranking and selection (R&S); and optimization via simulation (OvS). In both problems, we have a collection of *alternative systems*, and our goal is to determine which of them is the best. Typical applications include choosing the inventory policy with minimum expected cost or choosing the staffing levels in a call center that minimize the probability that a customer waits longer than a threshold. We indicate an alternative system with the notation x, and the set of all alternatives with $\mathscr{X}$.

There is a large body of research on R&S, beginning with Bechhofer (1954). Much of the work is discussed in the monograph Bechhofer, Santner, and Goldsman (1995), with other more recent surveys and tutorials including Swisher, Jacobson, and Yücesan (2003), Kim and Nelson (2006) and Kim and Nelson (2007). Tutorials and surveys that discuss OvS more broadly include Fu (1994), Andradóttir (1998), Fu (2002), Swisher, Hyden, Jacobson, Schruben, Hosp, and Fredericksburg (2000).

In R&S, the alternative systems are unordered, and it is unclear whether learning about the quality one alternative teaches us anything about other alternatives. For example, we might be comparing different configurations of an assembly line to find which one maximizes throughput, or comparing different queueing disciplines in a call center. In this situation, we refer to the alternatives with arbitrary integers, so $\mathscr{X} = \{1,2,\ldots,k\}$. Typically, the number of alternatives k is small ($k \leq 1000$).

In OvS, the alternatives have well-defined relationships with each other, or are ordered in a meaningful way, and observing one system teaches us about other systems. For example, we might be optimizing over the numbers of doctors and nurses within a hospital ward, where the average patient waiting times at similar staffing levels are likely to be similar. We will assume that the set of alternatives $\mathscr{X}$ is a finite subset of $\mathbb{Z}^d$, which is a common situation in OvS. Typical OvS problems have many more alternatives (thousands, tens of thousands, or even millions) than R&S problems, and methods for solving OvS problems explicitly use the relationships between the alternatives.

While the author believes that the above distinctions between R&S and OvS is commonly understood within the simulation community, this distinction is not observed uniformly within the literature. For example, Frazier, Powell, and Dayanik (2009) uses the term R&S to refer to problems with millions of alternatives with well-defined relationships to each other.

From a Bayesian viewpoint, the critical difference between R&S and OvS is in the prior distribution. In a R&S problem, we use an independent prior distribution (Section 4), in which information about one alternative does not affect our belief about other alternatives. In an OvS problem, we use a correlated prior distribution (Section 7), which incorporates the relationship between the alternatives, and allows information about one alternative to alter our beliefs about closely related alternatives. Algorithms that use such correlated prior information can learn more with fewer samples.

In this tutorial, we consider a single output measure, and assume that the distribution of a sample of this output measure is normal. We refer to the mean and variance of the sampling distribution as the *sampling mean* and *sampling variance* and write them as $\theta(x)$ and $\lambda(x)$ respectively to indicate that they depend on the input x. We use the prefix "sampling" in these two terms to distinguish them from the means and variances of the Bayesian prior and posterior distributions, which will be discussed below. Thus, we can write the distribution of a sample $y(x)$ obtained from alternative x as

$$y(x) \sim \mathcal{N}(\theta(x), \lambda(x)).$$

Our goal in both R&S and OvS is to find the alternative x with the largest $\theta(x)$, i.e., to find

$$x_* \in \operatorname*{argmax}_x \theta(x).$$

Although estimation of x_* from observed simulation replications can be non-trivial in the OvS setting, the main question with which methods for R&S and OvS grapple is how to allocate sampling effort across the alternatives to best support this final determination. It is on this question that we will ultimately focus in this tutorial, but first we must describe the basics of Bayesian inference.

3 BAYESIAN INFERENCE: OVERVIEW

The central idea in Bayesian statistics is that any quantity whose value is unknown (for us, the sampling mean $\theta(x)$) can be treated by placing a probability distribution over the range of values that this unknown quantity can take. This probability distribution should correspond, as well as possible, to the simulation analyst's initial belief about the likelihoods of various values. This probability distribution is called the Bayesian *prior probability distribution*.

For example, if we are interested in the long-run average patient waiting time as could be simulated for a hospital ward, these probabilities would arise in the same spirit as statements like "I think it is likely that the average waiting time is less than 2 hours," and "I think it is unlikely that the average waiting time is less than 15 minutes." The prior probability distribution quantifies such statements.

After observing data, which in our case will be the output of our simulation, we calculate the conditional distribution of the unknown quantity of interest given the data, using Bayes rule. This conditional distribution is called the *posterior probability distribution*, or simply the *posterior*, and it quantifies our new belief about the unknown quantity given the newly observed data. If we then observe more data, we condition on this additional data, again using Bayes rule, to obtain an updated posterior distribution. As we observe more and more data, the posterior distribution concentrates at the true value of $\theta(x)$.

When faced with a decision (in our problem, we must decide how to allocate simulation effort, and on our final estimate of x_*), the Bayesian approach is to construct a utility function giving the utility of each possible decision under each possible state of the world, to compute the expected utility of each possible decision under the posterior distribution, and to choose the decision with the largest expected utility.

Many authors have argued for or against the Bayesian approach. The main objection to Bayesian methodology is that it is subjective, in the sense that the answers and decisions one gets from a Bayesian method depend on the prior distribution used. Responding to this objection are results on consistency of Bayesian estimates (see, e.g., Ghosh and Ramamoorthi (2003)) showing that as we obtain more data, Bayesian estimates depend less and less on the prior, and (under weak technical conditions) converge to the correct values. Also in response are analyses showing that, if a decision-maker's decisions satisfy certain apparently reasonable axioms, then those decisions must be consistent with a Bayesian analysis (see Chapter 4.1.IV of Berger (1985)). While a detailed review of the arguments for and against Bayesian methods is outside the scope of this article, we refer the interested reader to Chick (2005) and Berger (1985).

An important practical question is how one should choose the prior distribution. One approach is to find a functional form for the prior distribution that makes analysis convenient (in the case discussed below, this

functional form is the normal distribution), and then to choose the free parameters in this functional form based on one's best judgment and discussions with those familiar with the simulation and the real system it simulates. A second approach is to specify a *non-informative prior distribution*, which corresponds to specifying as little information in the prior as possible. Both approaches are discussed below in the context of the particular statistical model that we consider in detail.

4 BAYESIAN INFERENCE: KNOWN VARIANCE AND INDEPENDENT PRIOR

We now proceed to our detailed discussion of inferring the values of $\theta(x)$ using Bayesian statistics. We use a prior distribution appropriate for R&S problems. Because of its simplicity, we focus on the case where the sampling variances $\lambda(x)$ are known, and discuss the case with unknown sampling variance briefly at the end of the section. The material in this section is discussed in most textbooks on Bayesian statistics, e.g., Gelman et al. (2004) and DeGroot (1970).

Because the alternatives are modeled as unrelated in R&S (see Section 2), our prior distribution on each $\theta(x)$ is independent across x. Thus, if we later learn something about $\theta(x)$, this will not affect our posterior distribution on $\theta(x')$, where $x' \neq x$. For computational tractability, we further assume that the functional form of our prior distribution on each $\theta(x)$ is normal. This prior distribution can be written,

$$\theta(x) \sim \mathcal{N}(\mu_0(x), \sigma_0^2(x)), x = 1, 2, \ldots, k,$$

with independence across x, for some $\mu_0(x) \in \mathbb{R}$ and $\sigma_0^2(x) \in [0, \infty]$. The special case of $\sigma_0^2(x) = 0$ corresponds to knowing perfectly that $\theta(x) = \mu_0(x)$, and $\sigma_0^2(x) = \infty$ corresponds to a non-informative prior, which can be understood as corresponding to our knowing nothing about $\theta(x)$.

Suppose we take n_x samples from alternative x, and let $\bar{y}(x)$ be the average of these samples, so $\bar{y}(x) \sim \mathcal{N}(\theta(x), \lambda(x)/n_x)$. Using Bayes rule, we can calculate the density of the conditional distribution of $\theta(x)$ given $\bar{y}(x)$,

$$P(\theta(x) = u | \bar{y}(x)) = \frac{P(\bar{y}(x) | \theta(x) = u) P(\theta(x) = u)}{\int_{\mathbb{R}} P(\bar{y}(x) | \theta(x) = v) P(\theta(x) = v) dv}.$$

Through brute-force computation (see, e.g, Gelman et al. (2004)), the right-hand side can be shown to be equal to the density of another normal random variable, which has a different mean and variance. Specifically, the conditional distribution of $\theta(x)$ given the data $\bar{y}(x)$ is

$$\begin{aligned} \theta(x) |\ \bar{y}(x), n_x &\sim \mathcal{N}(\mu_1(x), \sigma_1^2(x)), \\ \mu_1(x) &= \frac{\sigma_0^{-2}(x)\mu_0(x) + n_x \lambda^{-1}(x) y(x)}{\sigma_0^{-2}(x) + n_x \lambda^{-1}(x)} \\ \sigma_1^2(x) &= \left[\sigma_0^{-2}(x) + n_x \lambda^{-1}(x)\right]^{-1}. \end{aligned} \tag{1}$$

This conditional distribution on $\theta(x)$ is called the *posterior distribution* on $\theta(x)$.

The mean of the posterior distribution (called the *posterior mean*) on $\theta(x)$ is a weighted average between the prior mean $\mu_0(x)$ and the data $\bar{y}(x)$. The weight on the prior mean is proportional to $1/\sigma_0^2(x)$, and the weight on the data is proportional to $n_x/\lambda(x)$. If our prior has a big variance, then this says that we have a relatively weak prior belief and we put much more weight on the data. If the prior has a small variance, then we have a strong prior belief and we put more weight on the prior. If n_x is bigger, or if the sampling variance $\lambda(x)$ is smaller, then we put more weight on the observation $\bar{y}(x)$, and decrease the posterior variance. As n_x increases, the posterior mean converges to the sample average $\bar{y}(x)$. This is an example of the broad statement made above that Bayesian methods put more weight on the data and less weight on the prior as we obtain more data.

If our prior distribution was non-informative, so $\sigma_0^2(x) = \infty$ and $\sigma_0^{-2}(x) = 0$, then (1) provides $\mu_1(x) = \bar{y}(x)$ and $\sigma_1^2(x) = \lambda(x)/n_x$. With this choice for the prior, the prior mean does not affect the posterior distribution.

The posterior distribution can also be computed recursively, by adding one sample at a time to the set of random variables upon which we condition. This recursive computation is quite useful later in Section 6 when analyzing sequential algorithms.

Let $x_1, x_2, \ldots, x_N$ be a sequence of alternatives that were sampled, so $\sum_{n=1}^{N} 1_{\{x_n=x\}} = n_x$ for each x, and let $y_1(x_1), y_2(x_2), \ldots, y(x_N)$ be the sequence of corresponding observed values. Then, one can use induction and a computation similar to the one used to obtain (1) to show that the posterior distribution given $x_1, \ldots, x_n$ and $y_1(x_1), \ldots, y_n(x_n)$ is

$$\begin{aligned}
&\theta(x)|x_1, x_2, \ldots, x_n, y_1(x_1), y_2(x_2), \ldots, y_n(x_n) \sim \mathcal{N}(\mu_n(x), \sigma_n^2(x)), \\
&\mu_n(x) = \begin{cases} \frac{\sigma_{n-1}^{-2}(x)\mu_{n-1}(x) + \lambda^{-1}(x) y_n(x_n)}{\sigma_0^{-2}(x) + \lambda^{-1}(x)} & \text{if } x_n = x, \\ \mu_{n-1}(x) & \text{if } x_n \neq x, \end{cases} \\
&\sigma_n^2(x) = \begin{cases} \left[\sigma_{n-1}^{-2}(x) + \lambda^{-1}(x)\right]^{-1} & \text{if } x_n = x, \\ \sigma_{n-1}^2(x) & \text{if } x_n \neq x. \end{cases}
\end{aligned} \tag{2}$$

This preceding analysis assumed that the sampling variance $\lambda(x)$ was known, while in practice this is never the case. One comman approach is to maintain an adaptively updated point estimate $\hat{\lambda}_n(x)$ based on the available data $(x_1, \ldots, x_n, y_1(x_1), \ldots, y_n(x_n))$. The posterior can then be computed from the prior using (1) under the assumption that the current point estimate is correct. Although we describe this approximate approach here, it does not adhere to the Bayesian philosophy, and gives only approximations to a posterior distribution derived in a more principled way.

In some cases this approximate approach works well and is desirable for its expediency. In other cases, however, this approach produces misleading results because it ignores the uncertainty in our estimate of the variance. It is possible to include this uncertainty into our analysis by placing a prior on both the sampling mean *and* the sampling variance. The prior on the sampling variance is often an inverse-gamma distribution for tractability. Bayesian inference in such settings is described in DeGroot (1970), and many other Bayesian references. For the use of this approach in R&S, see Chick and Inoue (2001), Chick, Branke, and Schmidt (2010) and Frazier and Powell (2008).

5 ONE-STAGE METHODS

We now apply Bayesian inference to the R&S problem. We first consider a so-called *one-stage* or *batch* situation in which we make a decision at a single point in time about which simulation replications to obtain. These simulation replications will then be used to update our prior distribution to obtain a posterior, which we will then use to estimate the best alternative. We determine the *optimal* set of replications to make at this single point in time.

We use the independent normal prior on the sampling means and known sampling variance from Section 4, and independent samples, and assume a sampling budget of N samples. This setting is:

Time 0: Begin with the prior distribution on the sampling means $\theta(x)$, $x = 1, \ldots, k$,

$$\theta(x) \sim \mathcal{N}(\mu_0(x), \sigma_0^2(x))$$

Time 1:

(a) Choose a number of samples n_x to take from each alternative x, such that $\sum_x n_x \leq N$. This is called the *allocation decision*.

(b) For each alternative x, observe $\bar{y}(x)$, which is the average of n_x independent replications of $\mathcal{N}(\theta(x), \lambda(x))$.

(c) Calculate the posterior distribution on each $\theta(x)$, $x = 1,\ldots,k$, as

$$\theta(x) \mid \bar{y}(x), n_x \sim \mathcal{N}(\mu_1(x), \sigma_1^2(x)), \tag{3}$$

where $\mu_1(x)$ and $\sigma_1^2(x)$ are given by (2).

Time 2: Choose an alternative $\widehat{x}_*$ as our estimate of which alternative is the best, $\operatorname{argmax}_x \theta(x)$. This is called the *implementation decision*. Earn a reward $R = \theta(\widehat{x}_*)$ equal to the true value of the chosen alternative.

The prior distribution used at time 0 might be obtained as a true prior distribution based on the analyst's belief. However, the more typical case is that the practitioner decides on the prior distribution by taking an initial stage of measurements before time 0, say n_0 from each alternative (a common choice for n_0 is 10), sets the prior mean $\mu_0(x)$ to the sample average of the samples from alternative x, and sets $\sigma_0^2(x) = \lambda(x)/n_0$. This corresponds to beginning with a non-informative belief before time 0, and then setting the prior mean and variance μ_0, σ_0^2 to the mean and variance that would result from combining this non-informative belief with the n_0 samples from each alternative. This computation is as discussed in Section 4. When the prior is computed in this way using an initial stage of samples, the resulting procedure would be called a *two-stage* procedure.

Our goal in this section is to choose the allocation decision (made at time 1), and the implementation decision (made at time 2), to maximize in expectation the reward that we obtain at time 2. Although the reward is a function of only the implementation decision and θ, perhaps suggesting that the allocation decision does not matter, the information upon which the implementation decision is based depends upon the allocation decision.

5.1 Implementation Decision

In our analysis, we first track how the conditional distribution of $\theta(x)$ varies from time 0 to time 1, and show how this determines the optimal implementation decision at time 2. At time 0, in our Bayesian model, each $\theta(x)$ is drawn at random from the prior distribution. Once drawn, each $\theta(x)$ remains fixed throughout the rest of the steps. At time 0, because we have no information about $\theta(x)$ other than that it was drawn from the prior, our probability distribution over its values, given what we have observed (nothing), is simply the prior. At time 1, we have some additional information about $\theta(x)$ through observation of $\bar{y}(x)$, and our probability distribution over the values of $\theta(x)$ is now the conditional distribution given $\bar{y}(x)$ in (3), also called the posterior distribution. It is important to emphasize that, while the conditional distribution of $\theta(x)$ changes, $\theta(x)$ itself remains fixed. What changes is the information that we have about $\theta(x)$.

With this understanding of the dynamics of our conditional distribution on $\theta(x)$, we can now analyze how the implementation decision at time 2 should be made. Suppose we choose $\widehat{x}_* = x$ for some alternative x. Then, the expected value of the reward that we will earn, given what we know at the time we make the decision, is

$$E[\theta(x)|\bar{y}(1), n_1, \ldots, \bar{y}(k), n_k]. \tag{4}$$

The choice for $\widehat{x}_*$ that maximizes the expected reward that we will receive in time 2 (given what we know at the time we actually make this choice) is thus the alternative with the largest such conditional expected value, $\widehat{x}_* \in \operatorname{argmax}_{x=1,2,\ldots,k} E[\theta(x)|\bar{y}(1), n_1, \ldots, \bar{y}(k), n_k]$. We can write this decision more simply using the fact that $E[\theta(x)|\bar{y}(1), n_1, \ldots, \bar{y}(k), n_k]$ is the posterior mean of $\theta(x)$, for which we have the notation $\mu_1(x)$. Thus, our optimal implementation decision is

$$\widehat{x}_* \in \operatorname*{argmax}_{x=1,2,\ldots,k} \mu_1(x). \tag{5}$$

This recovers the well-known result (see, e.g., Berger (1985)) from Bayesian decision-theory that the optimal action to take is the one with the largest expected utility under the posterior distributions.

5.2 Allocation Decisions

We would now like to find the choice for $n_1,\ldots,n_k$ made in time 1 that will maximize the expected reward that we receive, $E[\theta(\widehat{x}_*)|n_1,\ldots,n_k]$. We have two options available to us for solving this problem.

The first way would be to build a Monte Carlo simulation that can evaluate the quality of a particular choice of $n_1,\ldots,n_k$. In each replication, this simulation would choose $\theta(1),\ldots,\theta(k)$ at random from the prior distribution, then simulate each $\bar{y}(x)$ based on n_x and the previously simulated value of $\theta(x)$, then choose $\widehat{x}_*$ according to (5), and would then provide the reward $\theta(\widehat{x}_*)$. Running many replications would then provide an estimate of the expected reward provided by that $n_1,\ldots,n_k$, where we average over both the simulation noise in $\bar{y}(x)$ and the different problem instances corresponding to different θ drawn from the prior. The best $n_1,\ldots,n_k$ could then be selected using an existing method for R&S, either frequentist or Bayesian (we would be using R&S to design a R&S algorithm!).

There is a second method for solving this problem that is more difficult to understand, but more efficient. This second method can be understood as marginalizing over the random variables $\theta(x)$ analytically, to estimate the expected reward $E[\theta(\widehat{x}_*)|n_1,\ldots,n_k]$ more precisely. This second method is also a stepping stone to analysis of sequential algorithms, considered in Section 6. We now consider this method in detail.

The first step is to write the conditional expectation of our final reward given what we know when we make the decision $\widehat{x}_*$ in time 2. This conditional expected value is the value (4) of the $\widehat{x}_* = x$. Call this value V_1. Since $\widehat{x}_*$ is the x with the largest value of $\mu_1(x)$, V_1 can be written

$$V_1 = \max_{x=1,\ldots,k} \mu_1(x).$$

We think of V_1 as the expected value of the best final decision rule, given what we know at time 1.

Now, at time 0, before the samples are observed, V_1 is random because it depends on the $\mu_1(x)$, which are themselves random. This randomness includes simulation noise, but also the randomness introduced by our lack of knowledge about the $\theta(x)$. We must determine the distribution of the $\mu_1(x)$, and the dependence of this distribution on $n_1,\ldots,n_k$. The distribution of $\mu_1(x)$ at time 0 is called the *pre-posterior distribution* in Bayesian statistics (DeGroot 1970; Raiffa and Schlaifer 1968), because it is the distribution of a parameter of the posterior, conditioning only on what is known before we observe the data upon which the posterior is based. Once we determine the distribution of $\mu_1(x)$, and from it the distribution of V_1, the optimal choice of $n_1,\ldots,n_k$ will be the one that maximizes the expected value of V_1.

Pick any x. It turns out that the distribution of $\mu_1(x)$ depends only on $\mu_0(x),\sigma^2(x),n_x$, and not on the values for other $x' \neq x$. To determine the distribution of $\mu_1(x)$, we have that

$$\begin{aligned}\theta(x) &\sim \mathcal{N}(\mu_0(x),\sigma_0^2(x))\\ \bar{y}(x)|\theta(x) &\sim \mathcal{N}(\theta(x),\lambda(x)/n_x),\end{aligned}$$

so if we let $\varepsilon(x) = \bar{y}(x) - \theta(x)$, then $\varepsilon(x)$ is independent of $\theta(x)$ and is normal with mean 0 and variance $\lambda(x)/n_x$. Thus, $\bar{y}(x) = \theta + \varepsilon(x)$ is the sum of two independent normal random variables, and is itself normal with a mean equal to the sum of the means, $\mu_0(x)+0$, and a variance equal to the sum of the two variances, $\sigma_0^2(x)+\lambda(x)/n_x$. From (2), $\mu_1(x)$ is a linear function of $\bar{y}(x)$, and so must itself be normally distributed as a linear function of a normal random variable is also normal. We can compute the mean and variance of $\mu_1(x)$ by direct computation from (2). This provides the sought-after distribution of $\mu_1(x)$,

$$\mu_1(x) \sim \mathcal{N}\left(\mu_0(x),\sigma_0^2(x)-\sigma_1^2(x)\right), \tag{6}$$

where $\sigma_1^2(x)$ is given by (2). Note that $\sigma_1^2(x)$ does not depend on the sampled value $\bar{y}(x)$ and can be computed without observing this sampled value.

The mean and variance of the distribution of $\mu_1(x)$ can be derived in a much more general way. First, the mean can be computed using the tower property of conditional expectation as $E[\mu_1(x)] = E[E[\theta(x)|n_x,\bar{y}(x)]] = E[\theta(x)] = \mu_0(x)$. Second, the variance can be computed using the conditional variance

formula as $\text{Var}[\mu_1(x)] = \text{Var}[E[\theta(x)|n_x,\bar{y}(x)]] = \text{Var}[\theta(x)] - E[\text{Var}[\theta(x)|n_x,\bar{y}(x)]] = \sigma_0^2(x) - E[\sigma_1^2(x)] = \sigma_0^2(x) - \sigma_1^2(x)$, where we have used that $\sigma_1^2(x) = \text{Var}[\theta(x)|n_x,\bar{y}(x)]$ is actually not random at time 0, as it does not depend on $\bar{y}(x)$.

With the distribution of $\mu_1(x)$ in hand, we can now write the expected reward that we will receive from a given choice of $n_1,\ldots,n_k$ as the expected value of $\max_x \mu_1(x)$ under this distribution, which can be written explicitly as an integral,

$$E\left[\max_x \mu_1(x)|n_1,\ldots,n_k\right] = \int_{\mathbb{R}^k}\left[\max_{x=1,\ldots,k}\mu_1(x)\right]\left[\prod_{x=1}^{k}\varphi(\mu_1(x);\mu_0(x),\sigma_0^2(x)-\sigma_1^2(x))\right]d\mu_1(1)\cdots d\mu_1(k)$$

where $\varphi(z;\mu,\sigma^2)$ is the density at z of a normal random variable with mean μ and variance σ^2.

The optimal choice for $n_1,\ldots,n_k$ is then the one that gives the largest expected reward,

$$\underset{(n_1,\ldots,n_k)\in\mathbb{Z}_+^k:\sum_x n_k\leq N}{\text{argmax}} E\left[\max_x \mu_1(x)|n_1,\ldots,n_k\right]$$

In general, this is a difficult optimization problem to solve, as evaluating the objective requires either performing a high-dimensional integral or using simulation, the objective is non-concave (Frazier and Powell 2010), and we have integrality constraints. Chick and Inoue (2001) performs a more general analysis for a version of this problem with unknown sampling variance, and computes an approximate analytic solution, which is called the LL(B) policy.

In the special case that $N = 1$, so only 1 sample is being allocated, the integral defining $h(n_1,\ldots,n_k)$ can be evaluated analytically (Gupta and Miescke 1996). Let x be the single alternative to which a sample is being allocated. Then, the integral becomes

$$E\left[\max_x \mu_1(x)|n_x = 1, n_{x'} = 0\ \forall x' \neq x\right] = \left[\max_{x'}\mu_0(x')\right] + \widetilde{\sigma}(x) f\left(-\frac{|\Delta(x)|}{\widetilde{\sigma}(x)}\right), \tag{7}$$

where

$$\widetilde{\sigma}(x) = \sqrt{\sigma_0^2(x)-\sigma_1^2(x)} = \sqrt{\sigma_0^2(x) - \left[\sigma_0^{-2}(x)+\lambda^{-1}(x)\right]^{-1}}$$

is the standard deviation of the distribution of $\mu_1(x)$, $\Delta(x) = \mu_0(x) - \max_{x'\neq x}\mu_0(x')$ is the difference in expected value between alternative x and the best of the remaining alternatives, and $f(z) = z\Phi(z)+\varphi(z)$. Φ is the standard normal cdf and φ is the standard normal pdf. Although it is rare that we encounter a R&S problem in which only one sample is to be allocated overall, this expression is very useful in a fully sequential method discussed below.

6 SEQUENTIAL METHODS

In the previous section, we considered how a collection of simulation replications should be allocated in a R&S problem, if they are to be allocated all at once. In many applications, however, we can allocate our simulation budget a little bit at a time, in a fully sequential algorithm. In such algorithms, we allocate one sample, observe the response, then based on this response allocate another sample. This process is:

Time 0: Let $n = 0$. We begin with our prior distribution on all of the sampling means $\theta(x)$, $x = 1,\ldots,k$,

$$\theta(x) \sim \mathcal{N}(\mu_0(x),\sigma_0^2(x))$$

Increment n.

Time n: While $n \leq N$,

(a) Based on the available data, summarized by the parameters μ_{n-1},σ_{n-1} of the most recent posterior distribution, choose the alternative x_n to sample next.

(b) Observe a sample from the sampling distribution of alternative x_n,

$$y_n|x_n,\theta(x_n) \sim \mathcal{N}(\theta(x_n),\lambda(x_n))$$

(c) Calculate the posterior distribution based on all of the available data

$$\theta(x) \mid x_1,\ldots,x_n,y_1,\ldots,y_n \sim \mathcal{N}(\mu_n(x),\sigma_n^2(x)), \tag{8}$$

The parameters $\mu_n(x),\sigma_n^2(x)$ can be computed recursively from $\mu_{n-1},\sigma_{n-1}^2,y_n$ using (2).

(d) Increment n.

Time N+1: Choose an alternative $\widehat{x}_*$ as our estimate of which alternative is the best, $x_* \in \operatorname{argmax}_x \theta(x)$. The optimal choice is

$$\widehat{x}_* \in \operatorname*{argmax}_x \mu_N(x)$$

We earn a reward $R = \theta(\widehat{x}_*)$.

Of course, when we have the opportunity to behave fully sequentially, we can decide up front how to allocate all of our samples, behaving in a one-step manner and ignoring the opportunity to behave sequentially. However, we may be giving up quite a bit of performance when choosing not to act in a fully sequential way. By allowing our allocation decisions to adapt to the most recent data, we have the opportunity to make allocations that more efficiently gather the information we need. Thus, while analysis of the optimal fully sequential algorithm is more difficult than for one-step algorithms, the performance boost that we receive makes this analysis worthwhile.

6.1 One-Step Approximate Methods

One approach to acting in sequential problems is to pretend, each time we take a sample x_n in Step a, that $N = n+1$ and we only have one sample remaining. Under this supposition, we can compute the optimal measurement to make using the one-step analysis from Section 5. Taking the same analysis that provided (7), but replacing $\mu_0(x)$ and $\mu_1(x)$ with $\mu_n(x)$ and $\mu_{n+1}(x)$, we find that the one-step value of choosing alternative x is

$$\left[\max_{x'} \mu_n(x')\right] + \widetilde{\sigma}_n(x) f\left(-\frac{|\Delta_n(x)|}{\widetilde{\sigma}_n(x)}\right),$$

where

$$\widetilde{\sigma}_n(x) = \sqrt{\sigma_n^2(x) - \left[\sigma_n^{-2}(x) + \lambda(x)\right]^{-1}}$$
$$\Delta_n(x) = \mu_n(x) - \max_{x' \neq x} \mu_n(x')$$

and $f(z) = z\Phi(z) + \varphi(z)$ is defined as before.

Since $[\max_{x'} \mu_n(x')]$ does not depend on the choice x of which alternative to sample, the resulting decision can be written as

$$x_n \in \operatorname*{argmax}_{x=1,\ldots,k} \widetilde{\sigma}_n(x) f\left(-\frac{|\Delta_n(x)|}{\widetilde{\sigma}_n(x)}\right).$$

This policy was introduced as the (R1,...,R1) policy by Gupta and Miescke (1996), and then analyzed within a dynamic programming framework under the name *knowledge-gradient policy* by Frazier, Powell, and Dayanik (2008).

Although the policy is not optimal in general, it turns out to be optimal in some special cases (Frazier, Powell, and Dayanik 2008), including the case when $k = 2$, where it corresponds to sampling the alternative with the largest posterior variance. It also works extremely well in the average-case, when compared with a number of other allocation algorithms (Frazier, Powell, and Dayanik 2008). This policy has been generalized to allow for a prior distribution on unknown sampling variance in Chick, Branke, and Schmidt (2010). The resulting policy is called the LL1 policy.

6.2 Optimal Methods

To find the optimal allocation policy in fully sequential settings, we must use dynamic programming. In this analysis, we calculate a sequence of functions, called *value functions*, that depend on the current point in time, and on the parameters of the current posterior distribution. The value function gives the expected value of pursuing the optimal policy, beginning from that point in time and that posterior distribution.

We actually did a simplified version of this in Section 5, where V_1 gave the conditional expectation of pursuing the optimal policy (which in this case was simply choosing $\widehat{x}_*$ optimally) forward from time 1, starting from the current posterior distribution. In that analysis, V_1 was implicitly a function of $\mu_1(1), \sigma_1(1), \ldots, \mu_1(k), \sigma_1(k)$.

We begin as we did in Section 5, where we note that the conditional expected reward of choosing implementation decision x is $E[\theta(x)|x_1, y_1, \ldots, x_N, y_N] = \mu_N(x)$, and so the optimal implementation decision is $\widehat{x}_* \in \text{argmax}_{x=1,\ldots,k} \mu_N(x)$, and the conditional expectation of the reward that we will receive from this implementation decision is $\max_{x=1,\ldots,k} \mu_N(x)$. We define a function equal to this conditional expected reward,

$$V_N(\vec{\mu}_N, \vec{\beta}_N) = \max_{x=1,\ldots,k} \mu_N(x),$$

where $\vec{\mu}_n = (\mu_n(1), \ldots, \mu_n(k))$ is the vector of posterior means and $\vec{\beta}_n = (\sigma_n^{-2}(1), \ldots, \sigma_n^{-2}(k))$ is the vector of posterior precisions (the precision is the inverse of the variance). The function V_N is called the value function at time N.

We will use this value function to decide which alternative to sample at time N, i.e., how to best choose x_N, but before doing so we adapt the pre-posterior analysis used to derive (6) to the setting where we take one measurement at time. By the same analysis used to derive (6) we have, for any $n = 1, \ldots, N$,

$$\begin{aligned}
\sigma_n^2(x) &= \sigma_{n-1}^2(x), && \text{if } x_n \neq x,\\
\mu_n(x) &= \mu_{n-1}(x), && \text{if } x_n \neq x,\\
\sigma_n^2(x) &= \left[\sigma_{n-1}^{-2}(x) + \lambda^{-1}(x)\right]^{-1}, && \text{if } x_n = x,\\
\mu_n(x) &\sim \mathcal{N}(\mu_{n-1}(x), \sigma_{n-1}^2(x) - \sigma_n^2(x)) && \text{if } x_n = x.
\end{aligned}$$

Letting $\vec{e}_x$ be the vector of all 0s, with a single 1 for entry x, and letting Z_n be an independent standard normal random variable, then we can write this previous set of equations more compactly as,

$$\begin{aligned}
\vec{\beta}_n &= \vec{\beta}_{n-1} + e_{x_n}\lambda^{-1}(x_n),\\
\vec{\mu}_n &= \vec{\mu}_{n-1} + e_{x_n}\widetilde{\sigma}(\vec{\beta}_{n-1}, x_n)Z_n,\\
\widetilde{\sigma}(\vec{\beta}_{n-1}, x_n) &= \sqrt{\sigma_{n-1}^2(x_n) - \sigma_n^2(x_n)} = \sqrt{\sigma_{n-1}^2(x) - \left[\sigma_{n-1}^{-2}(x) + \lambda^{-1}(x)\right]^{-1}}.
\end{aligned} \tag{9}$$

This set of equations tells us how the posterior distribution will change as a consequence of the decision x_n, for general n. In the dynamic program to follow (9) plays the role of the state transition equations.

We now return to the question of how to best choose x_N. From the pair of equations (9), we have the distribution of $\vec{\mu}_N, \vec{\beta}_N$ in terms of information available to us when we choose x_N (in terms of $\vec{\mu}_{N-1}, \vec{\beta}_{N-1}$, and x_N). We also have the conditional expectation of our overall reward in terms of $\vec{\mu}_N, \vec{\beta}_N$. Thus, to compute the conditional expectation of our overall reward in terms of information available to us when we choose x_N, we take the expectation as in

$$V_{N-1}(\vec{\mu}_{N-1}, \vec{\beta}_{N-1}, x_N) = \int_{\mathbb{R}} V_N\left(\vec{\mu}_{N-1} + e_{x_N}\widetilde{\sigma}(\vec{\beta}_{N-1}, x_N)z, \vec{\beta}_{N-1} + e_{x_N}\lambda^{-1}(x_N)\right)\varphi(z)\,dz,$$

where we have defined a new function $V_{N-1}(\vec{\mu}_{N-1}, \vec{\beta}_{N-1}, x_N)$ to be equal to this expectation (expressed as an integral). This new function is called the post-decision-state value function (Powell 2007), because it gives us the value after we have made the decision.

Now, since $V_{N-1}(\vec{\mu}_{N-1},\vec{\beta}_{N-1},x)$ gives us the expected value that results from choosing $x_N = x$, the optimal choice for x_N is simply

$$x_N \in \operatorname*{argmax}_{x=1,\ldots,k} V_{N-1}(\vec{\mu}_{N-1},\vec{\beta}_{N-1},x).$$

and the value that is obtained from this choice is

$$V_{N-1}(\vec{\mu}_{N-1},\vec{\beta}_{N-1}) = \max_{x=1,\ldots,k} V_{N-1}(\vec{\mu}_{N-1},\vec{\beta}_{N-1},x),$$

where we have defined a new function V_{N-1}, which is the value function at time $N-1$, and which gives us the conditional expected value of our final reward, given our posterior distribution $\vec{\mu}_{N-1},\vec{\beta}_{N-1}$.

We can repeat this process recursively, moving backward from $n = N$ to $n = 1$, as in

$$\begin{aligned} V_n(\vec{\mu}_n,\vec{\beta}_n,x_{n+1}) &= \int_{\mathbb{R}} V_{n+1}(\ \vec{\mu}_n + e_{x_{n+1}}\widetilde{\sigma}(\vec{\beta}_n,x_{n+1})z, \vec{\beta}_n + e_{x_{n+1}}\lambda^{-1}(x))\varphi(z)\,dz \\ x_{n+1} &\in \operatorname*{argmax}_{x=1,\ldots,k} V_n(\vec{\mu}_n,\vec{\beta}_n,x), \\ V_n(\vec{\mu}_n,\vec{\beta}_n) &= \max_{x=1,\ldots,k} V_n(\vec{\mu}_n,\vec{\beta}_n,x), \end{aligned} \tag{10}$$

This provides the optimal allocation decision at each point in time n, as a function of the posterior distribution.

The equations (10), which are Bellman's optimality equations for our problem, specify the value functions at time n in terms of the value functions at time $n+1$, and so in principle offer an explicit algorithm for computing the value functions and with them the optimal allocation decisions. However, in problems with large values of k, this is computationally challenging because of the so-called curse of dimensionality: the value functions must be computed for every possible pair of values for the vectors $\vec{\mu}_n,\vec{\beta}_n$, as well as at every time n, and the computational cost of doing so scales exponentially in the dimension of these vectors.

Dynamic programming has been used computationally to completely solve versions of the R&S problem with smaller values of k: Chick and Frazier (2012) and Chick and Gans (2009) solve slightly modified versions of the problem we have presented (they use a cost per sample and/or discounting in time of the final reward, rather than a fixed sampling budget) with one unknown alternative, and another alternative of fixed value. In some cases the dynamic program can be solved analytically, as done by Frazier, Powell, and Dayanik (2008) for certain special cases. For example, this dynamic programming analysis can be used to show that the average-case optimal policy for problems with $k = 2$ alternatives is to measure the alternative with the larger posterior variance. In other cases, the dynamic program can be analyzed theoretically and decomposed into a number of smaller dynamic programs that can be solved computationally. This is the approach of Xie and Frazier (2011), which computes average-case-optimal policies for the related problem of multiple comparisons with a known standard. Perhaps the most immediate practical benefit of this type of dynamic programming analysis is in the development of heuristics with better performance than the one-step heuristics discussed in the previous section. Chick and Frazier (2012) use this approach, developing methods with the currently best known average-case performance.

7 LARGE-SCALE PROBLEMS USING CORRELATED PRIORS

In this final section, we consider OvS problems in which alternatives correspond to points on an integer lattice, e.g., an alternative $x = (x^{(1)},x^{(2)})$ is the number of doctors $x^{(1)}$ and nurses $x^{(2)}$ that staff a hospital ward. In such situations, it is common that $\theta(x)$ and $\theta(z)$ will take similar values when x and z are close to each other. Incorporating this information into our algorithm greatly improves performance.

Space limitations require us to treat this topic only briefly. We refer the interested reader to Frazier, Powell, and Dayanik (2009) upon which this analysis is based, to more recent work (Frazier, Xie, and

Chick 2011) that address computational challenges, and to related work on Bayesian methods for global optimization of expensive deterministic computer codes (Brochu, Cora, and de Freitas 2009).

7.1 Bayesian Inference

We can incorporate the information that alternatives are related to each other into our Bayesian model through a *correlated* multivariate normal prior on the vector $\theta = (\theta(x) : x \in \mathscr{X})$. This correlated prior is really a Gaussian process prior (Rasmussen and Williams 2006) over a discrete set, where Gaussian process priors, and the related non-Bayesian technique of kriging, is an approach commonly employed for inferring the value of the output of (stochastic) simulation and deterministic computer codes, in applications outside of OvS. We provide a brief overview of how this is accomplished, but space prevents us from discussing many important practical details in how the prior should be chosen. We point the interested reader to the textbook Rasmussen and Williams (2006), and to work on kriging (Cressie 1993) and stochastic kriging (Ankenman et al. 2008; Ankenman et al. 2010).

To create an appropriate correlated multivariate normal prior on $\vec{\theta}$, we first choose a covariance function $\Sigma_0(\cdot,\cdot)$ that describes how closely related two alternatives are, as a function of their locations in the lattice. One common (though not always best) choice is

$$\Sigma_0(x,z) = \alpha_0 \exp\left(-\sum_{i=1}^{d} \alpha_i |x^{(i)} - z^{(i)}|^p\right)$$

where $\alpha_0, \alpha_1, \ldots, \alpha_d > 0$ and $p \in [1,2]$ are free parameters. This covariance function decreases as x and z move away from each other, with the speed of the drop controlled by the choice of the free parameters. We also choose a mean function, $\mu_0(\cdot)$, for which a common choice is

$$\mu_0(x) = \beta_0 + \sum_{\ell=1}^{L} \beta_\ell f_\ell(x),$$

where $\beta_0, \ldots, \beta_L$ are again free parameters and $f_1(\cdot), \ldots, f_L(\cdot)$ are basis functions.

Then, we define a vector $\vec{\mu}_0 = (\mu_0(x) : x \in \mathscr{X})$ and a matrix $\Sigma_0 = (\Sigma_0(x,z) : x,z \in \mathscr{X})$, and our prior distribution is multivariate normal,

$$\vec{\theta} \sim \mathscr{N}(\vec{\mu}_0, \Sigma_0).$$

The prior has a number of free parameters within it. The most common approach for setting these free parameters is to sample $y(x)$ at a few points x chosen uniformly at random, and to calculate maximum likelihood estimates of the free parameters. These maximum likelihood estimates can be updated adaptively, as more data are collected. A more principled approach is to put higher-level prior distributions on these parameters, resulting in a prior that is an mixture of multivariate normal distributions. The details of such computations are discussed in Rasmussen and Williams (2006).

Given this prior distribution, if we then observe samples at a sequence of points of our choosing, the posterior distribution will again be multivariate normal, but with a different mean vector and covariance matrix. After n samples, call the mean vector $\vec{\mu}_n$ and the covariance matrix Σ_n, so

$$\vec{\theta} | x_1, y_1(x_1), \ldots, x_n, y_n(x_n) \sim \mathscr{N}(\vec{\mu}_n, \Sigma_n).$$

This mean vector and covariance matrix can be computed recursively as,

$$\Sigma_n = \left[\Sigma_{n-1}^{-1} + \lambda(x_n)^{-1} \vec{e}_{x_n} \vec{e}_{x_n}^T\right]^{-1}, \quad \mu_n = \Sigma_n \left[\Sigma_{n-1}^{-1} \vec{\mu}_{n-1} + \lambda(x_n)^{-1} y_n(x_n) \vec{e}_{x_n}\right].$$

In problems where the number of alternatives in $\mathscr{X}$ is much larger than the number of alternatives that can be measured, it is more efficient to compute entries of the posterior distribution on the fly, rather than storing the full matrix Σ_n in memory. This is described in standard references on Gaussian process regression, such as Rasmussen and Williams (2006).

7.2 Allocation Decision

While these methods for Bayesian inference with correlated priors are substantially more complicated and computationally intensive than in the independent prior case appropriate for R&S, we can apply the same type of analysis to decide where to sample next. Here, we describe a one-step approximate method called the *knowledge-gradient policy for correlated beliefs* (Frazier, Powell, and Dayanik 2009).

This method can be used in a fully sequential setting, when we take one measurement at a time. To derive the allocation decision of this method, we imagine that $N = n$, and so after taking the next measurement we will be forced to make an implementation decision. The implementation decision that we will make, in this hypothetical situation, is $\operatorname{argmax}_{x\in\mathscr{X}} \mu_n(x)$ and its conditional expected value is $\max_{x\in\mathscr{X}} \mu_n(x)$.

One can show (Frazier, Powell, and Dayanik 2009), using a pre-posterior analysis similar to the one used earlier, that the conditional distribution of the vector $\vec{\mu}_n$, given $\vec{\mu}_{n-1}, \Sigma_{n-1}$ and x_n, is the same as

$$\vec{\mu}_{n-1} + \widetilde{\sigma}(\Sigma_{n-1}, x_n) Z_n$$

where Z_n is a scalar standard normal, and $\widetilde{\sigma}(\Sigma, x)$ is the vector $\widetilde{\sigma}(\Sigma, x) = \Sigma \vec{e}_x / \sqrt{\lambda(x) + \Sigma_{xx}}$. Thus, the conditional expected reward that we will receive if we choose to measure a particular alternative x_n is

$$\begin{aligned} \mathrm{KG}_n(x_n) &= E\left[\max_{x\in\mathscr{X}} \mu_n(x) | \vec{\mu}_{n-1}, \Sigma_{n-1}, x_n\right] \\ &= \int_{\mathbb{R}} \left[\max_{x\in\mathscr{X}} \mu_{n-1}(x) + \widetilde{\sigma}_x(\Sigma, x_n) z\right] \varphi(z)\, dz. \end{aligned}$$

A method for computing this integral analytically is given in Frazier, Powell, and Dayanik (2009). This quantity $\mathrm{KG}_n(x_n)$ is called the KG factor, and the KG policy is then to measure the alternative x_n for which this KG factor is largest,

$$x_n \in \operatorname*{argmax}_{x\in\mathscr{X}} \mathrm{KG}_n(x).$$

8 CONCLUSION

In this tutorial, we have provided an overview of Bayesian methods for optimization via simulation and ranking and selection. The analysis described provides a theoretical characterization, in terms of the solution of a dynamic program, of methods with optimal average-case performance. In some cases this dynamic program can be solved to give a practical implementation of this average-case optimal algorithm, while in other cases it provides a framework for creating heuristic procedures with good average-case performance.

While we have restricted our attention to standard formulations of OvS and R&S, the general roadmap that we describe can be applied to a much greater number of problems within simulation. For example, the same ideas apply to versions of optimization via simulation with common random numbers, steady-state simulation, or stochastic constraints. They also apply to other types of analysis methodology, such as screening, stochastic root-finding, and multiple comparisons with a standard.

We hope that this tutorial inspires the reader to go further into the literature on Bayesian methods and their application within analysis methodology, and to apply these powerful and flexible methods to their own problems.

ACKNOWLEDGMENTS

The author was supported by AFOSR YIP FA9550-11-1-0083.

REFERENCES

Andradóttir, S. 1998. "Simulation optimization". In *Handbook of simulation: Principles, methodology, advances, applications, and practice*, 307–333. New York: Wiley-Interscience.

Ankenman, B., B. L. Nelson, and J. Staum. 2008, December. "Stochastic kriging for simulation metamodeling". 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, 362–370. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58 (2): 371–382.

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

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

Berger, J. O. 1985. *Statistical decision theory and Bayesian analysis*. Second ed. New York: Springer-Verlag.

Berry, D. A., and B. Fristedt. 1985. *Bandit Problems: Sequential Allocation of Experiments*. London: Chapman & Hall.

Bertsekas, D. P. 2005. *Dynamic programming and optimal control, vol. I*. third ed. Athena Scientific.

Bertsekas, D. P. 2007. *Dynamic programming and optimal control, vol. II*. third ed. Athena Scientific.

Brochu, E., M. Cora, and N. de Freitas. 2009, November. "A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning". Technical Report TR-2009-023, Department of Computer Science, University of British Columbia.

Chick, S. 2005. *Subjective probability and Bayesian methodology*. Handbooks in Operations Research and Management Science. North-Holland.

Chick, S. E. 2000, December. "Bayesian methods: bayesian methods for simulation". In *Proceedings of the 2000 Winter Simulation Conference*, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 109–118. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chick, S. E. 2006, December. "Bayesian ideas and discrete event simulation: why, what and how". 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, 96–105. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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

Chick, S. E., and P. I. Frazier. 2012. "Sequential Sampling for Selection with Economics of Selection Procedures". *Management Science* 58 (3): 550–569.

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

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.

Cressie, N. 1993. *Statistics for Spatial Data, revised edition*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: Wiley Interscience.

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

Dynkin, E., and A. Yushkevich. 1979. *Controlled Markov Processes*. New York: Springer.

Frazier, P. I. 2010. *Wiley Encyclopedia of Operations Research and Management Science*, Chapter Decision-theoretic Foundations of Simulation Optimization. Wiley.

Frazier, P. I., and W. B. 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. I., and W. B. Powell. 2010. "Paradoxes in Learning and the Marginal Value of Information". *Decision Analysis* 7 (4): 378–403.

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

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

Frazier, P. I., J. Xie, and S. E. Chick. 2011, December. "Value of Information Methods for Pairwise Sampling with Correlations". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Fu, M. 1994. "Optimization via simulation: A review.". *Ann. Oper. Res.* 53 (1): 199–248.

Fu, M. 2002. "Optimization for simulation: Theory vs. practice". *INFORMS Journal on Computing* 14 (3): 192–215.

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

Ghosh, J. K., and R. V. Ramamoorthi. 2003. *Bayesian Nonparametrics*. Springer.

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

Kim, S. H., and B. L. Nelson. 2006. "Selecting the best system". In *Handbook in Operations Research and Management Science: Simulation*, edited by S. Henderson and B. Nelson, 501–534. Amsterdam: Elsevier.

Kim, S. H., and B. L. Nelson. 2007, December. "Recent advances in ranking and selection". 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, 162–172. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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

Powell, W., and I. Ryzhov. 2012. *Optimal Learning*. Wiley.

Powell, W. B. 2007. *Approximate Dynamic Programming: Solving the curses of dimensionality*. New York: John Wiley and Sons.

Powell, W. B., and P. I. Frazier. 2008. "Optimal Learning". *TutORials in Operations Research: State-of-the-Art Decision-Making Tools in the Information-Intensive Age*:213–246.

Raiffa, H., and R. Schlaifer. 1968. *Applied Statistical Decision Theory*. M.I.T. Press.

Rasmussen, C. E., and C. K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.

Ross, S. M. 1983. *Introduction to stochastic dynamic programming: Probability and mathematical*. Academic Press, Inc. Orlando, FL, USA.

Swisher, J. R., P. D. Hyden, S. H. Jacobson, L. W. Schruben, M. W. Hosp, and V. A. Fredericksburg. 2000, December. "A survey of simulation optimization techniques and procedures". In *Proceedings of the 2000 Winter Simulation Conference*, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, Volume 1, 119–128. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Swisher, J. R., S. H. Jacobson, and E. Yücesan. 2003. "Discrete-event simulation optimization using ranking, selection, and multiple comparison procedures: A survey". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 13 (2): 134–154.

Wetherill, G. B., and D. W. Brown. 1991. *Statistical process control: theory and practice*. Chapman & Hall.

Xie, J., and P. I. Frazier. 2011. "Sequential Bayes-Optimal Policies for Multiple Comparisons with a Known Standard". in review.

AUTHOR BIOGRAPHY

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.