UPPER BOUNDS ON THE BAYES-OPTIMAL PROCEDURE FOR RANKING & SELECTION WITH INDEPENDENT NORMAL PRIORS

Jing Xie
Peter I. Frazier

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

ABSTRACT

We consider the Bayesian formulation of the ranking and selection problem, with an independent normal prior, independent samples, and a cost per sample. While a number of procedures have been developed for this problem in the literature, the gap between the best existing procedure and the Bayes-optimal one remains unknown, because computation of the Bayes-optimal procedure using existing methods requires solving a stochastic dynamic program whose dimension increases with the number of alternatives. In this paper, we give a tractable method for computing an upper bound on the value of the Bayes-optimal procedure, which uses a decomposition technique to break a high-dimensional dynamic program into a number of low-dimensional ones, avoiding the curse of dimensionality. This allows calculation of the optimality gap for any given problem setting, giving information about how much additional benefit we may obtain through further algorithmic development. We apply this technique to several problem settings, finding some in which the gap is small, and others in which it is large.

1 INTRODUCTION

We consider the ranking and selection (R&S) problem, in which we wish to select the best among several competing alternatives, and the only way to evaluate the quality of an alternative is through stochastic simulation. Our goal in R&S is to allocate our simulation sampling effort efficiently among the alternatives, so as to accurately determine which alternative has the largest expected performance, while at the same time limiting simulation effort.

This problem has been considered by many authors, under four distinct mathematical formulations. We specifically consider the Bayesian formulation, for which early work dates to Raiffa and Schlaifer (1968), with recent surveys Chick (2006) and Frazier (2012). The other mathematical formulations of the problem are the indifference-zone formulation (see the monograph Bechhofer, Santner, and Goldsman 1995, and the survey Kim and Nelson 2006); the optimal computing budget allocation, or OCBA (Chen and Lee 2010); and the large-deviations approach (Glynn and Juneja 2004).

In the Bayesian formulation of the R&S problem, we place a prior distribution on the unknown true expected performance of each alternative, and our goal is to design an algorithm for allocating simulation effort with good average-case performance under the prior. While some work in this area, such as Raiffa and Schlaifer (1968), Chick and Inoue (2001b), and Chick and Inoue (2001a), considers two-stage algorithms, much of the recent work, such as Gupta and Miescke (1996), Chick, Branke, and Schmidt (2010), Frazier, Powell, and Dayanik (2008), Chick and Gans (2009), and Chick and Frazier (2012), has focused on sequential procedures, whose allocations of sampling effort are potentially more responsive to previous samples, and thus promise greater efficiency.

Bayes-optimal sequential R&S procedures are characterized by the dynamic programming equations, and given sufficient computational power, can be computed by solving these equations. These equations

have been used to compute Bayes-optimal procedures for problems with one alternative of unknown value and one of known value (Chick and Gans 2009, Chick and Frazier 2012), and for problems with two alternatives of unknown value (Frazier, Powell, and Dayanik 2008). However, for problems with more than a few alternatives, solving these dynamic programming equations becomes computationally infeasible, due to the curse of dimensionality (Powell 2007).

Thus, work in Bayesian R&S has focused in large part on developing sub-optimal procedures. These procedures are evaluated sometimes through theoretical investigations, but also by empirical comparison with previously developed procedures in simulation experiments. If a new procedure outperforms previously proposed procedures, then this is an improvement of the state-of-the-art. One can view the performance of each newly proposed procedure as a lower bound on the value of a Bayes-optimal procedure, and as more procedures are proposed, we may hope these lower bounds will get closer to this Bayes-optimal value.

In this paper, we focus on a complimentary approach: computing upper bounds on the value of a Bayes-optimal procedure. We focus on one version of the sequential Bayesian R&S problem, independent normal samples with known variance with an infinite horizon and a cost per sample, which was previously considered in Frazier and Powell (2008) and Chick and Frazier (2012). For this problem, we use a Lagrangian relaxation technique to obtain a computable upper bound on the value of a Bayes-optimal procedure. Our computational procedures build on recent work for the problem of sequential Bayesian multiple comparisons with a known standard, for which the Bayes-optimal procedure can be computed efficiently (Xie and Frazier 2013).

This allows computing an optimality gap, which is the distance between this upper bound and the expected performance of the best existing procedure (which may depend on the specific problem parameters used). This may be used to inform judgments of the value of continued algorithmic development. If this gap is small for a given set of problem parameters, it tells us that future procedures can improve only by a small margin over the current state-of-the-art. If this gap is large, this may be because existing procedures are far from optimal or because the upper bound is loose, or both. Being able to compute gaps as a function of problem's parameters will allow future researchers to focus development of improved procedures and upper bounds on regions of the problem parameter space where the gap is large.

The mathematical approach that we follow can be viewed as a Lagrangian relaxation of a stochastic dynamic program, which was used in Whittle (1980) to study restless bandit problems, and is also treated in Gittins, Glazebrook, and Weber (2011). Our focus on obtaining upper bounds for sequential decision-making problems is also similar in spirit to recent work on information relaxations in Brown, Smith, and Sun (2010), Brown and Smith (2011), Haugh and Kogan (2004), and Rogers (2002).

We begin in Section 2 by formulating the problem. We then describe our upper bound in Sections 3 and how to compute it in Section 4. In Section 5 we describe some special cases in which the bound is tight. In Section 6 we apply this bound to a variety of problems. In Section 7 we offer concluding remarks.

2 THE BAYESIAN RANKING & SELECTION PROBLEM

We would like to select the best among k alternative systems. We assume that samples from alternative x are normally distributed, with mean θ_x and variance λ_x, and independence across time and across alternatives. The means θ_x are unknown, while the sampling variances λ_x are assumed known. We let $\theta = (\theta_1, \ldots, \theta_k)$. We place a Bayesian prior distribution upon the unknown sampling means,

$$\theta_x \sim \mathcal{N}\left(\mu_{0,x}, \sigma_{0,x}^2\right), \qquad x = 1, \ldots, k,$$

with independence across alternatives. Our goal is to find the alternative with the largest mean θ_x, i.e., to find $x^* \in \operatorname{argmax}_x \theta_x$, and our main challenge in Bayesian R&S is to allocate simulation effort efficiently, so as to best support making this determination.

We assume no fixed computational budget, and instead assume that each sample of alternative x carries a cost $c_x > 0$ that may vary across alternatives. This assumption may be inappropriate when simulations are performed on hardware owned by the simulation analyst, and is instead intended to model the cost

structure of on-demand computation purchased through existing cloud computing services, in which the user pays a cost per hour of CPU time consumed.

We index time by $n = 1, 2, \ldots,$ and perform our simulations sequentially. At each time n, based on the samples observed so far, we either choose to stop sampling (see below), or we choose an alternative x_n to sample, paying a cost c_{x_n}, and observing a sampled value y_n, $y_n | x_n, \theta \sim \mathcal{N}(\theta_{x_n}, \lambda_{x_n})$. The posterior distribution that results from a sequence of observations obtained in this way is

$$\theta_x \mid x_1, y_1, \ldots, x_n, y_n \sim \mathcal{N}\left(\mu_{n,x}, \sigma_{n,x}^2\right), \qquad x = 1, \ldots, k, \quad n = 1, 2, \ldots,$$

where $\mu_{n,x}$ and $\sigma_{n,x}^2$ can be computed recursively from $\mu_{n-1,x}$, $\sigma_{n-1,x}^2$, x_n, and y_n (see, e.g., DeGroot 1970 or equation 2 in Frazier 2012). We define $x_n^* \in \text{argmax}_x \mu_{n,x}$.

At time n, if we choose to stop sampling, then we select an alternative as the best based on the previously collected samples, and receive a reward equal to the true value of that alternative. We call $\widehat{x_*}$ the selected alternative, so that the reward received from this selection is $\theta_{\widehat{x_*}}$. We call τ the total number of samples taken. We assume that $\widehat{x_*} = x_\tau^*$, and one can show formally that this choice is the best possible, as measured by expected reward under the prior (see, e.g., Frazier, Powell, and Dayanik 2008).

A procedure, or policy, for Bayesian R&S is then comprised of a sampling rule, for choosing each x_n based on the previous samples $(x_m, y_m : m < n)$, and a stopping rule, for choosing at each time n based on this same information whether to continue sampling or not, and thus implicitly for choosing the number of samples taken τ. (The selection rule is assumed to be $\widehat{x_*} = x_\tau^*$, as stated above.) We refer to such a policy with the notation π.

In Bayesian R&S, we measure the quality of a policy π by the expected net reward under the prior distribution, $\mathbb{E}^\pi[\theta_{\widehat{x_*}} - \sum_{n=1}^\tau c_{x_n}]$. where the expectation is taken both over randomness due to the stochasticity of the samples, and to the uncertainty about θ, and is written using the notation $\mathbb{E}^\pi$ when it depends upon the policy π. This reward includes both the reward due to selection, $\theta_{\widehat{x_*}}$, and the sampling costs, $\sum_{n=1}^\tau c_{x_n}$. We use the notation E_n^π to indicate the conditional expectation, with respect to the information available at time n, $(x_m, y_m : m \leq n)$.

This formulation of the sequential Bayesian R&S problem with independent normal samples, known sampling variance, independent normal prior, infinite horizon, and sampling costs, follows that of Frazier and Powell (2008), Chick and Frazier (2012), and is quite similar to the model in Chick and Gans (2009), which assumes a discount factor, and to the model in Frazier, Powell, and Dayanik (2008), which assumes a finite horizon and no discounting.

With this formulation, the expected value of a Bayes-optimal sampling policy is then

$$r := \sup_\pi \mathbb{E}^\pi \left[\theta_{x_\tau^*} - \sum_{n=1}^\tau c_{x_n} \right], \tag{1}$$

which depends implicitly on the number of alternatives k, and, the vectors composed of the prior mean $\mu_{0,x}$, prior variance $\sigma_{0,x}^2$, sampling variance λ_x, and sampling cost c_x of each alternative $x = 1, \ldots, k$.

This value r, understood as the solution to a stochastic dynamic programming problem, is characterized by the dynamic programming equations, e.g., as described in Chick and Frazier (2012), but actually computing r using existing methods is intractable except when k is very small. This intractability is caused (1) by the fact that the state space of the dynamic program is the set of all possible values of the vector of posterior means $(\mu_{nx} : x = 1, \ldots, k)$ and the vector of posterior variances $(\sigma_{nx}^2 : x = 1, \ldots, k)$, which has $2k$ dimensions; and (2) by the fact that computation required to solve a dynamic program scales badly with the dimensionality of its state space — a phenomenon which is referred to as the curse of dimensionality (see, e.g., Powell 2007). Our contribution in this paper is to provide a tractable method for computing an upper bound on r.

One simple upper bound is immediately apparent from (1). Sampling costs are positive, $c_x > 0$, so we have $\sum_{n=1}^{\tau} c_{x_n} \geq 0$. Also, $\theta_{x_\tau^*} \leq \max_x \theta_x$. Thus,

$$r \leq \mathbb{E} \left[\max_x \theta_x \right] := \mathrm{UB}^s$$

where the expectation does not depend on π, and so we use the notation $\mathbb{E}$ rather than $\mathbb{E}^\pi$. UB^s can be computed via numerical integration or Monte Carlo. This upper bound was used as a benchmark in Chick and Frazier (2012).

In the following sections, we will provide a tighter and more sophisticated upper bound than UB^s.

3 UPPER BOUND ON THE BAYES-OPTIMAL VALUE: STEP 1 (DECOMPOSITION)

In this section, we provide an upper bound on (1) in terms of a stochastic dynamic program with a special structure that admits solution through decomposition, avoiding the curse of dimensionality. Later, in Section 4, we show how to exploit this structure to allow efficient computation.

We first derive the dynamic program upper bound using a direct approach in Section 3.1, giving the bound below in equation (5). We then show an alternative derivation using a Lagrangian relaxation in Section 3.2, which is more complicated, but relates the upper bound to previous literature, and may also suggest generalizations.

3.1 Direct Approach

First, it is convenient to rewrite r as

$$r := \sup_\pi \mathbb{E}^\pi \left[\theta_{x_\tau^*} - \sum_{n=1}^{\tau} c_{x_n} \right] = \sup_\pi \mathbb{E}^\pi \left[\max_x \mu_{\tau,x} - \sum_{n=1}^{\tau} c_{x_n} \right], \tag{2}$$

where the second equation holds since $\mathbb{E}^\pi \theta_{x_\tau^*} = \mathbb{E}^\pi \left[\mathbb{E}_\tau^\pi \theta_{x_\tau^*} \right] = \mathbb{E}^\pi \left[\mu_{\tau,x_\tau^*} \right] = \mathbb{E}^\pi \left[\max_x \mu_{\tau,x} \right]$ by the tower property of conditional expectation.

We now state the following lemma, which bounds the reward received from the selection decision, and whose proof involves simple algebraic manipulations.

Lemma 1 For any $d \in \mathbb{R}$,

$$\max_x \mu_{\tau,x} \leq d + \sum_{x=1}^{k} (\mu_{\tau,x} - d)^+. \tag{3}$$

This inequality holds with equality if and only if $\mu_{\tau,x_\tau^{**}} \leq d \leq \mu_{\tau,x_\tau^*}$, where $x_\tau^{**} = \operatorname{argmax}_{x \neq x_\tau^*} \mu_{\tau,x}$.

Proof. The right-hand side of (3) can be rewritten as

$$d + \sum_{x:\ \mu_{\tau,x} \geq d} (\mu_{\tau,x} - d).$$

Consider two cases. If $\max_x \mu_{\tau,x} \geq d$, then this quantity is greater than or equal to $d + \max_x \mu_{\tau,x} - d$, which is equal to the left-hand side of (3). If not, so $\max_x \mu_{\tau,x} < d$, then the right-hand side of (3) is equal to d, which is greater than the left-hand side of (3) by our supposition. In both cases, the right-hand side of (3) is greater than or equal to the left-hand side.

Furthermore, $\sum_{x=1}^{k} (\mu_{\tau,x} - d)^+ = \max_x \mu_{\tau,x} - d$ if and only if $\mu_{\tau,x_\tau^{**}} \leq d \leq \mu_{\tau,x_\tau^*}$, hence the result follows. $\qquad\square$

It follows from (2) and Lemma 1 that, for any $d \in \mathbb{R}$,

$$r \leq d + \sup_\pi \mathbb{E}^\pi \left[\sum_{x=1}^{k} (\mu_{\tau,x} - d)^+ - \sum_{n=1}^{\tau} c_{x_n} \right] := R(d), \tag{4}$$

where we have defined the quantity $R(d)$.

We will see in Section 4 that $R(d)$ can be computed efficiently. Moreover, since the bound (4) holds for any $d \in \mathbb{R}$, it follows that

$$r \leq \inf_d R(d) = \inf_d \left\{ d + \sup_\pi \mathbb{E}^\pi \left[\sum_{x=1}^k (\mu_{\tau,x} - d)^+ - \sum_{n=1}^\tau c_{x_n} \right] \right\} := \text{UB}^*. \tag{5}$$

We will see in Section 4 that, in addition to being able to compute $R(d)$ efficiently for any d, we can also take the infimum efficiently over d to calculate UB^*. The bound UB^* is the main focus of this paper.

3.2 Lagrangian Approach

Suppose that we enlarge the set of decisions made by each policy π to include an additional variable $a_x \in [0,1]$ for each alternative x, whose value is determined at time τ when sampling stops. Let Π^0 be the set of such policies satisfying the constraint $\sum_{x=1}^k a_x = 1$ almost surely. That is, $\Pi^0 = \left\{ \pi : \sum_{x=1}^k a_x = 1 \right\}$. It follows immediately that

$$r = \sup_{\pi \in \Pi^0} \mathbb{E}^\pi \left[\sum_{x=1}^k a_x \mu_{\tau,x} - \sum_{n=1}^\tau c_{x_n} \right]. \tag{6}$$

We now apply a Lagrangian relaxation to (6). Let Π^1 be the set of policies that relaxes the constraint on a_x to hold only in expectation, and not almost surely, so $\Pi^1 = \left\{ \pi : \mathbb{E}^\pi \left[\sum_{x=1}^k a_x \right] = 1 \right\}$. It follows that $\Pi^0 \subseteq \Pi^1$. Thus, since, taking a supremum over a larger set provides an upper bound, we know that for any $d \in \mathbb{R}$,

$$r \leq \sup_{\pi \in \Pi^1} \mathbb{E}^\pi \left[\sum_{x=1}^k a_x \mu_{\tau,x} - \sum_{n=1}^\tau c_{x_n} \right] = \sup_{\pi \in \Pi^1} \left\{ \mathbb{E}^\pi \left[\sum_{x=1}^k a_x \mu_{\tau,x} - \sum_{n=1}^\tau c_{x_n} \right] - d \left[\mathbb{E}^\pi \left(\sum_{x=1}^k a_x \right) - 1 \right] \right\}$$

$$= d + \sup_{\pi \in \Pi^1} \mathbb{E}^\pi \left[\sum_{x=1}^k a_x (\mu_{\tau,x} - d) - \sum_{n=1}^\tau c_{x_n} \right] = d + \sup_{\pi \in \Pi^1} \mathbb{E}^\pi \left[\sum_{x=1}^k (\mu_{\tau,x} - d)^+ - \sum_{n=1}^\tau c_{x_n} \right]$$

$$= d + \sup_{\pi \in \Pi} \mathbb{E}^\pi \left[\sum_{x=1}^k (\mu_{\tau,x} - d)^+ - \sum_{n=1}^\tau c_{x_n} \right]$$

In the first equality, we have used $E^\pi [\sum_{x=1}^k a_x] = 1$ for all $\pi \in \Pi^1$. In the second equality, we have used the linearity of expectation and the fact that d does not depend on π to rearrange terms. In the third equality, we have used that the optimal choice of a_x in this supremum is to choose $a_x = 1$ when $\mu_{\tau,x} - d$ is positive, and $a_x = 0$ when it is negative. In the fourth and last equality, we have switched Π^1 to Π because the value whose supremum being taken does not depend on a_x, making it sufficient to consider $\pi \in \Pi$.

We have derived the same upper bound UB^* in (5), where d has played the role of a Lagrange multiplier on the constraint $E^\pi [\sum_{x=1}^k a_x] = 1$, using a technique similar to that used in Whittle (1980), Gittins, Glazebrook, and Weber (2011).

4 UPPER BOUND ON THE BAYES-OPTIMAL VALUE: STEP 2 (COMPUTATION)

In this section we give a tractable method for computing the upper bound UB^* on the expected value of the Bayes-optimal policy. This method has two components. First, we show that $R(d)$ can be computed efficiently for any given d, using a method developed in Xie and Frazier (2013). Second, we show that $d \mapsto R(d)$ is convex in d, allowing efficient computation of UB^* with a standard method for minimization of a one-dimensional convex function that uses only function values, such as Fibonacci search or golden section search (Kiefer 1953).

4.1 Computation of $R(d)$

As written in (4), computation of $R(d)$ requires solving a dynamic program whose state space includes every possible value of the $2k$-dimensional vector $(\mu_{nx}, \sigma_{nx}^2 : x = 1, \ldots, k)$, for which memory requirements and computation time scale exponentially in k, making computation intractable except when k is very small.

The essential idea behind this technique is to use the fact that alternatives are independent of each other, and costs are additive, to rewrite $R(d)$ as the sum of the values of k different sub-problems, each of which is much easier to solve than the original problem,

$$R(d) = \sum_{x=1}^{k} R_x(d), \tag{7}$$

where $R_x(d)$ is

$$R_x(d) = d/k + \sup_{\pi \in \Pi^x} \mathbb{E}^{\pi} \left[(\mu_{\tau,x} - d)^+ - \sum_{n=1}^{\tau} c_x \right],$$

and Π^x is the set of policies with $x_n = x$ for all x, i.e., that only measure alternative x. Calculating $R_x(d)$ requires solving a dynamic program whose state space is only two-dimensional, as it contains only $\mu_{n,x}$ and $\sigma_{n,x}^2$ for a single x. Solving such a low-dimensional dynamic program *is* tractable, and the computation to solve k 2-dimensional dynamic programs scales only linearly in k.

Figure 1 shows R_x as a function of d, with k taking values $1, 2, 3$ and 100, and the other parameters fixed to $\mu_{0x} = 0$, $\sigma_{0,x}^2 = 1$, $\lambda_x = 10$, $c_x = e^{-3}$. The figure suggests that R_x is convex in d, foreshadowing the result on convexity of R (though not R_x) to come in Section 4.2.
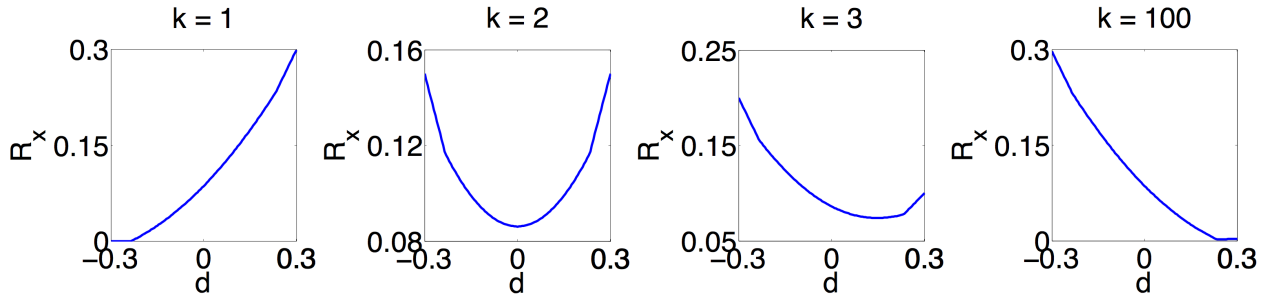

Figure 1: $R_x(d)$ as a function of d, when $\mu_{0x} = 0$, $\sigma_{0,x}^2 = 1$, $\lambda_x = 10$, $c_x = e^{-3}$, for different values of k. From left to right, $k = 1, 2, 3, 100$. $R(d) = \sum_{x=1}^{k} R_x(d)$, and our upper bound on the value of a Bayes-optimal procedure is $\text{UB}^* = \inf_d R(d)$.

The decomposition (7) was previously reported, and justified formally, in Xie and Frazier (2013), which considered the related problem of multiple-comparisons with a known standard (MCS). The quantity $R(d) - d$ is actually the value of a Bayes-optimal procedure for a variant of this MCS problem, which Xie and Frazier (2013) refers to as the variant for normal sampling, linear terminal payoff, and infinite horizon.

Here, we briefly describe this variant of the MCS problem. It arises when we sample exactly as in Section 2, paying a cost for each sample as before, but when sampling stops, our goal is not to find the alternative with the best true mean θ_x, but is instead to find the set of alternatives whose true means are above a threshold d, $\{x : \theta_x \geq d\}$. At time τ, in this problem, the decision-maker chooses a set of alternatives, and earns a reward of $\theta_x - d$ for every alternative selected, and a reward of 0 for every alternative not selected. The Bayes-optimal way to make this selection decision is to choose to include an alternative x iff $0 \leq E_\tau[\theta_x - d] = \mu_{\tau,x} - d$, making the optimal set of alternatives to include $B_\tau := \{x : \mu_{\tau,x} - d \geq 0\}$. When the selection decision is made in this way, the resulting expected reward is $\mathbb{E}^{\pi} \left[\sum_{x \in B_\tau} (\theta_x - d) \right] = \mathbb{E}^{\pi} \left[\mathbb{E}_\tau \left[\sum_{x \in B_\tau} (\theta_x - d) \right] \right] = \mathbb{E}^{\pi} \left[\sum_{x \in B_\tau} (\mu_{\tau,x} - d) \right] = \mathbb{E}^{\pi} \left[\sum_{x=1}^{k} (\mu_{\tau,x} - d)^+ \right]$, by the

tower property of conditional expectation. Including the sampling costs, the value of a Bayes-optimal policy for this variant of the MCS problem is $\sup_{\pi \in \Pi} \mathbb{E}^{\pi} \left[\sum_{x=1}^{k} (\mu_{\tau,x} - d)^{+} - \sum_{n=1}^{\tau} c_{x_n} \right]$, which is exactly $R(d) - d$.

It is also useful for development in Section 5 to define $V_x(d) = R_x(d) - d/k - (\mu_{0,x} - d)^{+}$. We have subtracted from $R_x(d)$ the term d/k, as well as the value $(\mu_{0,x} - d)^{+}$ that we would receive in expectation in the MCS problem if we were forced to stop immediately and estimate whether θ_x is above d or below d. Thus, $V_x(d)$ can be seen as the optimal incremental reward that can be obtained through sampling, in an MCS problem with a single alternative. Xie and Frazier (2013), in its discussion of MCS with linear terminal payoff, normal sampling and infinite horizon, shows that V_x is non-negative, symmetric, maximized at $d = \mu_{0,x}$, with bounded support. The methods described in Xie and Frazier (2013) actually compute $V_x(d)$, from which $R_x(d)$ can be determined via

$$R_x(d) = d/k + (\mu_{0,x} - d)^{+} + V_x(d).$$

4.2 Convexity of $R(d)$

We now show that R is convex, allowing efficient computation of $\mathrm{UB}^{*} = \inf_d R(d)$ with Fibonacci search or golden section search.

Proposition 1 *R is a convex function.*

Proof.   Let Π denote the complete set of policies. Since point-wise supremum preserves convexity, it suffices to show that $g \colon \mathbb{R} \times \Pi \mapsto \mathbb{R}$, defined by

$$g(d, \pi) = \mathbb{E}^{\pi} \left[\sum_{x=1}^{k} (\mu_{\tau,x} - d)^{+} - \sum_{n=1}^{\tau} c_{x_n} \right],$$

is convex in d for any given π. Now for a given π,

$$g(d, \pi) = \int h(d, \vec{\omega}) \, p^{\pi}(\vec{\omega}) \, \mathrm{d}\vec{\omega},$$

where

$$\vec{\omega} = (\tau, x_1, \dots, x_{\tau}, \mu_{\tau,1}, \dots, \mu_{\tau,k}), \qquad h(d, \vec{\omega}) = \sum_{x=1}^{k} (\mu_{\tau,x} - d)^{+} - \sum_{n=1}^{\tau} c_{x_n},$$

and $p^{\pi}(\vec{\omega})$ is the probability distribution of $\vec{\omega}$ given the specified priors and the sampling policy π. Since h is convex in d for any given $\vec{\omega}$, its integral (infinite sum) g is also convex in d. $\quad\square$

5   SPECIAL CASES IN WHICH THE UPPER BOUND IS TIGHT

In general, the upper bound UB^{*} is not tight. However, the following theorems present two special cases in which the upper bound UB^{*} is tight, i.e., in which it is strictly equal to the optimal expected value r.

Theorem 1 *If $k = 1$, then $\mathrm{UB}^{s} = \mathrm{UB}^{*} = r$.*

Proof.   First note that $\mathrm{UB}^{s} = \mathbb{E}[\theta_x] = \mu_{0,x}$, and that the optimal policy is to stop without taking any samples at $\tau = 0$, with $r = \mathbb{E}[\theta_x] = \mu_{0,x}$.

$$R(d) = d + (\mu_{0,x} - d)^{+} + V_x(d) = \begin{cases} d + V_x(d), & \text{if } d \geq \mu_{0,x} \\ \mu_{0,x} + V_x(d), & \text{otherwise} \end{cases}.$$

Since V_x is symmetric and maximized at $d = \mu_{0,x}$, we know

$$\mathrm{UB}^{*} = \inf_d R(d) = R(-\infty) = \mu_{0,x} + V_x(-\infty) = \mu_{0,x}.$$
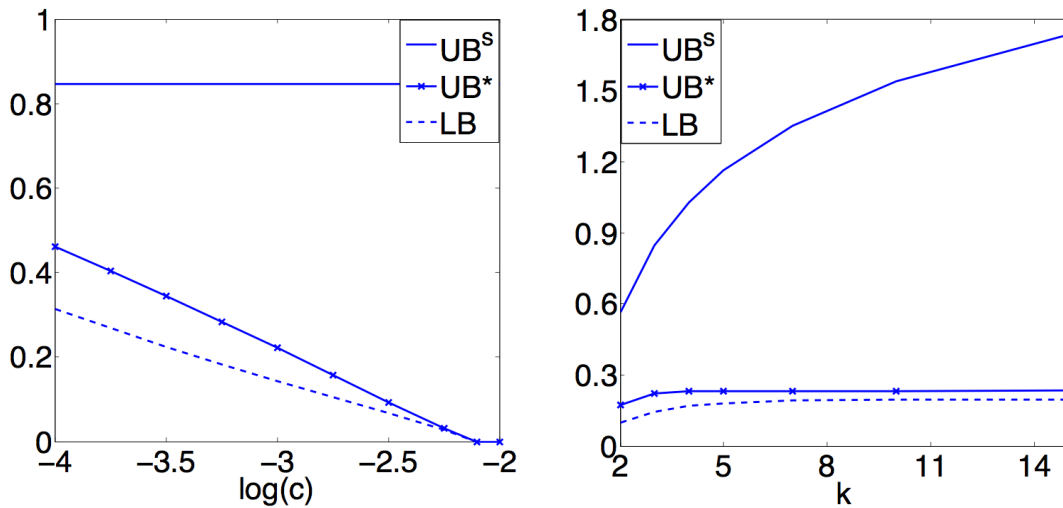
$\quad\square$

Figure 2: Upper bounds UB^* and UB^s and lower bound LB on the value of a Bayes-optimal procedure for R&S problems with $\mu_{0x} = 0$, $\sigma_{0,x}^2 = 1$, $\lambda_x = 10$, and $c_x = c$ for all x. UB^s is a simple upper bound computed by supposing that the best alternative is revealed without sampling, and UB^* is computed using the methods described in this paper. LB is the expected value of the best existing procedure for the given problem parameters, among a collection of procedures tested, as computed using Monte Carlo simulation. The left plot fixes $k = 3$ and varies c from e^{-4} to e^{-2}. The right plot fixes $c = e^{-3}$ and varies k.

Theorem 2 If $k = 2$ and $\sigma_{0,1}^2 = 0$, then $UB^* = r$.

Proof. Since alternative 1 has known value $\mu_{0,1}$, the optimal sampling policy only samples from alternative 2, and $\mu_{n,1} = \mu_{0,1}$ for all n. It follows from (2) that

$$r = \sup_\pi \mathbb{E}^\pi \left[\max\{\mu_{0,1}, \mu_{\tau,2}\} - \sum_{n=1}^\tau c_{x_n} \right] = \mu_{0,1} + \sup_\pi \mathbb{E}^\pi \left[(\mu_{\tau,2} - \mu_{0,1})^+ - \sum_{n=1}^\tau c_{x_n} \right].$$

By (5), we know

$$r \leq UB^* \leq \mu_{0,1} + \sup_\pi \mathbb{E}^\pi \left[(\mu_{\tau,2} - \mu_{0,1})^+ - \sum_{n=1}^\tau c_{x_n} \right] = r.$$

$\square$

6 NUMERICAL RESULTS

In this section we apply the technique in Section 4 for computing the proposed upper bound UB^* on several test problems, to bound the optimality gaps of existing R&S procedures.

In our numerical experiments, we implement a number of benchmarking policies, each of which is the combination of a sampling rule among KG_1 (Frazier and Powell 2008), KG_* (Frazier and Powell 2010), ESP_b (Chick and Frazier 2012), and a stopping rule among $EOC_{c,k}$ (Chick and Frazier 2012), KG_* (Frazier and Powell 2010), ESP_b (Chick and Frazier 2012). The best expected value of these policies serve as a lower bound on the expected value of the Bayes-optimal policy. We denote this lower bound by LB.

First, in Figure 2, we consider a collection of problems with homogeneous priors $\mu_{0x} = 0$, $\sigma_{0,x}^2 = 1$, on the unknown means, homogeneous sampling variances $\lambda_x = 10$, and homogeneous sampling costs, $c_x = c$ for all x. We first fix $k = 3$ and vary $\log(c)$ (where log indicates the natural logarithm) within $[-4, -2]$, and then fix $\log(c) = -3$ and vary k within $[2, 15]$. Figure 2 shows the resulting upper bounds UB^s, UB^*, and the lower bound LB, on the Bayes-optimal value.

Figure 2 shows that the proposed upper bound UB^* improves dramatically over the naive upper bound UB^s. Moreover, the optimality gap provided by UB^* vanishes as c increases, and stabilizes as k increases. We hypothesize that the optimality gap vanishes as c increases because, when c is large, both the Bayes-optimal R&S procedure, and the Bayes-optimal MCS procedure used to compute UB^*, stop sampling immediately, without taking any samples. When both procedures stop immediately, then $\mu_{\tau,x} = \mu_{0,x}$, and the bound is tight.
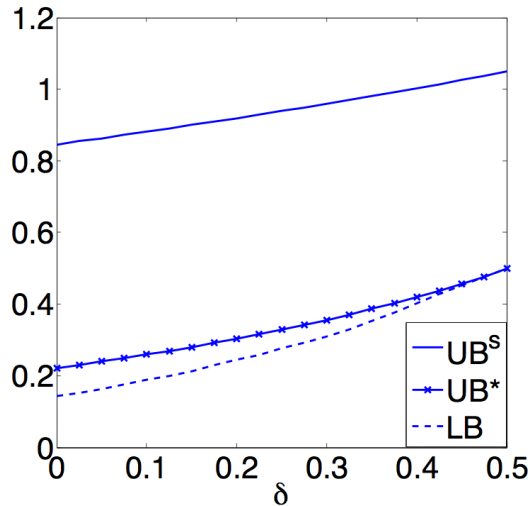

Figure 3: Upper bounds UB^* and UB^s and lower bound LB on the value of the Bayes-optimal procedure for R&S problems with heterogeneous priors. We set $\sigma_{0,x}^2 = 1$, $\lambda_x = 10$, $c_x = e^{-3}$ for all x, $k = 3$, $\mu_{0,1} = \mu_{0,2} = 0$, and $\mu_{0,3} = \delta$, and plot bounds as a function of δ.

Second, in Figure 3, we consider problems with non-homogeneous priors on the unknown means. We set $\sigma_{0,x}^2 = 1$, $\lambda_x = 10$, $c_x = e^{-3}$ for all x, fix $k = 3$, $\mu_{0,1} = \mu_{0,2} = 0$, and vary $\mu_{0,3} = \delta$ between $[0, 0.5]$. UB^* gives a significantly tighter bound than does UB^s, and the gap vanishes for sufficiently large δ. We hypothesize that the optimality gap vanishes for large δ because a large difference between the best and second-best *prior* allows a well-chosen value of d to be between the best and second-best values of $\mu_{\tau,x}$ with a probability close to 1, and when this occurs the bound in Lemma 1 is tight.

7   CONCLUSIONS

We have provided a computationally tractable method for computing upper bounds on the value of a Bayes-optimal procedure for the Bayesian R&S problem with independent normal samples, an independent normal prior, and an infinite horizon with a cost per sample. These upper bounds can be used to judge how far from optimality existing procedures are, for a given set of problem parameters, and can be used to judge where future algorithmic development can be directed.

ACKNOWLEDGMENTS

REFERENCES

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

Brown, D. B., and J. E. Smith. 2011. "Dynamic portfolio optimization with transaction costs: Heuristics and dual bounds". *Management Science* 57 (10): 1752–1770.

Brown, D. B., J. E. Smith, and P. Sun. 2010. "Information relaxations and duality in stochastic dynamic programs". *Operations research* 58 (4-Part-1): 785–801.

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

Chick, S. 2006. "Bayesian ideas and discrete event simulation: why, what and how". In *Proc. Winter simulation Conference*, edited by L. Perrone, F. Wieland, J. Liu, B. Lawson, D. Nicol, and R. Fujimoto, 96–105. Piscataway, NJ: IEEE.

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

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

Chick, S., and N. Gans. 2009. "Economic Analysis of Simulation Selection Problems". *Management Sci.* 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. 2012. "Tutorial: Optimization via Simulation with Bayesian Statistics and Dynamic Programming". In *Proceedings of the 2012 Winter Simulation Conference*. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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

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

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.

Gittins, J., K. Glazebrook, and R. Weber. 2011. *Multi-armed Bandit Allocation Indices*. 2nd ed. Wiley.

Glynn, P., and S. Juneja. 2004. "A large deviations perspective on ordinal optimization". In *Proc. Winter Simulation Conference*, edited by R. Ingalls, M. Rossetti, J. Smith, and B. Peters, 577–585. Piscataway, NJ: IEEE.

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.

Haugh, M. B., and L. Kogan. 2004. "Pricing American options: a duality approach". *Operations Research* 52 (2): 258–270.

Kiefer, J. 1953. "Sequential minimax search for a maximum". *Proceedings of the American Mathematical Society* 4 (3): 502–506.

Kim, S., and B. Nelson. 2006. *Handbook in Operations Research and Management Science: Simulation*, Chapter Selecting the best system, 501–534. Amsterdam: Elsevier.

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

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

Rogers, L. C. 2002. "Monte Carlo valuation of American options". *Mathematical Finance* 12 (3): 271–286.

Whittle, P. 1980. "Multi-armed bandits and the Gittins index". *Journal of the Royal Statistical Society. Series B (Methodological)* 42 (2): 143–149.

Xie, J., and P. Frazier. 2013. "Sequential Bayes-Optimal Policies for Multiple Comparisons with a Known Standard". *Operations Research*. to appear.

AUTHOR BIOGRAPHIES

JING XIE is a PhD student in the School of Operations Research and Information Engineering at Cornell University. She received a B.S. in mathematics from Fudan University at Shanghai, China in 2009. She has a research interest in simulation optimization and Bayesian statistics. Her e-mail is <jx66@cornell.edu> and her web page is <http://people.orie.cornell.edu/jx66/>.

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 recepient of an AFOSR Young Investigator Award, and an NSF CAREER Award. His research interest is in dynamic programming and Bayesian statistics, focusing on the optimal acquisition of information. He works on applications in simulation, optimization, operations management, and medicine. His email address is <pf98@cornell.edu> and his web page can be found via <www.orie.cornell.edu>.