Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

UTILITY-BASED STATISTICAL SELECTION PROCEDURES

Guowei Sun

Department of Mathematics
Institute for Systems Research
University of Maryland
College Park, MD 20742, USA

Yunchuan Li

Department of Electrical and Computer Engineering
Institute for Systems Research
University of Maryland
College Park, MD 20742, USA

Michael C. Fu

Robert H. Smith Business School
Institute for Systems Research
University of Maryland
College Park, MD 20742, USA

ABSTRACT

We present two sequential allocation frameworks for selecting from a set of competing alternatives when the decision maker cares about more than just the simple expected rewards. The frameworks are built on general parametric reward distributions and assume the objective of selection, which we refer to as utility, can be expressed as a function of the governing reward distributional parameters. The first algorithm, which we call utility-based OCBA (UOCBA), uses the Δ-technique to find the asymptotic distribution of a utility estimator to establish the asymptotically optimal allocation by solving the corresponding constrained optimization problem. The second, which we refer to as utility-based value of information (UVoI) approach, is a variation of the Bayesian value of information (VoI) techniques for efficient learning of the utility. We establish the asymptotic optimality of both allocation policies and illustrate the performance of the two algorithms through numerical experiments.

1 INTRODUCTION

Ranking and Selection (R&S) refers to the procedure of selecting the best system from a usually finite set of competing alternatives, where performance measures are expensive to collect and subject to noise. The vast majority of R&S literature works with the expectation of the random output. For example, consider the problem of selecting an *optimal* route for a delivery service (Kawabe et al. 2015). Modeling the traveling times on a designed route as a random variable to capture the unpredictable effects from factors such as weather and traffic conditions, the problem of testing a set of candidate routes to quickly identify the one with the smallest mean delivery time could then be treated as a R&S problem. Efficient testing strategies include optimal computing budget allocation (OCBA) and Bayesian VoI-based allocation algorithms. However, the mean delivery time may not be the appropriate measurement of route quality in this scenario. A route with a slightly higher mean but much less variance in delivery time, could be preferable compared to one with a smaller mean but much larger variance, as an unusually long delivery could cause packages to be delayed to the next day. Similar problems are also found in many other applications. In financial applications, value at (VaR) and conditional value at risk (CVaR) are two popular objectives when comparing different pricing strategies, as financial institutions are extremely sensitive to risks (Rockafellar and Uryasev 2000).

In behavioral economics, a cumulative prospect theoretic (CPT) utility is often used to properly capture people's perception of random rewards in games such as lotteries and gambling (Tversky and Kahneman 1992). In such scenarios, a utility function could be used to capture the problem-specific preferences of decision makers. In a similar route selection scenario, a (CPT) utility was applied in the works of Jie et al. (2018) in the multi-armed bandit setting for avoiding extraordinarily long traveling times. Prior research on ranking and selection algorithms designed for objectives other than expected values includes Trailovic and Pao (2004) for minimizing variance and Pasupathy et al. (2010) for quantile. In this paper, we consider more general objective functions.

There is a rich literature on solving R&S problems with simple expectation being the utility. The OCBA framework maximizes probability of correct selection (PCS) under a budget constraint to find the asymptotically optimal policy, proposing sequential allocation algorithms using plug-in estimates for the unknown parameters in the optimal allocation (Chen et al. 2000). The indifference-zone (IZ) approach provides a frequentist confidence guarantee for PCS under the assumption that there exists a gap of δ in expected performance between sub-optimal alternatives and the optimal one (Kim and Nelson 2001). Another line of research, which is often referred to as the Bayesian value of information (VoI) approach, works under a Bayesian framework for efficiently learning the expected value of the unknown random reward. At each step, the alternative that contains the most *information* (variously defined) is selected. Two notable examples are the expected improvement (EI) and knowledge gradient (KG) policies, which are shown to be more efficient in the finite-budget domain compared to asymptotically optimal policies such as OCBA (Chick 2006; Chick et al. 2010; Powell and Ryzhov 2012). Recent results in Ryzhov (2016) and Peng and Fu (2017) connected the EI policy asymptotically to the OCBA policy, providing theoretical support to its empirical performance. Another advantage of the Bayesian framework is its flexibility in incorporating problem-specific information such as correlations into the allocation procedures, such as in Frazier et al. (2009) and Qu et al. (2015), where similarity is modeled as correlation in the prior beliefs to facilitate better selection.

In this paper, we tackle the R&S problem where the quality of each alternative is measured by a utility function. Without assuming specific forms of the function, we first establish the asymptotically optimal allocation using techniques similar to OCBA and propose sequential selection algorithms based on the results. We then develop a Bayesian VoI approach for efficient learning of the utility rather than the simple expectation and establish the equivalence between the two approaches. We also discuss the issue of numerical computations and point out scenarios where the Bayesian approach could fail. Two numerical experiments using utility functions found in economics and operations research were performed validating the proposed algorithms.

The paper is organized as follows: Section 2 formulates the R&S problem. Section 3 finds the asymptotically optimal allocation configuration for maximizing PCS with a given utility function. Section 4 designs a Bayesian VoI dynamic allocation procedure by providing an information measure for selecting the next alternative. Section 5 formally states two algorithms based on the theoretical derivations, and Section 6 illustrates the performance of the algorithms on two simulation problems.

2 PROBLEM FORMULATION

Let $\mathscr{A} = \{1, 2...k\}$ denote the set of candidate alternatives, each associated with an unknown random outcome $Y_i, i \leq k$, where Y_i follows a distribution with unknown parameter θ_i. Before formulating the problem, we list a set of notations used throughout the paper.

- $U(\cdot)$: the known utility function,
- U_i : the true utility of alternative i computed as $U_i \equiv U(\theta_i)$,
- y_{ij} : the jth sample obtained for alternative i,
- $\hat{\theta}_i$: an estimator of θ_i,
- $\hat{U}_i$: an estimator of U_i,

- N : the total budget,
- i_n : alternative selected at allocation step n,
- $y^{(n)}$: the sample obtained at allocation step n from the chosen alternative i_n,
- n_i: the budget allocated to alternative i.

We assume U is known and the goal is to maximize the utility. For instance, in the case of quantile selection in Pasupathy et al. (2010) with normal random rewards where $\theta_i = \{\mu_i, \sigma_i\}$, the utilities are $U_i = \mu_i + \alpha\sigma_i$, where α is the quantile coefficient of a standard normal density. Without loss of generality, we assume that $U_1 \geq U_2 \geq ... \geq U_k$, so that alternative 1 is the best. Upon exhausting the budget N, we make the final selection as

$$i^N = \arg\max_{i \leq k}\{\hat{U}_i\},$$

where $\hat{U}_i$ is chosen as $U(\hat{\theta}_i)$ in the UOCBA approach and $\mathbb{E}[U(\hat{\theta}_i)]$ in the Bayesian VoI approach. Then, the problem of designing an allocation that maximizes PCS under a budget constraint N could be formulated as

$$\begin{aligned} &\max_{n_1,n_2..n_k} P\{\bigcap_{2\leq i\leq k} \hat{U}_1 \geq \hat{U}_i\} \\ &\text{s.t. } \sum_{i=1}^{k} n_i = N. \end{aligned} \tag{1}$$

In Section 3, we solve (1) in the asymptotic domain with $n_i \to \infty$ when $\hat{\theta}$ is the maximum likelihood estimator (MLE). In Section 4, we work in the Bayesian framework where posterior densities are assumed on θ_i and updated upon receiving new samples, and design information criterion for dynamically allocating the simulation budget.

3 UTILITY-BASED OPTIMAL COMPUTATION BUDGET ALLOCATION

In this section, we consider the frequentist problem setting and explicitly find the asymptotically optimal allocation configuration by solving Equation 1 when $n_i \to \infty, \forall i \leq k$. Under the assumption that $\hat{\theta}_i$ is the MLE of θ, the asymptotic distribution of $\hat{U}$ can be shown with the Δ-technique to be normal. Then we approximate PCS with its Bonferroni lower bound and derive the optimal allocation using standard techniques from OCBA.

3.1 Asymptotic Distribution of Plug-in Utility Estimator

Most R&S literature works with normal random rewards. In cases without normality, a simple batching procedure could be applied to obtain approximately normal samples (Law and Kelton 2007). We do not assume the normality of Y_i, but restrict them to the family of random distributions with the following property.

Assumption 1 $\{Y_i\}_{i=1}^k$ belong to the family of random variables such that $\hat{\theta}_i$ has the asymptotic distribution

$$\sqrt{n_i}(\hat{\theta}_i - \theta_i) \xrightarrow{\mathscr{D}} \mathscr{N}(0, I^{-1}(\theta_i)) \text{ as } n_i \to \infty,$$

where $I(\theta_i) = \mathbb{E}_{\theta_i}[(\nabla_\theta \log f(X|\theta)|_{\theta=\theta_i})(\nabla_\theta \log f(X|\theta)|_{\theta=\theta_i})^T]$ is the corresponding Fisher information matrix.

In the case of normal random outcomes, we have $(\hat{\mu}_i, \hat{\sigma}_i) = \left(\sum_{j=1}^{n_i} y_{ij}/n_i, \sum_{j=1}^{n_i}(y_{ij} - \hat{\mu}_i)^2/n_i\right)$ and the asymptotic distribution of $\hat{\mu}_i$ and $\hat{\sigma}_i$ is well-established. For densities other than normal, we refer the readers to Casella and Berger (2002) for conditions on which Assumption 1 will hold. Setting $\hat{U}_i$ to be the plug-in estimator $U(\hat{\theta}_i)$, we establish the asymptotic normality of $\hat{U}_i$ with the following lemma.

Lemma 1 If $U(\cdot)$ is a differentiable function, then

$$\sqrt{n_i}(\hat{U}_i - U_i) \xrightarrow{\mathscr{D}} \mathscr{N}(0, v_i^2) \text{ as } n_i \to \infty,$$

where

$$v_i^2 = \nabla^T U(\theta_i) I^{-1}(\theta_i) \nabla U(\theta_i). \tag{2}$$

Proof. A direct application of the Δ-technique (Casella and Berger 2002) will prove the Lemma. □

Lemma 1 establishes the normality of $U(\hat{\theta}_i)$, which allows us to find an approximation of PCS when $n_i \to \infty$ and explicitly solve (1).

3.2 Asymptotically Optimal Allocation Policy

With $n_i \to \infty$, we first construct an approximation of PCS in (1) using the asymptotic normality results in Lemma 1. Using the well-known Bonferroni lower bound, we have

$$\begin{aligned} PCS = P\left\{\bigcap_{2\le i\le k} \hat{U}_1 \ge \hat{U}_i\right\} &= 1 - P\left\{\bigcup_{2\le i\le k}\left(\bigcap_{j\ne i} \hat{U}_i > \hat{U}_j\right)\right\} = 1 - \sum_{2\le i\le k} P\left\{\bigcap_{j\ne i} \hat{U}_i \ge \hat{U}_j\right\} \\ &\ge 1 - \sum_{2\le i\le k} P\left\{\hat{U}_i \ge \hat{U}_1\right\} = APCS. \end{aligned}$$

We assume $\{\hat{U}_i\}_{i=1}^k$ are independent. With the normality results in Lemma 1, the term $P\{\hat{U}_i \ge \hat{U}_1\}$ can be expressed in terms of the standard normal cumulative distribution function. Letting

$$\delta_i = U_i - U_1,\ \zeta_i = \frac{\delta_i}{\sqrt{v_i^2/n_i + v_1^2/n_1}}, \tag{3}$$

APCS can be written compactly as

$$APCS = 1 - \sum_{2\le i\le k} \Phi(\zeta_i), \tag{4}$$

where $\Phi(\cdot)$ is the cumulative distribution function for a standard normal distribution. Reaching (4) required two approximations: (1) approximating PCS with its Bonferroni lower bound, and (2) approximating $\{\hat{U}_i\}_{i=1}^k$ with their asymptotic normal densities. The quality of the approximations will depend on the exact value of n_i, the true θ values and the utility function U. Careful evaluation of the quality of approximations and their effect on the allocation performance is an active area in OCBA-related research, but out of scope for this paper. The optimization problem in (1) can now be approximated by

$$\min \sum_{2\le i\le k} \Phi(\zeta_i) \text{ subject to } \sum_{i=1}^k n_i = N, \tag{5}$$

by replacing PCS with APCS. The approximate problem has the analytical solution presented in the following theorem.

Theorem 1 (Optimal Allocation for a Given Utility) Under the conditions of Lemma 1, the APCS is maximized as $N \to \infty$ when the allocations satisfy the conditions

$$n_1 = \sqrt{v_1 \left\{ \sum_{i=2}^{k} \frac{n_i^2}{v_i^2} \right\}}, \tag{6}$$

$$\frac{n_i}{n_j} = \left(\frac{\delta_i}{\delta_j}\right)^2 \cdot \frac{v_j^2}{v_i^2},\ 2 \le i, j \le k, \tag{7}$$

$$\sum_{i=1}^{k} n_i = N, \tag{8}$$

where v_i and δ_i are defined in (2) and (3), respectively.

Sketch of proof. Introducing Lagrange multiplier λ, the Lagrangian of the optimization problem in (5) can be written as

$$L(n_1, n_2, .., n_k, \lambda) = \sum_{2 \le i \le k} \Phi(\zeta_i) - \lambda\Big(\sum_{1 \le i \le k} n_i - N\Big).$$

Next, write the Karush-Kuhn-Tucker (KKT) conditions

$$\frac{\partial L}{\partial n_1} = -\sum_{2 \le i \le k} \frac{\partial \Phi(\zeta_i)}{\partial \zeta_i} \frac{\partial \zeta_i}{\partial n_1} - \lambda = 0,$$
$$\frac{\partial L}{\partial n_i} = -\frac{\partial \Phi(\zeta_i)}{\partial \zeta_i} \frac{\partial \zeta_i}{\partial n_i} - \lambda = 0, \forall 2 \le i \le k,$$

and apply the standard derivations in Chen et al. (2000) will yield the desired results. □

It is worth mentioning in the case of Y_i having normal densities with expected value being the utility, Theorem 1 reduces to the usual OCBA optimal allocation results.

4 UTILITY-BASED BAYESIAN VoI

We also attempt to tackle the problem by developing a variation of the Bayesian VoI technique for efficient learning of the unknown utility. In the frequentist setting, the utility is viewed as a function of fixed unknown distributional parameters, whereas in the Bayesian setting, we treat the utilities as random variables and design an information criterion for dynamically allocating the simulation budget. We first present two Bayesian models which we later use in our numerical experiments, then propose the expected utility improvement and establish its asymptotic equivalence with the UOCBA allocation results in Theorem 1.

4.1 Bayesian Models and VoI

Two common Bayesian models are the Normal-Normal and Beta-Bernoulli models where tractable posterior updates are readily available.

Normal-Normal posterior updates: In the Bayesian model with known normal priors on the mean parameters $\mu_i \sim \mathcal{N}(t_i^0, (\tau_i^0)^2)$ and normally distributed samples $Y_i \sim \mathcal{N}(\mu_i, \sigma_i)$ where σ_i is known, the posterior of μ_i at allocation step n is $\mathcal{N}(t_i^n, (\tau_i^n)^2)$ with the updates on the parameters

$$t_i^{n+1} = \begin{cases} \frac{(\tau_i^n)^{-2} t_i^n + \sigma_i^{-2} y^{n+1}}{(\tau_i^n)^{-2} + \sigma_i^{-2}}, & i_n = i \\ t_i^n, & i_n \ne i, \end{cases} \qquad \left(\tau_i^{n+1}\right)^2 = \begin{cases} \left((\tau_i^n)^{-2} + \sigma_i^{-2}\right)^{-1}, & i_n = i \\ (\tau_i^n)^2, & i_n \ne i. \end{cases} \tag{9}$$

Beta-Bernoulli Posterior Updates: Assuming the probabilities of success p_i have $Beta(\alpha_i^0, \beta_i^0)$ priors, upon receiving y^n which is either 0 or 1, the posterior of p_i is $Beta(\alpha_i^n, \beta_i^n)$ with the parameter updates

$$\alpha_i^{n+1} = \begin{cases} \alpha_i^n + \mathbb{1}_{\{y^n=1\}}, & i_n = i, \\ \alpha_i^n, & i_n \neq i, \end{cases} \quad \beta_i^{n+1} = \begin{cases} \beta_i^n + \mathbb{1}_{\{y^n=0\}}, & i_n = i, \\ \beta_i^n, & i_n \neq i, \end{cases} \tag{10}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function.
VoI tends to favor alternatives with higher uncertainty or higher estimated mean. Two important examples are Expected Improvement (EI) and Knowledge Gradient (KG). In this paper, we use a variant of EI for its simplicity in computation and its connection with the UOCBA policy.

4.2 Expected Utility Improvement

Denote the priors on $\theta_i,\ i \leq k$ by f_i^0 and the posteriors at step n as f_i^n. VoI seeks to measure the potential gain of learning θ_i by balancing the exploration-exploitation trade-off (Powell and Ryzhov 2012). We propose the expected utility improvement (EUI)

$$g_i^{EUI,n} = \mathbb{E}\left[(U(\theta_i) - U^*)^+\right], \tag{11}$$

where the expectation is taken with respect to (w.r.t.) $\theta_i \sim f_i^n$ and $U^* = \max_{i \leq k}\{\mathbb{E}[U(\theta_i)]\}$ is the current expected optimal utility under the posteriors $\{f_i^n\}_{i=1}^k$. The EUI-based policies select the alternative with maximum EUI at allocation step n as

$$i_n = \arg\max_{1 \leq i \leq k}\{g_i^{EUI,n}\}.$$

In the special case of Normal-Normal Bayesian models, similar to the results in Ryzhov (2016), we have the following theorem relating the asymptotic allocations of EUI and UOCBA policies.

Theorem 2 In the case of Normal-Normal Bayesian model with $\tau_i^0 \to \infty$ and $N \to \infty$, let n_i^{EUI} and n_i^{UOCBA} denote, respectively, the budget allocated to alternative i under the EUI and UOCBA policies in Equation (11) and Theorem 1, we have

$$\frac{n_i^{UOCBA}}{n_j^{UOCBA}} \to \frac{n_i^{EUI}}{n_j^{EUI}},\ \forall i \neq j,\ i \neq 1,\ j \neq 1.$$

Proof. When $\tau_i^0 \to \infty$, we have $t_i^n = \bar{y}_i, \tau_i^n = \frac{\sigma_i}{\sqrt{n_i}}$. When $n_i \to \infty$, a direct application of the Δ-technique yields $U(\mu_i) \xrightarrow{\mathscr{D}} \mathscr{N}\left(U(\bar{X}_i), \left(|\frac{\partial U}{\partial \mu_i}| \cdot \frac{\sigma_i}{\sqrt{n_i}}\right)^2\right)$. The EUI computation is then effectively a usual EI computation on the random variable U with normal posterior densities. Apply the results in Ryzhov (2016) on convergence rates of EI methods yields the theorem. □

Theorem 2 only holds for sub-optimal alternatives ($i, j \neq 1$). Peng and Fu (2017) derived variants of EI policies that achieve asymptotically optimal allocation ratio for all alternatives, which can be easily applied to our scenario. We use the most basic EI policy for simplicity.

4.3 Practical Computation of Expected Utility Improvement

The original expected improvement policy in Jones et al. (1998) was developed under a normal distribution assumption and has the closed-form expression

$$g^{EI} = (t - t^*)\Phi(\frac{t - t^*}{\tau}) + \tau\phi(\frac{t - t^*}{\tau}),$$

where t and τ are the mean and standard deviation, respectively, of the posterior normal density and t^* is the current threshold for improvement. It is easy to see that g^{EI} increases with τ for any $t \le t^*$, therefore favors alternatives with higher uncertainty. However, in Equation (11), a higher uncertainty in θ does not necessarily lead to higher EUI. One such example is $U(\theta) = -e^{-4\theta} - \theta$ which we tested in Section 6.2. As illustrated in Figure 1, a higher uncertainty in θ in terms of higher variance leads to a smaller g^{EUI},

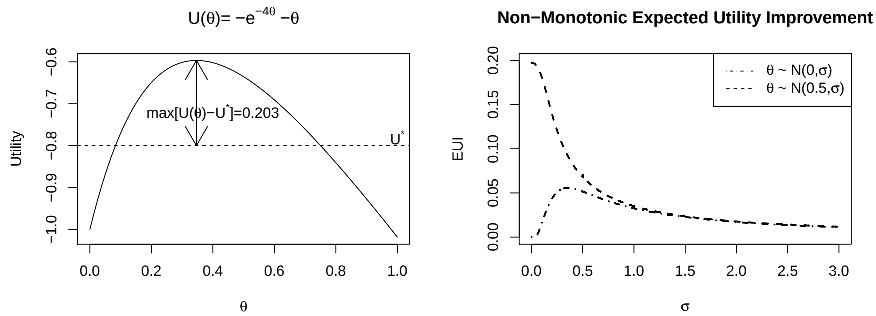

Figure 1: A utility function (left), and higher uncertainty leads to lower EUI (right).

causing the EUI aproach to fail in the numerical experiments in Section 6.2. The EUI approach no longer encourages exploration and we defer addressing the issue to future research.

5 UTILITY-BASED ALLOCATION ALGORITHMS

Using Theorem 1 and Equation (11), we design two fully sequential procedures for utility-based allocation problems, which we refer to as most-starving-UOCBA (MS-UOCBA) and EUI.

MS-UOCBA: An initial sampling budget n_0 is allocated to each alternative to obtain estimates of the unknown parameters in Theorem 1. At each time step, Equations (6) to (8) are solved to find the estimated optimal allocations under the given total budget. The alternative that is furthest away from its currently estimated optimal allocation is selected. This "most-starving" implementation is fully sequential except for the initialization batch of samples (Chen and Lee 2011).

EUI: The algorithm requires inputs for specifying the priors and posteriors updates, and therefore depends on the specific Bayesian model used for specific application problems.

Algorithm 1: $MS-UOCBA$

Input: total budget N, initial budget n_0, utility function U
Output: the final selection i^N

Allocate n_0 to each alternative
Compute $\hat{\theta}_i$, $U(\hat{\theta}_i)$ and $\nabla U(\hat{\theta}_i)$
Set counter $n \leftarrow kn_0$ and $m_i \leftarrow n_0,\ i = 1,2,\ldots,k$
while $n \le N$ **do**
  compute n_i by solving Equations (6) to (8)
  select $i_n \leftarrow \arg\max\{n_i - m_i\}$
  update $m_{i_n} \leftarrow m_{i_n} + 1$ and $n \leftarrow n+1$
  update $\hat{\theta}_{i_n}$, $U(\hat{\theta}_{i_n})$, $I(\hat{\theta}_{i_n})$, and $\nabla U(\hat{\theta}_{i_n})$
return $i^N = \arg\max_i\{U(\hat{\theta}_i)\}$

Algorithm 2: EUI

Input: the priors f_i^0, total budget N, utility function U
Output: the final selection i^N

Initialization: set $n = 0$ and select $i_0 = \arg\max_{i \le k} v_i^{EUI,0}$
while $n \le N$ **do**
  collect one sample on alternative i_n as x_{i_n}
  update $f_i^n(\theta_i)$ for $i = i_n$
  set $f_i^n = f_i^{n-1}$ for $i \ne i_n$
  compute $v_i^{EUI,n}$ with Equation (11)
  select $i_n = \arg\max_{i \le k}\{v_i^{EUI,n}\}$
  update $n \leftarrow n+1$
return $i^N = \arg\max_{i \le k} \mathbb{E}[U(\theta_i)]$.

Computational considerations: in the MS-UOCBA algorithm, the most computationally intensive step would be solving for optimal allocation using Equations (6) to (8). However, for the Bayesian approach, the posterior updates and computation of $v_i^{EUI,n}$ could be non-trivial, depending on the exact form of utility and prior-posterior pairs. We assume relevant computations are more efficient compared to obtaining an output from the simulation model, thus justifying the overhead for efficient sequential allocation.

6 NUMERICAL EXPERIMENTS

We test the performance of Algorithm 1 and Algorithm 2 on two simulated experiments by comparing their performance with the simple equal allocation (EA) and usual OCBA allocation policies. The first experiment selects from alternatives with binary rewards and employs the beta-Bernoulli conjugate pair outlined in Equation (10) when implementing EUI. The second works with continuous random rewards and uses the Normal-Normal Bayesian model in Equation (9).

6.1 Binary Rewards with Prospect Theoretic Utility

Let $\mathscr{A} = \{1,2,\ldots,k\}$ denote a set of lotteries and $Y_i, i \leq k$ be the Bernoulli random variable representing the outcome of a lottery ticket. We have

$$Y_i = \begin{cases} 0, & \text{w.p. } 1-p_i, \\ 1 & \text{w.p. } p_i, \end{cases}$$

where p_i are the unknown winning probabilities. Let a_i denote the winning prize of lottery i and b_i be the cost of buying a lottery ticket. The prospect-theoretic utility (Tversky and Kahneman 1992) for a lottery has the form

$$U(p_i) = (a_i - b_i)p_i^{w_1} - b_i(1-p_i)^{w_2}$$

with weights w_1 and w_2 reflecting people's perception of gains and losses. In the context of R&S, we assume a customer chooses from 19 different lotteries before committing to a favorite one with preference modeled by a prospect-theoretic utility with weights $w_1 = 1.1$ and $w_2 = 100$. The winning probabilities are set to be $\{0.05, 0.1, 0.15..., 0.90, 0.95\}$ with rewards $a_i = 1/p_i$ and cost $b_i = 1$, $\forall 1 \leq i \leq 19$, such that the expected net gain of all lotteries will be 0. Under the utility above, the optimal lottery will be the 2nd with $p_2 = 0.1$ and $a_2 = 10$ in this setting, as it offers a large reward as well as a reasonable chance of winning. The selection problem is further illustrate in Figure 2 (Left).

Implementation of MS-UOCBA: Given observed outcomes of Y_i, UOCBA first estimates the winning probabilities, and then updates the estimation of v_i, n_i for dynamic allocation. For lottery i, after n_i trials, we have

$$\hat{p}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}, \quad v_i = |w_1(a_i - b_1)p_i^{w1-1} + b_i(1-p_1)^{w2-1}|\sqrt{\hat{p}_i(1-\hat{p}_i)}).$$

We choose the initialization budget kn_0 to be 20% of the total budget N.

Implementation of EUI: We set the priors on p_i to be the flat beta distribution $Beta(1,1)$ for all $i \in \mathscr{A}$. Upon collecting new observations, the update is performed according to Equation (10). Given a beta posterior with shape parameters α and β, the posterior expected utility has the closed-form formula

$$\mathbb{E}[U(p_i)] = \frac{(a_i - b_i)B(\alpha + w_1, \beta) - b_i B(\alpha, \beta + w_2)}{B(\alpha, \beta)},$$

where $B(\alpha,\beta)$ is the Beta-function defined as $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$. The expected improvement in Equation (11) is computed with numerical integration routine **integrate** in **R**. For posteriors with large shape parameters ($> 1,000$ in our experiments), a Monte Carlo integration is performed to compute the posterior expected utility and expected improvement with $1,000$ samples to avoid the numerical instability with extraordinarily small ($< 10^{-100}$) $B(\alpha, \beta)$.

We compare the performance of MS-OCBA and EUI using the above implementations with the Equal Allocation (EA) policy and the MS-OCBA policy for budgets ranging from 100 to $10,000$. For the MS-OCBA policy, we use the same implementation with MS-UCOA with $U(\hat{p})$ set to be $\hat{p}$ and v_1 set as $\sqrt{p_i(1-p_i)}$. The PCS for each allocation algorithm is estimated using $1,000$ simulation replications. The

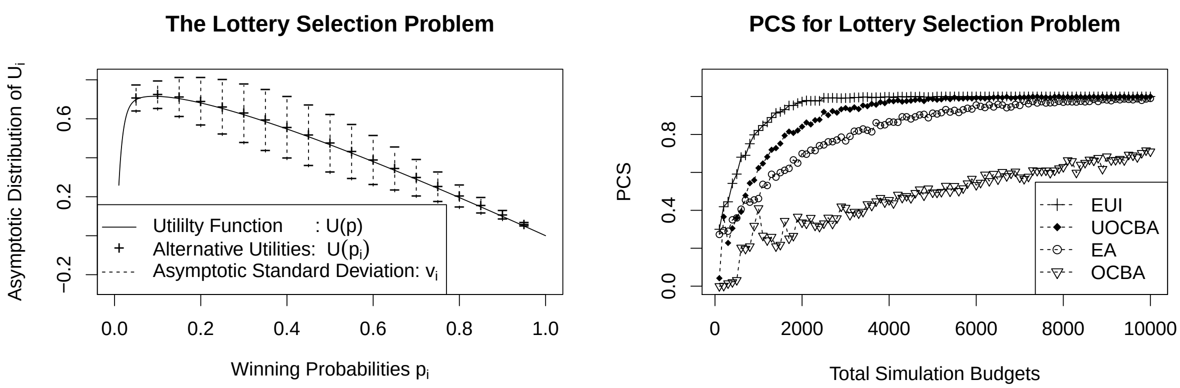

Figure 2: The prospect utility function $U(\cdot)$ and asymptotic distribution of the utility estimator $U(\hat{p}_i)$ (left), and simulated PCS (right).

simulation results are presented in Figure 2 (right). EUI has the best performance among all policies, with MS-UOCBA being the second best. The usual MS-OCBA policy has the worst performance, as it devotes most of its allocation budget on alternatives with large p_i, which have relatively low values under the given utility measure.

6.2 Staffing with a Cost Utility

A company staffing a service center is often faced with the trade off between quality of services and staffing costs (such as training and compensation). Let Y_i denote the service times, we set

$$Y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

where μ_i is unknown and σ_i is assumed to be 1. A smaller value of μ_i indicates higher service quality but requires higher training and wage costs. A utility function of the form $U(\mu_i) = -C(\mu_i) - Q(Y_i)$ in Fu et al. (2005) is often used to capture the trade off, where $C(\cdot)$ denote the cost and $Q(\cdot)$ denote the service quality. We use negative terms to make the problem a maximization rather than minimization, for consistency with our problem setting. We test with two utilities

$$U_1(\mu) = e^{10\mu - 10}, \qquad U_2(\mu) = -e^{-4\mu} - \mu,$$

where U_1 monotonically increases with μ and U_2 balances between cost and quality.
Implementation of MS-OCBA: Given collected samples drawn from Y_i, MS-UOCBA estimates μ_i with sample averages, and v_i for two two utilities can be easily computed to be

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}, \quad v_i^{U_1} = |10e^{10\mu-10}|\sigma, \quad v_i^{U_2} = |4e^{-4\mu} + 1|\sigma.$$

The initialization budget kn_0 is chosen to be 20% of the total budget N..
Implementation of EUI: Given normal random observations, we use the Normal-Normal model outlined in Equation (9). The priors are set to be $t_i^{(0)} = 0$ and $\tau_i^{(0)} = 1,000$ to create flat priors. Given a posterior $\mathscr{N}(t_i^n, (\tau_i^n)^2)$, the posterior expected utilities can be computed as

$$\mathbb{E}[U_1(\mu_i)] = e^{10t_i^n+50(\tau_i^n)^2}, \quad \mathbb{E}[U_2(\mu_i)] = -t_i^n - e^{-4t_i^n+2(\tau_i^n)^2(t_i^n)^2}.$$

EUI in Equation (11) for U_1 and U_2 under the assumed Normal-Normal conjugate pair also has the closed form expressions

$$EUI(U_1^*) = e^{-10}[1-\Phi(\frac{t_i^n - x_1^c}{\tau_i^n})] + e^{10t_i^n+50(\tau_i^n)^2-10}[1-\Phi(\frac{t+10(\tau_i^n)^2 - x_1^c}{\tau_i^n})],$$
$$EUI(U_2^*) = (U_2^* + t_i^n)\left(\Phi\left(\frac{x_2^l - t_i^n}{\tau_i^n}\right) - \Phi\left(\frac{x_2^u - t_i^n}{\tau_i^n}\right)\right) + \tau_i^n\left(\phi\left(\frac{x_2^l - t_i^n}{\tau_i^n}\right) - \phi\left(\frac{x_2^u - t_i^n}{\tau_i^n}\right)\right)$$
$$+ e^{4t_i^n+8(\tau_i^n)^2}\left(\Phi\left(\frac{x_1^l - t_i^n - 4(\tau_i^n)^2}{\tau_i^n}\right) - \Phi\left(\frac{x_1^u - t_i^n - 4(\tau_i^n)^2}{\tau_i^n}\right)\right),$$

where $U_1(x_1^c) = U_1^*$ and $U_2(x_2^l) = U_2(x_2^u) = U_2^*$ with $x_2^l \le x_2^u$. Under this notation $[x_1^c, \infty)$ and $[x_2^l, x_2^u]$ will be the region where improvement function is positive for computing EUI for utility functions U_1 and U_2. The priors are chosen to be $\mathscr{N}(0,4)$ for initializing the unknown μ_i inside the *interesting* region. At each time step, the alternative with maximum EUI is chosen. If there is a tie in EUI, one of the alternatives with the maximum EUI is randomly selected. In both tests, EUI allocation steps are terminated early due to its clear convergence.
Implementation of MS-OCBA and EA: The MS-OCBA algorithms are implemented using the same setup for MS-UOCBA. For total simulation budgets ranging from 100 to 10,000, each policy is simulated 1,000 replications to estimate PCS. The numerical results are presented in Figure 3.

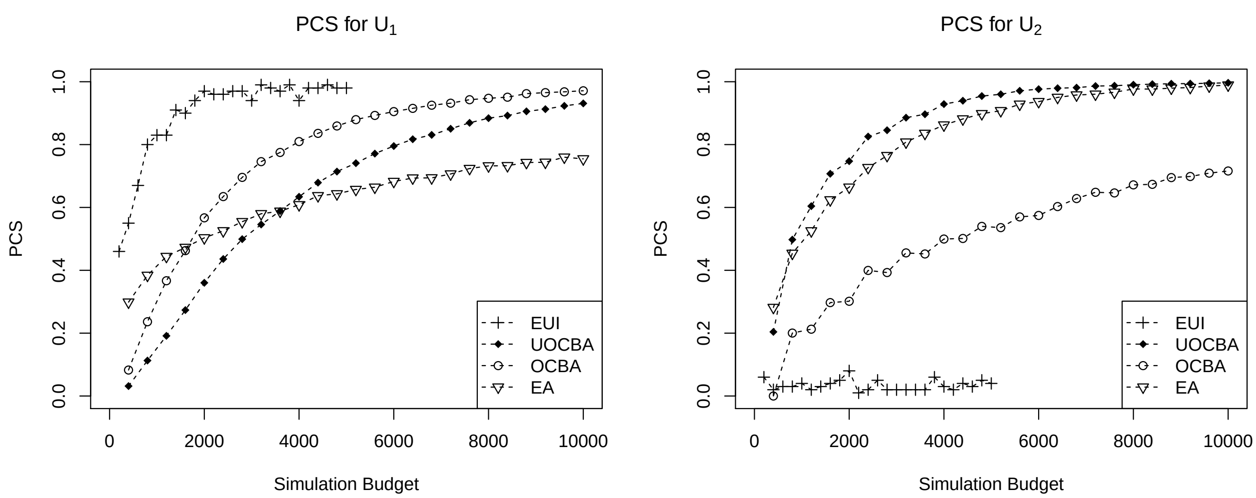

Figure 3: Performance of allocation algorithms for the two utilities U_1 and U_2.

For U_1, the EUI algorithm outperforms all other algorithms. MS-OCBA outperforms MS-UOCBA, as U_1 being a monotoe function, both MS-OCBA and MS-UOCBA will treat alternative 20 as the true optimal choice and allocate budgets on alternatives closer to 20. Despite being optimal under their respective selection rule (OCBA with $i^N = \arg\max\{\hat{\mu}_i\}$ and UOCBA with $i^N = \arg\max\{U_1(\hat{\mu}_i)\}$), there is no easy theoretical analysis on their relative performance. The EA policy outperforms both MS-OCBA and MS-UOCBA when the total budget is small, as OCBA policies are known to be sensitive to initialization noise. For U_2, MS-UOCBA is the best among all tested algorithms. Despite the asymptotic optimality result in Theorem 2, the shape of U_2 causes the EUI policy to fail in this problem, as higher uncertainty leads to a smaller expected utility improvement, causing the algorithm to perform similar to a uniform random choice, as explained in detail in Section 4.3.

7 CONCLUSION

In this paper, we considered the R&S problem where the selection objective can be expressed as a utility function of the observed random samples. We established a plug-in utility estimator to derive an asymptotically optimal allocation policy and provided insight on how a utility function would affect the optimal allocation policy. The main result in Theorem 1 can be easily extended to cases where the MLE is not easy to obtain, but some other parameter estimator with the same convergence rate, such as the methods of moments estimator, could still be applied. We also developed a variation of the Bayesian VoI approach that showed better finite budget performance. We proposed two sequential allocation algorithms and discussed their practical implementations, including a scenario where the Bayesian VoI approach could fail. To the best of our knowledge, this is the first attempt at extending R&S techniques for general utility measures.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant CMMI-1434419 and by the Defense Advanced Research Projects Agency (DARPA) under Grant N660011824024. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

REFERENCES

Casella, G., and R. L. Berger. 2002. *Statistical Inference*, Volume 2. Boston: Cengage Learning, Inc.

Chen, C.-H., and L. H. Lee. 2011. *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*, Volume 1. Singapore: World Scientific.

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

Chick, S. E. 2006. "Subjective Probability and Bayesian Methodology". *Handbooks in Operations Research and Management Science* 13:225–257.

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

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

Fu, M. C., F. W. Glover, and J. April. 2005. "Simulation Optimization: A Review, New Developments, and Applications". In *Proceedings of the 2005 Winter Simulation Conference*, edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines: Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Jie, C., L. Prashanth, M. Fu, S. Marcus, and C. Szepesvári. 2018. "Stochastic Optimization in a Cumulative Prospect Theory Framework". *IEEE Transactions on Automatic Control* 63(9):2867–2882.

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

Kawabe, T., Y. Kobayashi, S. Tsuruta, Y. Sakurai, and R. Knauf. 2015. "Case Based Human Oriented Delivery Route Optimization". In *IEEE Congress on Evolutionary Computation*, 2368–2375.

Kim, S.-H., and B. L. Nelson. 2001. "A Fully Sequential Procedure for Indifference-zone Selection in Simulation". *ACM Transactions on Modeling and Computer Simulation* 11(3):251–273.

Law, A. M., and W. D. Kelton. 2007. *Simulation Modeling and Analysis*. 4th ed. New York: McGraw-Hill, Inc.

Pasupathy, R., R. Szechtman, and E. Yücesan. 2010. "Selecting Small Quantiles". In *Proceedings of the 2010 Winter Simulation Conference*, edited by S. Jain, R. Creasey, J. Himmelspac, K. White, and M. Fu, 2762–2770: Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Peng, Y., and M. C. Fu. 2017, April. "Myopic Allocation Policy with Asymptotically Optimal Sampling Rate". *IEEE Transactions on Automatic Control* 62(4):2041–2047.

Powell, W. B., and I. O. Ryzhov. 2012. *Optimal Learning*, Volume 841. New Jersy: John Wiley & Sons.

Qu, H., I. O. Ryzhov, M. C. Fu, and Z. Ding. 2015. "Sequential Selection with Unknown Correlation Structures". *Operations Research* 63(4):931–948.

Rockafellar, R. T., and S. Uryasev. 2000. "Optimization of Conditional Value-at-risk". *Journal of Risk* 2:21–42.

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

Trailovic, L., and L. Y. Pao. 2004. "Computing Budget Allocation for Efficient Ranking and Selection of Variances with Application to Target Tracking Algorithms". *IEEE Transactions on Automatic Control* 49(1):58–67.

Tversky, A., and D. Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty". *Journal of Risk and Uncertainty* 5(4):297–323.

AUTHOR BIOGRAPHIES

GUOWEI SUN is a Ph.D. candidate in Applied Mathematics, Statistics and Scientific Computation at the Department of Mathematics at the University of Maryland, College Park. His interests lie in the broad area of stochastic optimization and applied probability. His email is gwsun@math.umd.edu.

YUNCHUAN LI is a Ph.D. candidate in the Department of Electrical and Computer Engineering at the University of Maryland, College Park. His email is yli93@umd.edu.

MICHAEL C. FU holds the Smith Chair of Management Science in the Robert H. Smith School of Business, with a joint appointment in the Institute for Systems Research, A. James Clark School of Engineering, University of Maryland, College Park. He has a Ph.D. in applied math from Harvard and degrees in math and EECS from MIT. He attended his first WSC in 1988 and served as WSC2011 Program Chair, NSF Operations Research Program Director, *Management Science* Stochastic Models and Simulation Department Editor, and *Operations Research* Simulation Area Editor. He is a Fellow of **INFORMS** and **IEEE**. His e-mail address is mfu@umd.edu, and his Web page is https://www.rhsmith.umd.edu/directory/michael-fu.