Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

SAMPLE ALLOCATION FOR MULTIPLE ATTRIBUTE SELECTION PROBLEMS

Dennis D. Leber

National Institute of Standards and Technology
100 Bureau Drive, MS 8980
Gaithersburg, MD 20899, USA

Jeffrey W. Herrmann

A. James Clark School of Engineering
University of Maryland
Martin Hall Room 2181
College Park, MD 20742, USA

ABSTRACT

Prior to making a multiple attribute selection decision, a decision-maker may collect information to estimate the value of each attribute for each alternative. In this work, we consider a fixed experimental sample budget and address the problem of how best to allocate this budget across three attributes when the attribute value estimates have a normally distributed measurement error. We illustrate that the allocation choice impacts the decision-maker's ability to select the true best alternative. Through a simulation study we evaluate the performance of a common allocation approach of uniformly distributing the sample budget across the three attributes. We compare these results to the performance of several allocation rules that leverage the decision-maker's preferences. We found that incorporating the decision-maker's preferences into the allocation choice improves the probability of selecting the true best alternative.

1 INTRODUCTION

The problem of selecting the system (alternative) with the largest probability of actually being the best is known as the multinomial selection problem (Kim and Nelson 2006). In the selection problem, the performance measure must be inferred by sampling using a simulation model or other stochastic process. The problem is made more complicated when the performance measure is a function of multiple, uncertain attributes that are sampled separately. We call the process of sampling one attribute a "measurement." Leber and Herrmann (2013, 2014) described the challenge of selecting a radiation detection system in which this problem occurred, but the problem is not limited to this particular application.

The variability in each measurement process generates a range of results, which leads to uncertainty about the attributes that are relevant to the selection problem. (If the measurement process had no variability, then the measured value would equal the attribute's true value, and there would be no uncertainty.) The decision-maker can reduce the attribute value uncertainty by making additional measurements, which yields more information. When the budget for measurements is limited, however, tradeoffs must be made. Thus the allocation of measurement effort (sample allocation) across the multiple decision attributes plays an important role in maximizing the probability of selecting the truly best alternative.

While much of the recent ranking and selection literature has focused on problems as they pertain to computer simulation experiments, the procedures are also applicable to physical experimentation (see Bechhofer, Santner, and Goldsman 1995). It is this setting for which the work desribed in this paper is most obviously applicable. Consider, for example, the selection problem faced by the Domestic Nuclear Detection Office (DNDO) of the U.S. Department of Homeland Security when the United States Congress mandated that the DNDO work with the U.S. Customs and Border Protection (CBP) to evaluate and improve radiation detection systems in U.S. based international airports. As a result of this mandate, the DNDO initiated the PaxBag pilot program to identify the best possible system design for detecting, identifying, and localizing illicit radiological or nuclear material entering the United States through international passenger and baggage screening. This challenge was met by testing and evaluating, in a laboratory environment, available radiation detection equipment suitable for such an application, followed by an operational demonstration of the system that displayed the strongest potential for improved capability over currently deployed technology. To select the radiation detection system to put forth for the operational demonstration, DNDO and CBP formulated a multiple attribute decision model and developed a laboratory experimental plan to support the estimation of the true attribute values. This led to the following question: how should the limited laboratory experimental budget be allocated across the multiple alternatives and multiple attributes to generate information that leads to selecting the true best system? This question, which is not limited to the selection of a radiation detection system, applies to all decision processes where the true values of multiple attributes are estimated based upon experimental evaluations.

In the following section we indicate how the problem of allocating the measurement effort (the sample allocation problem) considered in this paper differs from the extensive work done in the field of ranking and selection. The study described in this paper extends the results from our previous work with pass-fail testing on two attributes (Leber and Herrmann 2013) and normally distributed measurement error with two attributes (Leber and Herrmann 2014) to address the sample allocation problem for a three attribute selection decision with normally distributed measurement error where the measurement variance is assumed to be known. Details of this problem setting are provided in Section 3. Section 4 describes the simulation study that we designed to determine how well different procedures used to determine the allocation of experiments for the evaluation of the attribute values perform and is part of a larger study of this problem. Results from our simulation study and conclusions are presented in the final sections of this paper.

2 RANKING-AND-SELECTION AND STATISTICAL EXPERIMENT DESIGN

The problem studied herein is a type of ranking-and-selection problem. The ranking problem is to generate a complete ordering of a set of alternatives, when performance is a random variable and an alternative's true performance must be estimated using experimentation – either physical measurements or computer simulation. The selection problem is to find the best of these alternatives. The result of an experiment can be used to estimate $y_j = f\left(A_j\right)$, where y_j is the true value of the response variable (performance) for A_j, the j^{th} alternative within the given set of alternatives. When the total number of available experimental runs (samples) is limited, the problem is to determine how many experimental runs should be allocated to each alternative. The indifference zone (IZ), the expected value of information procedure (VIP), and the optimal computing budget allocation (OCBA) are sequential approaches that have been developed to find good allocation solutions (see Bechhofer, Santner, and Goldsman 1995; Kim and Nelson 2006; Branke, Chick, and Schmidt 2007). In these approaches, the problem is to determine which alternatives should be observed (simulated) next and when to stop. Computational results presented by Branke, Chick, and Schmidt (2007) demonstrated the strengths and weaknesses of these procedures. Laporte, Branke, and Chen (2012) developed a version of OCBA that is useful when the computing budget is extremely small. Chen et al. (2008) developed a version of OCBA that can be used to find the best m alternatives efficiently. Lee et al. (2004, 2010) considered the problem of finding the set of non-dominated alternatives when there are multiple objectives and developed approaches for allocating simulation replications to different alternatives. Although these approaches have some similarities to the problem that the current paper considers, we are exploring how the allocation of simulation replications to different attributes (which are combined in a single aggregate value function) affects the probability of selecting the truly best alternative. This paper describes a computational study; analytical approaches like OCBA are being developed and will be described in future work.

As described in the next section, the selection problem considered here is concerned with the allocation of information-gathering resources across the different attributes, not the different alternatives. Given a set of alternatives, each described by k attributes, the decision-maker's value for a particular alternative A_j may be represented by $y_j = f\left(A_j\right) = v\left(x_{j1},\ldots,x_{jk}\right)$. Instead of directly observing and estimating the alternative's performance measure, y_j, we can only estimate the alternative's multiple true attribute values, $x_{j1},\ldots,x_{jk}$, based on different information-gathering tasks (e.g., experiments). The estimated attribute values are then combined through a multiple attribute decision (value or utility) model to provide an alternative's overall performance measure (see Butler, Morrice, and Mullarkey 2001 as an example of this approach for the selection problem). Our challenge is to determine how many experiments should be allocated to the evaluation of each attribute.

The statistical design of experiments provides the foundation for defining experimental factors and levels in developing a design space, identifying optimal locations to sample within the design space, and determining the appropriate sample size. Box, Hunter, and Hunter (2005) and Montgomery (2013) provide extensive guidance for the principles and methods of statistical design of experiments. These problems can be represented by $y = f\left(l_1,\ldots,l_p\right)$, where y is the response variable of interest, p is the number of multiple level experimental factors under study, and l_i is the level of the i^{th} experimental factor. A primary focus of the design of experiments discipline is how to best allocate the total budget of N measurements across the design space defined by the factors and their levels. The designer must choose which particular combinations of factors and levels will be included in the experiment. Bayesian experimental design (Chaloner and Verdinelli 1995) is an alternative to classical experimental design that leverages the information available prior to experimentation to find the best set of factors and levels, and to determine the appropriate sample size.

3 PROBLEM STATEMENT

As classified by Roy (2005), the decision problem we consider is one of choice: given a set of alternatives, $\left\{A_1,\ldots,A_m\right\}$, $m \geq 2$, the decision-maker will select a single alternative. Each alternative A_j is described by attributes, $X_1,\ldots,X_k$, $k \geq 2$, which are quantified by specific attribute values, $x_{j1},\ldots,x_{jk}$, and by its overall value (utility), as determined by $y_j = v\left(x_{j1},\ldots,x_{jk}\right)$. The decision-maker prefers the alternative that has the greatest overall value. We assume that the corresponding tradeoffs condition is satisfied (Keeney and Raiffa 1993), and hence an additive value function of the form displayed in Equation (1) is a valid model of the decision-maker's preferences. Let x_i be the value of attribute X_i, let λ_i be the weight of attribute X_i, and let $v_i\left(x_i\right)$ be the individual value function for attribute X_i, for $i = 1,\ldots,k$. Then the decision-maker's overall value for alternative A_j is:

$$y_j = v\left(x_{j1},\ldots,x_{jk}\right) = \lambda_1 v_1\left(x_{j1}\right) + \cdots + \lambda_k v_k\left(x_{jk}\right) \tag{1}$$

The individual value functions $v_i\left(x_i\right)$ in Equation (1) map the attribute values, which are determined by the characteristics of the alternative, to decision values, and are scaled such that $v_i\left(x_i^0\right) = 0$ for the least desirable attribute value, x_i^0, and $v_i\left(x_i^*\right) = 1$ for the most desirable attribute value, x_i^*. The attribute weights, λ_i, reflect the decision-maker's preferences and satisfy the constraint $\sum_{i=1}^{k} \lambda_i = 1$.

While true values for the k attributes exist for each alternative, they are unknown to the decision-maker and will be estimated through a series of experiments (measurements). In this setting, a "measurement" is an information-gathering activity that provides a value for one attribute of one alternative. Due to randomness in the measurement process, the observed value is a random variable that is influenced by the measurement process but depends primarily upon the true value of the attribute for that alternative. The uncertainty associated with the attribute (*attribute value uncertainty*) is a function of the values that are collected from experimentation. (More measurements gather more information about an attribute and will reduce the uncertainty of the estimate for the true attribute value.) The information that is gathered (the measurements or experimental results) are used to model the uncertainty of the estimated attribute values. This uncertainty leads to uncertainty in an alternative's overall value.

We assume that the decision-maker is concerned with finding the best alternative and is thus facing a selection problem. Furthermore, we assume that, to make his decision, the decision-maker prefers (and will select) the alternative that has the greatest probability of being the best among the given set of alternatives. (Of course, there are other preferences that may be considered, each with their own virtues, but that is beyond the scope of this paper.) To estimate this probability, while propagating the attribute value uncertainty through the decision model, we use a very generalizable Monte Carlo approach. Further details of this approach are provided in Section 4.2, with a complete discussion found in (Leber and Herrmann 2012).

If the budget for measurements is sufficiently large, then the decision-maker can gather enough information about every attribute of every alternative to reduce the attribute value uncertainty to a point where it is clear which alternative is truly the best. In practice, however, especially when measurement or experiments are expensive, this is not possible. For this work, we assume that the budget is fixed and all measurements (experimentation) will occur in a single phase. We will be considering sequential allocation policies in future work.

The sample allocation problem for multiple attribute selection problems can be stated as follows: The overall budget in terms of measurements, B, is fixed and will be divided equally among the m alternatives. The budget for each alternative must be further divided among the k attributes. In general, the budgets for different alternatives could be divided differently, but we made the simplifying assumption that the allocation is the same for all alternatives (this constraint will be relaxed in future work). For a given alternative, let n_i denote the number of measurements (samples) of attribute X_i. Let $N = B/m$ denote the total number of measurements for each alternative, thus, $n_1 + \cdots + n_k = N$. The problem is to find values $n_1, \ldots, n_k$ that maximize the probability that the decision-maker will choose the truly best alternative (the probability of correct selection), given the decision-maker's values and preferences.

4 SIMULATION STUDY

In general, obtaining more measurements on those attributes that have the most uncertainty and are the most important to the decision-maker is an obvious strategy for allocating the overall budget. To test this intuition, we conducted a simulation study to understand how the sample allocation affects the probability of correct selection. The following subsections briefly describe the details of the simulation study and the sample allocation rules that were tested.

We considered the situation in which an alternative is described by three attributes, X_1, X_2, and X_3, and each attribute is measured using a different technique. The error of each measurement technique is normally distributed with the variance assumed to be known. The alternatives, when characterized by their true values of X_1, X_2, and X_3, form a concave efficient frontier in R^3 space. The attributes share a common domain and the individual value functions $v_1(x_1)$, $v_2(x_2)$, and $v_3(x_3)$ were defined to be linear. The overall value for alternative A_j can be expressed as $y_j = v(x_{j1}, x_{j2}, x_{j3}) = \lambda_1 x_{j1} + \lambda_1 x_{j2} + \lambda_1 x_{j3}$.

4.1 Training Cases and Measurement Error

We generated a set of 20 training cases (sets of alternatives), evaluated every possible sample allocation, and used the results to generate insights for developing sample allocation rules. Each training case consisted of five alternatives described by three attributes. The true values of the attributes were randomly assigned from the domain of [100, 200], subject to the constraints necessary for non-dominance and concavity. The algorithm used to generate a concave efficient frontiers in R^3 space is as follows:

1. An attribute space was defined for each attribute X_i, i = 1, 2, 3, by:
   a. The distance between the minimum attribute value and the maximum attribute value, denoted $dist_i$, was randomly selected from a *Uniform*[0,100] distribution.
   b. The minimum attribute value, x_{imin}, was randomly selected from a *Uniform*[100, 200 – $dist_i$] distribution.
   c. The maximum attribute value, x_{imax}, was determined by $x_{imin} + dist_i$.
2. A normalized space was defined such that the domain of each variable, Z_i, i = 1, 2, 3 is [0, 1].
3. A random concave surface in normalized space was defined by the curve $z_1^s + z_2^s + z_3^s = 1$, where s was generated by randomly selecting a value r from a *Beta*[1, 2] distribution and setting $s = 9r + 1$ so that min(s) = 1 and max(s) = 10. (The expected value of s was 4.)
4. The normalized attribute values (z_1, z_2, z_3) for each of five alternatives were randomly selected from the concave surface. For each alternative the following steps were performed:
   a. A value of z_1 was randomly drawn from a *Uniform*[0, 1] distribution.
   b. A value of z_2 was randomly drawn from a $Uniform\left[0, \sqrt[s]{1 - z_1^s}\right]$ distribution.
   c. $z_3 = \sqrt[s]{1 - z_1^s - z_2^s}$.
5. The normalized attribute values were translated to the attribute space that was defined in step 1 by:
   a. Assigning $x_1 = z_a$, $x_2 = z_b$, $x_3 = z_c$, where (a, b, c) is a random permutation of (1, 2, 3), with each permutation having equal probability.
   b. Scaling (by $dist_i$) and shifting (by x_{imin}) each x_i, i = 1, 2, 3.

For our simulation, each attribute was measured with a different measurement technique and it was assumed that the technique maintained a measurement variability that was consistent across all alternatives measured. We set the actual measurement variance of each of the three attributes (σ_1^2, σ_2^2, and σ_3^2) to one of 10^2 or 30^2, which created $2^3 = 8$ different "measurement error scenarios."

4.2 Evaluating Sample Allocations

A sample generates one (random) measurement of one attribute of one alternative. Given a budget of $N = n_1 + n_2 + n_3 = 9$ samples for each alternative, the problem is to determine n_1, n_2, and n_3, the number of samples of attribute 1, attribute 2 and attribute 3, to maximize the probability of correct selection. That is, the decision-maker wants to maximize the likelihood of selecting the alternative whose true values of the attributes yield the greatest overall value as defined by Equation (1). As mentioned before, we assume that, given the uncertainty in the attribute values, the decision-maker prefers the alternative that is most likely to have the greatest overall value (the best performer) in any single trial.

We evaluated, using the 20 training cases, all of the possible sample allocations (55) for N = 9 total samples per alternative, (n_1, n_2, n_3) = (0, 0, 9), (0, 1, 8), …, (9, 0, 0), over a range of values of λ_1 , λ_2 and λ_3, the weights in the decision value function. In particular, we considered 39 decision weight pairs (λ_1, λ_2, λ_3) = (0.1, 0.1, 0.8), (0.1, 0.2, 0.7), …, (0.8, 0.1, 0.1), (0.05, 0.05, 0.9), (0.05, 0.9, 0.05), (0.9, 0.05, 0.05). To do this, for each case (20), measurement error scenario (8), and sample allocation (55) – a total of 8800 combinations – we simulated 1000 sets of measurements. (Henceforth, a case under a particular measurement error scenario is referred to as a subcase.) Each set included 45 measurements, 9 for each of the 5 alternatives, with n_1 measurements observed from attribute 1, n_2 measurements observed from attribute 2, and n_3 measurements observed from attribute 3. Each measurement was created by observing a single random draw from a normal distribution with a mean equal to the true attribute value and a variance defined by the measurement error scenario.

Upon observing the sample measurements, we modeled the attribute value uncertainty, propagated this uncertainty through the decision model and selected an alternative for each set of sample measurements. The uncertain attribute values were modeled, a priori, with a normal distribution with mean of 150 and variance of 35^2. A Bayesian conjugate prior model for normally distributed data (Gelman, et al. 2004) was then used to update the attribute value models based on the observed sample measurements to provide posterior distributions.

The uncertainty was propagated through the decision model and onto the decision value parameter by drawing 1000 Monte Carlo samples from the posterior distributions of each of the three attributes and calculating the overall decision value of the alternative using each of the 39 decision value functions (as defined by the 39 decision weight triplets). For each decision weight triplet, the alternative that most frequently displayed the best (largest) decision value across the Monte Carlo replications was selected and checked whether this alternative was the true best (the alternative whose true values of the attributes yield the greatest overall decision value for the given decision weight triplet). Repeating this selection process over all 1000 sets of measurements allowed us to define the *frequency of correct selection* (*fcs*) evaluation measure as the number of times that the best alternative had been selected divided by 1000 sets.

The result of this simulation was an *fcs* value for each of the 55 sample allocations, for each of the 39 decision weights, across the 160 subcases. For each of the 160 subcases and each of the 39 decision weights, there is at least one optimal sample allocation that produced the maximum *fcs* value. This optimal sample allocation should maximize the probability of choosing the true best alternative. For each subcase and decision weight, we defined the *relative frequency of correct selection* (*rel fcs*) for each sample allocations as the ratio of the frequency of correct selection for that sample allocation to the frequency of correct selection for the optimal allocation. Within the confines of the problem which include the alternatives' attribute values and the total budget, this relative frequency of correct selection measure allows us to quantify how much better the selection could have been if a different sample allocation were chosen.

The *rel fcs* values produced by the training cases were illustrated through a series of contour plots such as those presented in Figure 1. Each panel of Figure 1 displays the *rel fcs* values for the indicated training subcase under a single decision model defined by the decision weight pair λ_1 and λ_2 (recall that $\lambda_3 = 1 - \lambda_1 - \lambda_2$). Within each panel, the shaded contours present the *rel fcs* values as a function of n_1 and n_2, ranging from dark (low *rel fcs* values) to light (high, desirable *rel fcs* values). Note that results are only feasible in the region $n_2 \leq 9 - n_1$ since the overall budget $N = n_1 + n_2 + n_3 = 9$. The solid squares within the plots denote the optimal sample allocation for the decision model. For each decision model there is at least one, but potentially more than one optimal sample allocation.

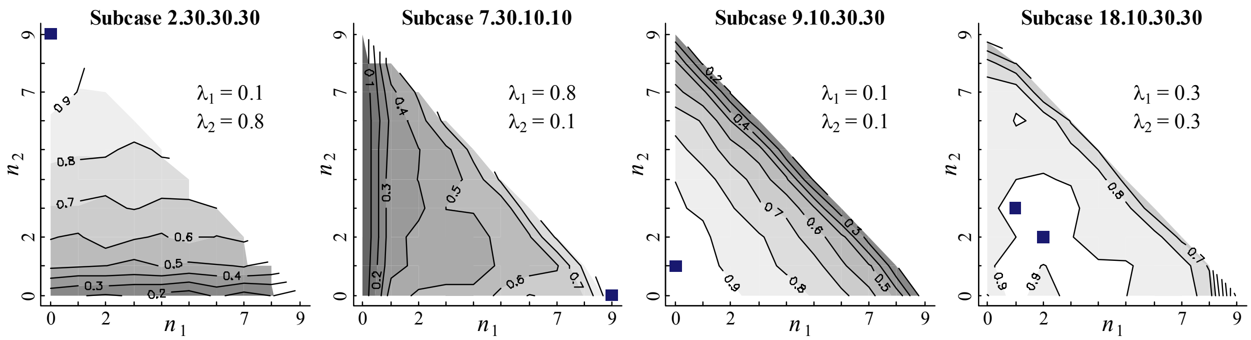

Figure 1: Contour plots displaying *rel fcs* as a function of n_1 and n_2 for training subcase 2.30.30.30 (case 2 with measurement error $\sigma_1^2 = 30^2, \sigma_2^2 = 30^2, \sigma_3^2 = 30^2$) under decision model ($\lambda_1 = 0.1$, $\lambda_2 = 0.8$, $\lambda_3 = 0.1$), subcase 7.30.10.10 under decision model (0.8, 0.1, 0.1), subcase 9.10.30.30 under decision model (0.1, 0.1, 0.8), and subcase 18.10.30.30 under decision model (0.3, 0.3, 0.4). The solid squares denote the optimal sample allocation for the decision model.

The immediate observation to be made from Figure 1 is that the choice in sample allocation matters. That is, the *rel fcs* for the selection problem is impacted by the choice in sample allocation. Consider, for example, Subcase 7.30.10.10 in Figure 1 where a sample allocation of $n_1 = 9$, $n_2 = 0$, $n_3 = 0$ is indicated to the be the optimal sample allocation. If a different sample allocation is selected, say $n_1 = 1$, $n_2 = 6$, $n_3 = 3$, then the *rel fcs* would be approximately 0.3 and hence, the probability of selecting the true best alternative (correct selection) would be reduced by nearly 70 %.

A second observation that can be made from the plots in Figure 1 is that when the decision models are such that high weight (high λ value) is placed on one of the attributes and the other two attributes receive low weight, the optimal allocation is to allocate all or nearly all of the budget (N samples) to the highly weighted attribute. This trend was seen repeatedly throughout the 160 training subcases.

4.3 Sample Allocation Rules

In general, the optimal sample allocation rule depends upon the information that the decision-maker has. If he has no information, the decision-maker will have no reason to allocate more samples to any attribute and would use a balanced allocation of $n_1 = n_2 = n_3 = N/3$. We refer to this sample allocation as the *uniform allocation rule*. This allocation is consistent with the principle of balance in the traditional design of experiments discipline.

If the decision weight values of λ_1, λ_2, and λ_3 are available, then the decision-maker may choose to assign n_1, n_2, and n_3 proportional to λ_1, λ_2, and λ_3. Observations made from contour plots resulting from the training cases (e.g., Figure 1) showed that, in the optimal sample allocation, the allocation to attribute i generally increased as λ_i increased. Since n_1, n_2, and n_3 must be integer values, rounding is necessary, e.g., $n_1 = \text{round}(\lambda_1 N)$, $n_2 = \text{round}(\lambda_2 N)$, $n_3 = N - n_1 - n_2$. We refer to this sample allocation approach as the *proportional allocation rule*. As an example of this allocation rule, when the decision weights are (0.1, 0.5, 0.4) and the budget $N = 9$, then the sample allocation equals (1, 5, 3).

The results from the training cases also showed that “extreme allocations” that allocate all of the budget to only one attribute (while the others are not evaluated) were optimal allocations for some of the 39 decision weight triplets, especially those in which one weight is near 1 while the other two weights are near 0. This observation was consistent with observations in previous work involving two attributes. We thus created two “zone” allocation rules that determined the allocation based on the decision weight values of λ_1, λ_2, and λ_3.

The *three-zone allocation rule* assigns the allocation (n_1, n_2, n_3) = (9, 0, 0) to decision weight triplets in which λ_1 is near 1, assigns the allocation (0, 9, 0) to decision weight triplets in which λ_2 is near 1, and assigns the allocation (0, 0, 9) to decision weight triplets in which λ_3 is near 1. The *four-zone allocation rule* assigns the same allocation as the three-zone allocation rule except for decision weight triplets in which all of the weights are between 0.2 and 0.4; to these triplets the rule assigns the allocation $(n_1, n_2, n_3) = (3, 3, 3)$. Figure 2 illustrates the sample allocations provided by the three- and four-zone allocation rules as a function of decision model.

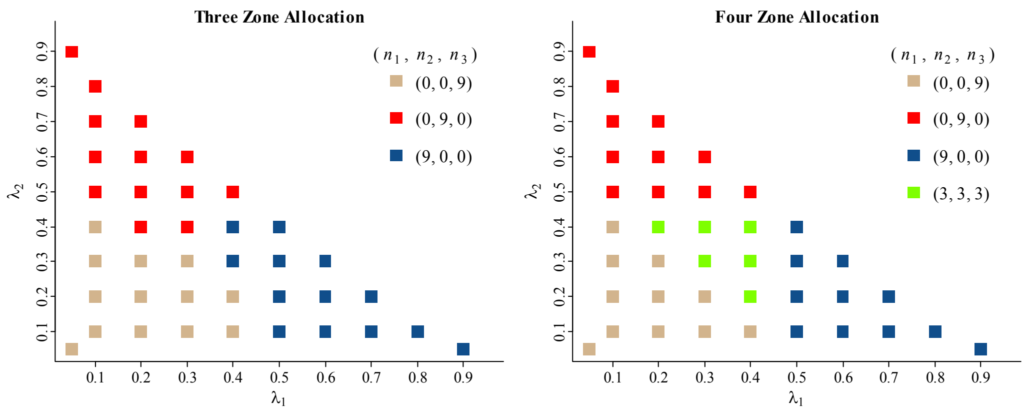

Figure 2: Sample allocation definitions for the three-zone (left) and four-zone (right) allocation rules.

4.4 Testing the Sample Allocation Rules

To test the sample allocation rules, we generated 500 new concave frontiers (testing cases). Each case was a set of 5 randomly generated alternatives. Again, the frontier generation process ensured that the alternatives formed a concave efficient frontier with attribute values restricted to the domain of [100, 200] by following the generation algorithm described in Section 4.1.

We tested the sample allocation rules using all 500 testing cases and 39 decision weight triplets in the value function; $(\lambda_1, \lambda_2, \lambda_3)$ = (0.1, 0.1, 0.8), (0.1, 0.2, 0.7), …, (0.8, 0.1, 0.1), (0.05, 0.05, 0.9), (0.05, 0.9, 0.05), (0.9, 0.05, 0.05). To each of the 500 testing cases, we assigned a triplet of measurement variability values, (σ_1^2, σ_2^2, and σ_3^2), to be associated with the three attributes, X_1, X_2, and X_3. The assigned σ_i values (i = 1, 2, 3) were independent, random draws from a uniform distribution with parameters min = 1 and max = 30. Then, for each of the 55 possible sample allocations from an overall budget $N = 9$, for each of the 39 decision weight triplets, across the 500 testing cases, we evaluated the performance of the sample allocation using the process described in Section 4.2 and obtained a *rel fcs* value. For each testing case and decision weight combination, we used each of the sample allocation rules to produce a sample allocation. From the evaluations of the 55 possible sample allocations and 39 decision weight triplets, the *rel fcs* values for the allocations resulting from the sample allocation rules were identified. The performance of a rule, for each decision weight, was defined to be the average *rel fcs* of its sample allocation across the 500 test cases. The uncertainties in the average *rel fcs* were expressed as 95 % confidence intervals based upon the normality assumption as justified by the Central Limit Theorem.

5 RESULTS

The *uniform* and *proportional* allocation rules provided larger average *rel fcs* values than an arbitrary (random) allocation of samples over the range of decision weights studied. When $\lambda_1 = \lambda_2 = \lambda_3$ the proportional allocation rule and the uniform allocation rule provide the same sample allocation ($n_1 = n_2 = n_3 = N/3$) and thus the rules displayed similar performance near these decision weight values. Otherwise, the proportional allocation rule provided *rel fcs* values that exceeded those provided by the uniform allocation rule. This underscores the importance of the sample allocation decision when embarking upon a data collection exercise to support a selection decision. Figure 3 illustrates these general conclusions by displaying, for each of the four allocation rules studied and the random allocation (provided as a reference), the relative frequency of correct selection averaged across all test cases and the 95 % confidence interval at each decision weight value.

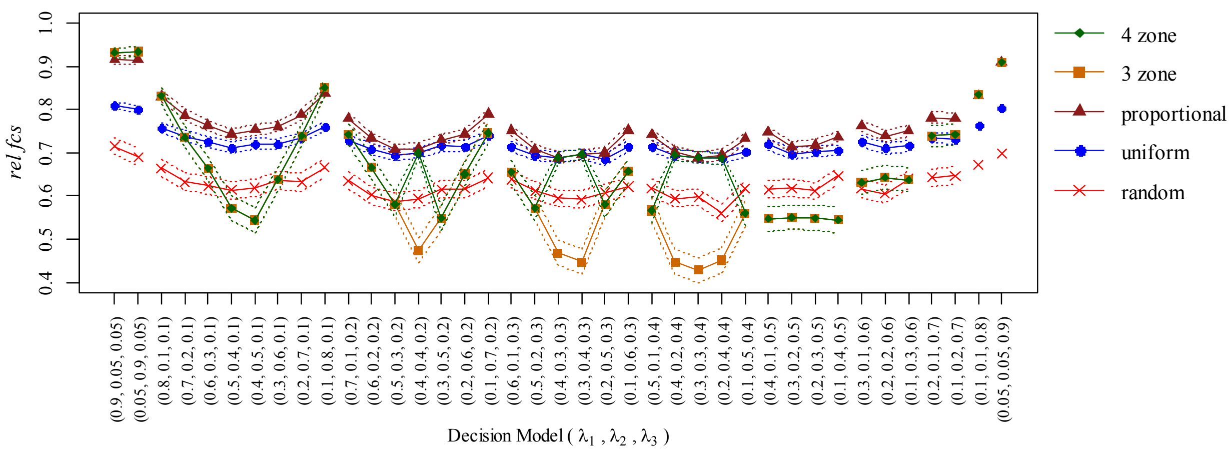

Figure 3: Relative frequency of correct selection for each allocation rule averaged across all testing cases for each decision weight value. The dotted lines represent the 95 % confidence intervals.

The three- and four-zone allocation rules, which leverage extreme sample allocations, provided the largest *rel fcs* value as λ_i approaches 1 for any $i = 1, 2, 3$. However, as the λ_i move away from 1 and approach equality at $\frac{1}{3}$, the performance of the three- and four-zone allocation rules rapidly decreases. With few exceptions, when $0.3 \leq \lambda_i \leq 0.6$ for any $i = 1, 2, 3$, the average *rel fcs* values provided by the three- and four-zone allocation rules are either lower than or indistinguishable from the average *rel fcs* values provided the random allocation. Only when $0.2 \leq \lambda_i \leq 0.4$ for all $i = 1, 2, 3$ does the performance of the four-zone allocation rule exceed that of the three-zone allocation rule. It is within this range of λ_i that the four-zone allocation utilizes the uniform allocation.

6 SUMMARY AND CONCLUSIONS

The ultimate goal of this research is to provide guidance on allocating a fixed budget (for measurements or experiments) across multiple attributes – when collecting data to support a selection decision – to maximize the probability that the decision-maker will choose the true best alternative. Through a simulation study, we have demonstrated that the allocation of samples across the multiple attributes does indeed impact the ability of the decision-maker to choose the true best alternative when the estimated attribute values are subject to normally distributed measurement error. As shown by the contour plots in Figure 1, for a given set of decision weights the relative frequency of correct selection can vary considerably based on the sample allocation. We have shown that a sample allocation based upon the decision model weights (proportional allocation rule) improves the probability of selecting the true best alternative over a sample allocation that does not consider this information (the uniform allocation rule). This emphasizes the importance for projects focused on a selection decision to be managed so that the decision modeling and the experimental (or measurement) planning are done jointly rather than in isolation (which, unfortunately, is currently not uncommon). Such a cooperative approach can improve the overall selection results of the project.

For the three attribute case where the decision alternatives form a concave efficient frontier and the attribute value estimates are subject to normally distributed measurement error, we evaluated four sample allocation rules: uniform allocation, proportional allocation, three-zone allocation, and four-zone allocation. When the experiment or measurements are planned without any knowledge of the decision model or the alternatives' attribute values, then the uniform allocation rule would be a reasonable approach for allocating the budget. We have displayed, however, that this allocation rule nearly always provides an allocation that is sub-optimal. By simply defining the decision model prior to the data collection phase, the proportional allocation rule can be utilized, providing sample allocations that improve the probability of correct selection over those provided by the naïve uniform allocation rule.

The three-zone and four-zone allocation rules implement "extreme allocations" that perform very well for some decision models but very poorly for others. The four-zone allocation rule dominated the three-zone allocation rule, providing more favorable results when all of the weights in the decision model were near $1/3$. Moreover, a hybrid allocation rule that suggests "extreme allocations" when the weight for one attribute is very high (near 1) but proportional allocations when all of the attribute weights are moderate may prove to be valuable.

We expect that these results will hold in cases with more than five alternatives and decision situations with more than three attributes. Nonlinear individual value functions may alter the influence of attribute value uncertainty, however, which could influence the impact of the sample allocation. In situations with a non-additive value function, the trends described here may not hold.

Broadening our understanding of how the frontier characteristics impact the ideal sample allocation and incorporating these findings into an allocation rule is part of our ongoing work on this sample allocation problem for the three attribute selection problem. We will focus efforts on developing a better understanding of the impact the measurement uncertainty of each attribute plays on the optimal allocation. And finally, while our work to this point has focused on single-phased experiments with equal allocations across alternatives, our future work will consider sequential allocation policies and allow for varying allocations across alternatives.

REFERENCES

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

Box, G. E., J. S. Hunter, and W. G. Hunter. 2005. *Statistics for Experimenters*. 2nd ed. Hoboken, NJ: John Wiley & Sons Inc.

Branke, J., S. E. Chick, and C. Schmidt. 2007. "Selecting a Selection Procedure." *Management Science* 53(12):1916-1932.

Butler, J., D. J. Morrice, and P. W. Mullarkey. 2001. "A Multiple Attribute Utility Theory Approach to Ranking and Selection." *Management Science* 47(6):800-816.

Chaloner, K., and I. Verdinelli. 1995. "Bayesian Experimental Design: A Review." *Statistical Science* 10(3):273-304.

Chen, C.-H., D. He, M. Fu, and L. H. Lee. 2008. "Efficient Simulation Budget Allocation for Selecting an Optimal Subset." *INFORMS Journal on Computing* 20(4):579-595.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. *Bayesian Data Analysis*. 2nd ed. New York: Chapman & Hall/CRC.

Keeney, R. L., and H. Raiffa. 1993. *Decisions with Multiple Objectives - Preferences and Value Tradeoffs*. 2nd ed. New York: Cambridge University Press.

Kim, S.-H. and B. L. Nelson. 2006. "Selecting the Best System." In *Handbooks in Operations Research and Management Science, vol. 13: Simulation*, edited by S. G. Henderson and B. L. Nelson, 501-534. Oxford: Elsevier.

Laporte, G. J., J. Branke, and C.-H. Chen. 2012. "Optimal Computing Budget Allocation for Small Computing Budgets." In *Proceedings of the 2012 Winter Simulation Conference*, edited by C.

Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 3324-3336. Piscataway, New Jersey: Institute for Electrical and Electronics Engineers.

Leber, D. D., and J. W. Herrmann. 2012. "Incorporating Attribute Value Uncertainty into Decision Analysis." In *Proceedings of the 2012 Industrial and Systems Engineering Research Conference*, edited by G. Lim and J. W. Herrmann, 1253-1262. Orlando, FL: Institute of Industrial Engineers, Inc.

Leber, D. D., and J. W. Herrmann. 2013. "Allocating Attribute-Specific Information-Gathering Resources to Improve Selection Decisions." In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 101-112. Piscataway, New Jersey: Institute for Electrical and Electronics Engineers.

Leber, D. D., and J. W. Herrmann. 2014. "Resource Allocation for Selection Decisions with Measurement Uncertainty." In *Proceedings of the 2014 Industrial and Systems Engineering Research Conference*, edited by Y. Guan and H. Liao, Montreal, Canada: Institute of Industrial Engineers, Inc.

Lee, L. H., E. P. Chew, S. Y. Teng, and D. Goldsman. 2004. "Optimal Computing Budget Allocation for Multi-objective Simulation models." In *Proceedings of the 2004 Winter Simulation Conference*, edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 586-594. Piscataway, New Jersey: Institute for Electrical and Electronics Engineers.

Lee, L. H., E. P. Chew, S. Y. Teng, and D. Goldsman. 2010. "Finding the Non-dominated Pareto Set for Multi-objective Simulation Models," *IIE Transactions* 42(9):656-674.

Montgomery, D. C. 2013. *Design and Analysis of Experiments*. 8th edition. New York: John Wiley & Sons.

Roy, B. 2005. "Paradigms and Challenges." In *Multiple Criteria Decision Analysis State of the Art Surveys*, edited by J. Figueira, S. Greco and M. Ehrgott, 3-24. New York: Springer.

AUTHOR BIOGRAPHIES

DENNIS D. LEBER is a statistician in the Statistical Engineering Division at the National Institute of Standards and Technology. He is also a Ph.D. candidate in the Department of Mechanical Engineering at the University of Maryland where his research includes incorporating measurement uncertainty into decision analysis. He holds a M.S. degree in Statistics from Rutgers University, a M.S. degree in Mechanical Engineering from the University of Maryland, and is a member of INFORMS and IIE. His email address is dennis.leber@nist.gov and his web page is http://www.nist.gov/itl/sed/gsg/leber.cfm.

JEFFREY W. HERRMANN is an associate professor at the University of Maryland, where he holds a joint appointment with the Department of Mechanical Engineering and the Institute for Systems Research. He is the Associate Director of the University of Maryland Quality Enhancement Systems and Teams (QUEST) Honors Fellows Program. He is a member of IIE, INFORMS, ASME, and ASEE. In 2012 he and Gino Lim were the conference chairs for the Industrial and Systems Engineering Research Conference. His email address is jwh2@umd.edu and his web page is http://www.isr.umd.edu/~jwh2/jwh2.html.