# ALLOCATING ATTRIBUTE-SPECIFIC INFORMATION-GATHERING RESOURCES TO IMPROVE SELECTION DECISIONS

Dennis D. Leber

Jeffrey W. Herrmann

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

A. James Clark School of Engineering
University of Maryland
College Park, MD 20742, USA

## ABSTRACT

When collecting data to select an alternative from a finite set of alternatives that are described by multiple attributes, one must allocate effort to activities that provide information about the value of each attribute. This is a particularly relevant problem when the attribute values are estimated using experimental data. This paper discusses the problem of allocating an experimental budget amongst two attributes when the non-dominated decision alternatives form a concave efficient frontier. The results of a simulation study suggested allocation rules that take advantage of knowledge of the decision model and, when available, knowledge about the general shape of the frontier. These rules were compared to a default rule that equally allocated the experimental budget across the attributes. A proportional rule that allocated samples based on the value function weights performed well only in some cases; a more sophisticated step rule increased the frequency of correct selection across all weights.

## 1 INTRODUCTION

In 2008 the United States Congress mandated that the Domestic Nuclear Detection Office (DNDO) of the U.S. Department of Homeland Security 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, 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 decision model and developed a laboratory experimental plan to support the estimation of the decision model attribute values. This led to the following question: how should the limited laboratory experimental budget be allocated to best support the decision process? This question, which is not limited to the selection of a radiation detection system, applies to all decision processes where the values of multiple attributes are estimated based upon experimental evaluations.

When attribute values are estimated using experimental measurement data, measurement uncertainty associated with the estimates is present and relevant to the decision model. In particular, this attribute value uncertainty can limit the decision model's ability to identify the alternative that truly maximizes the decision-maker's value (or utility). Since the decision-maker can reduce the amount of attribute value uncertainty associated with each attribute by increasing the amount of information used in its assessment, the allocation of experimental effort across the decision attributes plays an important role in maximizing the probability of selecting the truly best alternative.

Leber and Herrmann (2012) introduced the problem of decision-making with attribute value uncertainty and described a method to propagate this uncertainty and estimate, for each alternative, the likelihood that it is truly the best one. This paper discusses the results of a simulation study to investigate the impact of sample allocation in a two-attribute selection decision problem. We present details of the study along with the findings and resulting methods for sample allocation.

## 2    EXPERIMENTAL DESIGN FOR DECISION MAKING

An entire discipline is devoted to the importance of obtaining the best set of observations in an experimental setting. 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. Classic references such as Box, Hunter and Hunter (2005) and Montgomery (2013) provide extensive guidance for the principles and numerous example applications of the methods of statistical design of experiments. Problems in this domain span the realm of comparing entities, quantifying the impact of various experimental factors, and estimating functional relationships. These problems can be represented by $y = f(l_1, \ldots, l_p)$, where $y$ is the response variable of interest, there are $p$ experimental factors that each have multiple levels, and $l_i$ is the level of the $i$-th experimental factor. A main focus of the design of experiments discipline is how to best allocate the total experimental budget of $N$ observations across the design space defined by the factors and their levels while adhering to the underlying principles which, for example, minimize estimation variability and maximize hypothesis testing power. In this regard, the designer must choose which particular combinations of factors and levels will be included in the experiment. The response variable can be either a single response or multiple responses, with all responses measured over each of the identified design points. While the principles are much the same, an alternative to the traditional design of experiments approach is that of Bayesian experimental design (Chaloner and Verdinelli 1995). In Bayesian design, information available prior to experimentation is leveraged in identifying optimal locations to sample within the design space and determining the appropriate sample size.

Finding the best alternative when performance is a random variable (and an alternative's true performance must be estimated using simulation) is known as the selection problem. A simulation run can generate a value that can be used to estimate $y_j = f(A_j)$, where $y_j$ is the value of the response variable (performance) for $A_j$, an individual alternative within the given set of alternatives. When the total number of simulation runs is limited, the problem is to determine how many simulation 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) approaches have been developed to find good solutions (see Branke, Chick and Schmidt (2007) and the references therein). In these sequential approaches, the problem is to determine which alternatives should be simulated next and when to stop. The 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.

As described in the next section we also address a selection problem, but our work is concerned with gathering information about different attributes, not 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 $y_j = f(A_j) = v(x_{j1}, \ldots, x_{jk})$. The alternative's true attribute values $x_{j1}, \ldots, x_{jk}$ are unknown and are estimated based on different information gathering tasks (e.g., experiments). The challenge is to determine how many experiments should be allocated to the evaluation of each attribute value. The simulation study that we describe in this current paper was 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.

## 3  PROBLEM STATEMENT

As classified by Roy (2005), the decision problem we consider is one of *choice*: given non-dominated alternatives $A_1, \ldots, A_m$, the decision-maker will select a single alternative. Each alternative $A_j$ is described by attributes $X_1, \ldots, X_k$, which are quantified by specific attribute values $x_{j1}, \ldots, x_{jk}$, and by its overall value, which equals $v\left(x_{j1}, \ldots, x_{jk}\right)$. The decision-maker will select the alternative that provides the greatest overall value. We assume that the corresponding tradeoffs condition is satisfied, and hence an additive value function of the form displayed in Equation (1) is valid to model the decision (Keeney and Raiffa 1976). Let $x_i$ be the value of attribute $X_i$, let $\lambda_i$ 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,

$$v\left(x_1, \ldots, x_k\right) = \lambda_1 v_1\left(x_1\right) + \cdots + \lambda_k v_k\left(x_k\right) \tag{1}$$

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 $v\left(x_i^*\right) = 1$ for the most desirable attribute value $x_i^*$; $i = 1, \ldots, k$. The attribute weights $\lambda_i$ reflect the decision-maker's preferences and satisfy the constraint $\sum_{i=1}^{k} \lambda_i = 1$.

The best alternative is the one that has the greatest overall value, which is a function of its true attribute values. There are no other influences (e.g., uncertain events) relevant to the decision that must be resolved. 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. In this setting, an "experiment" is an information-gathering activity that provides a value for one attribute of one alternative. Due to randomness in the experiment, the value returned is a random variable that depends upon the true value of the attribute for that alternative. The uncertainty associated with the attribute is a function of the values that are collected from experimentation. (More experiments gather more information about an attribute and will reduce the uncertainty about the estimate for the true attribute value.) After the information is gathered, the results of the experiments are used to model the uncertainty about the estimated attribute values. We assume that the experimentation will occur in a single phase. (We will be considering sequential allocation policies in future work.)

We assume that the decision-maker is concerned about the true rank of the alternative selected and will select the alternative that is most likely to be the true best; thus, he will use the *Rank 1* decision rule described by Leber and Herrmann (2012). (Further details of this approach are provided in Section 4.2.2.) The likelihood that the decision-maker selects the best alternative depends not only upon the decision rule but also upon the experiments selected to gather information about the alternatives and their attributes.

Clearly, if the experimental budget 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 experiments are expensive, this is not possible.

The information-gathering resource allocation problem can be stated as follows: The overall experimental budget will be divided equally among the $m$ alternatives. The experimental budget for each alternative must be divided among the $k$ attributes. In general, the budgets for different alternatives could be divided differently, but we assumed that the allocation is the same for all alternatives. These assumptions are reasonable for a situation in which there is no prior information about the alternatives that would suggest treating them differently. For a given single alternative, let $c_i$ denote the cost of an experiment that generates a single value of attribute $i$, and let $n_i$ denote the number of times that this experiment will be repeated. Let $C$ denote the total budget for one alternative. Thus, 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), such that $c_1 n_1 + \cdots + c_k n_k \leq C$.

## 4    SIMULATION STUDY

Intuitively, it would seem reasonable to perform more experiments on those attributes that have the most uncertainty and are the most important in the decision model. To test this intuition and understand better the behavior of the probability of correct selection as a function of sample allocation across the attributes, we conducted a simulation study. The following subsections describe the details of the study, the insights gained, and the sample allocation rules suggested.

We considered the following simpler situation in which each alternative is described by two attributes, $X_1$ and $X_2$, and the two attributes are the alternative's probability of success in two different tasks. The alternatives, when characterized by their true values of $X_1$ and $X_2$, form a concave efficient frontier in $R^2$ space. An experiment tests one alternative on one task, and the alternative either succeeds or fails. The experiments cost the same amount. Given an overall experimental budget of $N = 10$ Bernoulli trials for each alternative, the problem is to determine the number of trials to be allocated to attribute 1, $n_1$, and the number of trials to be allocated to attribute 2, $n_2$, (where $n_1 + n_2 = N$) 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 defined by Equation (1).

### 4.1    Non-dominated Decision Alternatives

While the efficient frontier formed by a set of non-dominated alternatives described by two attributes can take on several forms (Keeney and Raiffa 1976), we focused specifically on the case of a concave efficient frontier. To form the test bed for our initial experimentation, we developed 20 concave efficient frontiers (cases) where each case included five alternatives characterized by two attributes. The attribute values associated with each alternative were randomly generated subject to the constraints necessary for non-dominance and concavity. The left panel of Figure 1 displays the attribute values of the 20 cases considered in our initial experimentation.
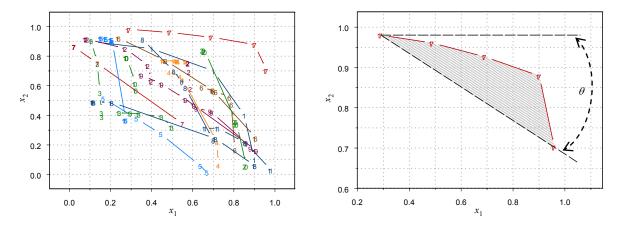


Figure 1: Twenty cases considered in initial experimentation (left panel) with the case number displayed by the numeric plotting character. The right panel illustrates the measures of nonlinearity (scaled shaded area) and general angle ($\theta$).

Given the random nature of the 20 cases, many aspects of potential alternative sets are captured such as frontiers with great curvature, nearly linear frontiers, both horizontally and vertically situated frontiers, frontiers that span a small region, and those that span a larger region. Two characteristics were quantified and used to describe each case: a measure of nonlinearity and a measure of general angle. The measure of

nonlinearity is defined as a scaled area between the piecewise linear concave curve formed by the alternatives on the concave frontier and the line segment connecting the alternatives defined by the attribute pairs $(x_{1[1]}, x_{2[5]})$ and $(x_{1[5]}, x_{2[1]})$, where the bracketed subscript denotes the $i^{\text{th}}$ ordered statistic, $i = 1, \ldots, 5$. The measure of general angle is defined as the acute angle formed by the line segment connecting the attribute pairs $(x_{1[1]}, x_{2[5]})$ and $(x_{1[5]}, x_{2[1]})$ and the horizontal line $x_2 = x_{2[5]}$. The measures of general angle, $\theta$, and nonlinearity, $NL$, are related by $\max NL(\theta) = \frac{1}{4}\sin\left(\frac{\theta\pi}{90}\right)$. An illustration of the nonlinearity and general angle measures is given in the right panel of Figure 1.

## 4.2 Decision Simulation

To measure the performance of a particular allocation of experimental effort, we tallied how often the decision rule selected the alternative that is truly the best for that case, given the value function weights $\lambda_1$ and $\lambda_2 = 1 - \lambda_1$. This measure we called the *frequency of correct selection*. The following subsections provide details on the generation of the experiment realizations and the implementation of the subsequent decision rule that provide frequency of correct selection values.

### 4.2.1 Experiment Realizations

For each of the $j = 1, \ldots, 5$ alternatives within each of the $i = 1, \ldots, 20$ cases we simulated 1000 pairs of binomial experiments that represented experimental results from the pair of attributes, $X_1$ and $X_2$. Each pair of binomial experiments consisted of a total of $N = 10$ Bernoulli trials that were divided between attribute 1 and attribute 2 as $n_1$ and $n_2 = N - n_1$. We considered all 11 possible sample allocations defined as $(n_1, n_2)$ pairs: (0, 10), (1, 9), …, (10, 0). The Bernoulli probability of success associated with each binomial experiment was provided as the true attribute value for the alternative as specified by the initial test bed. The result obtained from each of the simulated binomial experiments is denoted as $y_{ijkrs}$, the number of successes for attribute $k$ from alternative $j$ in test bed case $i$ under sample allocation $s = n_1 + 1$ in simulation replicate $r = 1, \ldots, 1000$.

### 4.2.2 Decision Rule

We considered $d = 1, \ldots, 19$ decision models of the form displayed in Equation (1), each with a unique pair of weighting parameters $\lambda_1$ and $\lambda_2 = 1 - \lambda_1$, where $(\lambda_1, \lambda_2) = (0.05, 0.95), (0.10, 0.90), \ldots,$ (0.95, 0.05). The individual value functions $v_1(x_1)$ and $v_2(x_2)$ were both defined to be linear in all nineteen models. Provided the binomial outcomes $y_{ijkrs}$ from each simulated experiment, the decision method described in Leber and Herrmann (2012) was implemented and an alternative was selected for each case, sample allocation, and decision model combination. Specifically, the uncertain attribute values were modeled with a uniform prior distribution and updated based upon the observed binomial outcomes to provide the Bayesian posterior distribution $Beta\left(1 + y_{ijkrs}, 1 + n_{ikrs} - y_{ijkrs}\right)$. 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 two attributes and calculating the value of the alternative for each sample using the value function. This resulted in a distribution of 1000 values for each of the five alternatives within a case.

For each of the 1000 samples, the *Rank 1* selection rule noted the alternative with the best (largest) value, and the alternative that had the best value most often was selected for this simulation replicate. The *Rank 1*-selected alternative was identified for each case, sample allocation, and decision model for each of the 1000 simulated binomial experiments and compared to the true best alternative for the decision model (the alternative known to maximize the value function given its attribute values provided by the test bed). The frequency of correct selection was then defined as the percentage of *Rank 1* selected alternatives that were truly the best alternative out of the 1000 simulated binomial experiments.
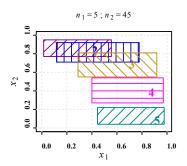
### 4.2.3 Pseudo Code and Performance

1. For each case $i = 1, \ldots, 20$, perform steps 2 to 7.
2. For each decision model $d = 1, \ldots, 19$,
   (a) Determine $j^*$, the alternative that maximizes $v_d(x_{ij1}, x_{ij2}) = \lambda_{1d}x_{ij1} + \lambda_{2d}x_{ij2}$; and
   (b) Perform steps 3 to 7.
3. For each allocation $s = 1, \ldots, 11$,
   (a) Set $n_1 = s - 1$ and $n_2 = N - n_1$;
   (b) Set $C_{dijs} = 0$ for alternatives $j = 1, \ldots, 5$; and
   (c) Perform steps 4 to 7.
4. For each simulation replication $r = 1, \ldots, 1000$,
   (a) Set $M_{dijrs} = 0$ for alternative $j = 1, \ldots, 5$;
   (b) For each alternative $j = 1, \ldots, 5$, select a sample value for $y_{ij1rs}$ from the *Binomial*$(n_1, x_{ij1})$ distribution and a sample value for $y_{ij2rs}$ from the *Binomial*$(n_2, x_{ij2})$ distribution; and
   (c) Perform steps 5 and 6.
5. For each Monte Carlo sample $t = 1, \ldots, 1000$,
   (a) For each alternative $j = 1, \ldots, 5$, select a sample value for $z_{ij1rst}$ from the posterior distribution *Beta*$(1 + y_{ij1rs}, 1 + n_1 - y_{ij1rs})$ and a sample value $z_{ij2rst}$ from the posterior distribution *Beta*$(1 + y_{ij2rs}, 1 + n_1 - y_{ij2rs})$ and calculate $v_{jt} = \lambda_{1d}z_{ij1rst} + \lambda_{2d}z_{ij2rst}$.
   (b) Identify $j^{**}$ as the alternative that maximizes $v_{jt}$ and add one to $M_{dij^{**}rs}$.
6. Identify $j^{***}$ as the alternative that maximizes $M_{dij^{**}rs}$ (this is the alternative selected in replication $r$) and add one to $C_{dij^{***}s}$.
7. Let $F_{dis} = C_{dij^*s}$. (This is the frequency of selecting the correct alternative for this case, decision model, and sample allocation)
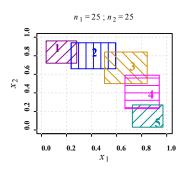
The simulation code was implemented in the freely available statistical computing software R, 64-bit version 2.15.3 and run on a laptop PC with a 2.80 GHz processor and 8.00 GB of memory. The user time (the CPU time charged for the execution of user instructions of the calling process) of approximately 22 hours was required to complete this simulation consisting of 20 five-alternative frontier cases, 19 decision models, 11 sample allocations, 1000 simulation replications, and 1000 Monte Carlo replications.

### 4.3 The Impact of Sample Allocation

Before exploring the results of the decision simulation, we gained some insight about the impact of sample allocation in this two-attribute problem when utilizing the decision method of Leber and Herrmann (2012). Using the set of five alternatives with attribute value pairs {(0.1, 0.9), (0.4, 0.85), (0.7, 0.7), (0.85, 0.4), (0.9, 0.1)} we investigated the impact of sample allocation on the expected Bayesian posterior distributions.

As with our general problem, we allocated $N$ samples (in this case 50) across the two probability of success attributes as $n_1$ and $n_2 = N - n_1$, with the allocation the same for all alternatives. For each alternative $j = 1, \ldots, 5$ and attribute $k = 1, 2$, we defined the expected Bayesian posterior distribution to be the posterior distribution obtained when the proportion of successes observed in the binomial experiment of size $n_k$ equals $x_{jk}$, the true attribute value, i.e., $Beta\left(1 + n_k x_{jk}, 1 + n_k\left(1 - x_{jk}\right)\right)$. The 95 % expected credible interval was calculated for each alternative and attribute as the interval from the lower 2.5 % quantile to the upper 97.5 % quantile of the expected Bayesian posterior distribution. We defined a 95 % expected credible box to be the rectangle whose edges are equal to the end points of the 95 % expected credible intervals for each attribute in $R^2$ space. The probability that the alternative's true attribute values are believed to be contained within this credible box equals $(0.95)^2$. Figure 2 displays graphically the expected credible boxes for each of the five alternatives for three different sample allocations.
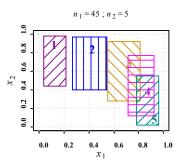
Figure 2: Expected credible boxes for three different sample allocations of 50 total samples (when the proportion of successes observed equals the true attribute value). The plotting character (1 to 5) indicates the true attribute values for each alternative.

When moving from the left panel (sample allocation $n_1 = 5$, $n_2 = 45$) to the center panel ($n_1 = 25$, $n_2 = 25$) and to the right panel ($n_1 = 45$, $n_2 = 5$) of Figure 2, it is seen that the ability to discern alternatives improves when considering attribute 1 (width of boxes decrease) and declines when considering attribute 2 (height of boxes increase). This suggests that allocations placing more samples with attribute 2, such as in the left panel, would provide the ability to better identify the true best alternative for a decision model with a large emphasis on attribute 2 (i.e., large $\lambda_2$ value). Whereby the opposite is suggested for allocations placing more samples with attribute 1 (better suited for decision models with large $\lambda_1$ value).

## 4.4 Frequency of Correct Selection Analysis

For each of the 20 cases and each of the 19 decision models (defined by $\lambda_1$, $\lambda_2$ pairs) considered in the decision simulation, there is at least one sample allocation value of $n_1$, denoted $n_1^*$, that produced the maximum frequency of correct selection (in some cases, for some values of $\lambda_1$, there were multiple values of $n_1^*$). This optimal sample allocation should maximize the probability of choosing the true best alternative. Since $\lambda_1 = 1 - \lambda_2$ and $n_1 = N - n_2$, it was sufficient to consider only $\lambda_1$ and $n_1$ when exploring the frequency of correct selection results. The relationship between $n_1^*$ and $\lambda_1$ followed a general trend in which $n_1^*$ increased as $\lambda_1$ increased, often in a manner that could be well represented by an "*S*-curve" such as the logistic function. The shape and location of the "*S*-curves" varied but displayed dependencies on $\theta$, the frontier characteristic measure of overall angle.

For each of the 11 sample allocations for each case and decision model, we defined the relative frequency of correct selection as the ratio of the frequency of correct selection for that sample allocation to the frequency of correct selection that results from the optimal sample allocation. This measure provided a continuous response variable, as a function of $n_1$ and $\lambda_1$, that indicates the relative quality of each sample allocation versus the optimal sample allocation. The relative frequency of correct selection measure allows us, within the confines of the problem which include the alternatives' attribute values and the total experimental budget, to quantify how much better the selection could be if a different sample allocation was chosen. By viewing the relationship between $n_1$, $\lambda_1$, and the relative frequency of correct selection as a contour plot for each case, further insights were gained.

The shaded contour plots of Figure 3 present the relative frequency of correct selection as a function of $n_1$ and $\lambda_1$ ranging from dark (low values) to light (high, desirable values). The solid squares within the plots denote $n_1^*$, the sample allocation that produced the maximum frequency of correct selection for each $\lambda_1$ considered in the simulation study. The contour plots of Figure 3 serve to illustrate three general trends observed across the 20 test bed cases.
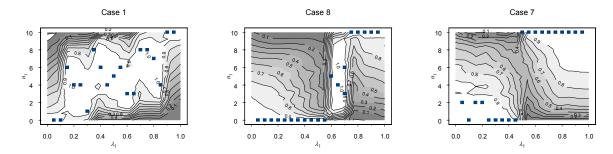
Figure 3: Contour plots for the relative frequency of correct selection as a function of $n_1$ and $\lambda_1$ for three selected test bed cases. The solid squares denote $n_1^*$ for the $\lambda_1$ values evaluated.

The left panel in Figure 3 displays the contour plot as a result of case 1. The general increasing trend of $n_1^*$ as a function of $\lambda_1$ is observed, but further, it is observed that favorable values of relative frequency of correct selection can be obtained through much of the middle of the graph; in other words, choosing a sample allocation that performed poorly would be difficult for nearly all values of $\lambda_1$. The center panel in Figure 3 displays the contour plot as a result of case 8. Again the general increasing trend of $n_1^*$ as a function of $\lambda_1$ is observed, but here the optimal "path" from $\lambda_1 = 0$ to $\lambda_1 = 1$ is much more defined with less room for error, and thus more careful consideration is necessary when choosing a sample allocation. Finally, the rightmost panel in Figure 3 is the contour plot as a result of case 7. Here, favorable relative frequency of correct selection values at low values of $\lambda_1$ correspond to low values of $n_1$ (lower left region of graphic) and at high values of $\lambda_1$ favorable relative frequency of correct selection values correspond to high values of $n_1$ (upper right region of graphic). But unlike the first two cases, these two regions are disconnected.

## 4.5 Sample Allocation Rules

In general, the optimal allocation rule has dependencies on the degree of information possessed by the decision-maker. In the absence of any information, including knowledge of the decision model, the decision-maker will have no reason to allocate more samples to either attribute. Thus, a balanced allocation of $n_1 = n_2 = \frac{1}{2}N$ would be implemented as the experimental plan to collect data to support the estimation of each attribute value and ultimately the selection decision. 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 knowledge of the decision model – specifically the values of $\lambda_1$ and $\lambda_2$ – is available, the decision-maker may choose to implement an allocation rule that assigns $n_1$ and $n_2$ proportional to $\lambda_1$ and $\lambda_2$. Such an approach is supported by the general insights discussed in Section 4.3 as well as the observations made is Section 4.4 that the optimal allocation $n_1^*$ increased as $\lambda_1$ increased. Since $n_1$ and $n_2$ must be integer values, rounding is necessary, e.g., $n_1 = round(\lambda_1 N)$, $n_2 = N - n_1$. We refer to this sample allocation approach as the *proportional allocation rule*.

If, in addition to knowledge of the decision model, the decision-maker is able to make some general statements about the shape of the concave frontier, such as the value of the general overall angle $\theta$, a case-specific allocation rule may be utilized. It was observed in Section 4.4 that as a function of $\lambda_1$, the optimal sample allocation $n_1^*$ was reasonably represented by an "*S*-curve" with location and shape dependent on the general overall angle of the concave frontier. One simplification of an "*S*-curve" is a step-like function with the horizontal steps connected by a line which may or may not be vertical. We considered a sample allocation *step rule* with $n_1$ defined by Equation (2) and $n_2 = N - n_1$.

$$n_1 = \begin{cases} 0 & \text{if } \lambda_1 < c_1 \\ round\left(\dfrac{\lambda_1 - c_1}{c_2 - c_1} N\right) & \text{if } c_1 \leq \lambda_1 \leq c_2 \\ N & \text{if } \lambda_1 > c_2 \end{cases} \qquad (2)$$

For each of our 20 test bed cases, we determined the values of $c_1$ and $c_2$ in Equation (2) that maximized the average relative frequency of correct selection across the 19 evaluated values of $\lambda_1$. We selected the pair of second order polynomial models $c_1(\theta) = \frac{1}{225}\theta + \frac{1}{13500}\theta^2$ and $c_2(\theta) = \frac{7}{450}\theta - \frac{1}{20250}\theta^2$ to describe the relationship between these optimal values of $c_1$ and $c_2$ and the general overall angle of the concave frontiers. Note that when $\theta = 0$ (a horizontal frontier), $c_1 = c_2 = 0$ and when $\theta = 90$ (a vertical frontier), $c_1 = c_2 = 1$; this results in the sample allocations ($n_1 = 10$, $n_2 = 0$) and ($n_1 = 0$, $n_2 = 10$), respectively, for any decision model values of $\lambda_1$ and $\lambda_2$.

## 5    ALLOCATION RULE COMPARISON

The second part of our study compared the performance of the allocation rules. One hundred new concave frontiers, each containing 5 alternatives were randomly generated for the comparison of the three sample allocation rules described in Section 4.5 (*uniform allocation, proportional allocation,* and *step rule*). The attribute values of the 100 evaluation cases and their nonlinearity and general angle measure values are displayed in Figure 4.
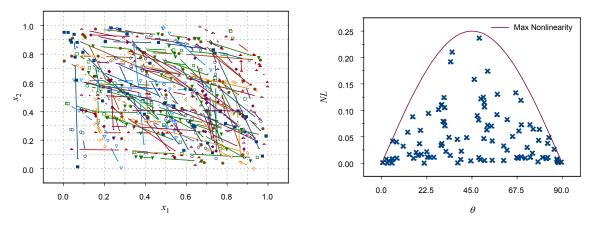


Figure 4: Attribute values (left panel) and nonlinearity and general angle measure values (right panel) for the 100 concave frontiers in the evaluation set.

The sample allocations dictated by each of the three allocation rules were determined for each of the 100 evaluation cases. Based on these sample allocations, the decision simulation described in Section 4.2, again with 1000 simulation replicates and 1000 Monte Carlo samples, was used to evaluate the frequency of correct selection (and relative frequency of correct selection) for each evaluation case and each of the 19 decision models ($\lambda_1$ values). We then examined, on a case-by-case basis, the relative frequency of correct selection that resulted from the simulation, as a function of $\lambda_1$. At each $\lambda_1$ value, we calculated the average relative frequency of correct selection across the 100 evaluation cases. The uncertainties in the average relative frequency of correct selection were expressed as 95 % confidence intervals based upon the normality assumptions provided by the Central Limit Theorem.

While it varied from case to case and across $\lambda_1$ values, in general, the step allocation rule provided the largest relative frequency of correct selection values. The proportional allocation rule displayed similar performance to the step rule with $\lambda_1$ values near 0 and near 1, but presented lower relative frequency of correct selection values elsewhere. The uniform allocation rule nearly always produced the smallest

relative frequency of correct selection values. Further, in many evaluation cases, the sample allocation resulting from the uniform allocation rule failed to provide the best possible allocation (a relative frequency of selection equal to one) for any values of $\lambda_1$. Figure 5 illustrates these general conclusions by displaying, for each of the three allocation rules, the relative frequency of correct selection averaged across all evaluation cases at each $\lambda_1$ value.
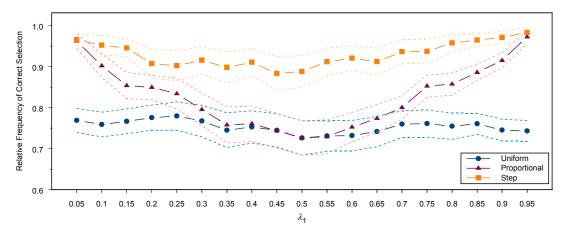


Figure 5: Relative frequency of correct selection averaged across all evaluation cases for each $\lambda_1$ value. The dashed lines represent the 95 % confidence intervals.

It is observed from Figure 5 that sample allocations provided by the uniform allocation rule lead to, on average, a relative frequency of correct selection near 0.75, and this performance is nearly constant over all values of $\lambda_1$. The sample allocations provided by the proportional allocation rule near $\lambda_1 = 0.5$ are identical to the uniform allocation, and hence the performance of the proportional and uniform allocation rules are comparable in this region. The relative frequency of correct selection for the proportional allocation rule increases as $\lambda_1$ moves away from 0.5 and towards either $\lambda_1 = 0$ or $\lambda_1 = 1$. Overall, the average relative frequency of correct selection for the proportional allocation rule is approximately 0.83. The step rule provides a maximum average relative frequency of correct selection value of 0.98 at $\lambda_1 = 0.95$, a minimum average value of 0.88 at $\lambda_1 = 0.5$, and an overall average relative frequency of correct selection of 0.93. We thus arrive at the important conclusion that for nearly all values of $\lambda_1$, the step rule provides average relative frequency of correct selection values that are statistically distinguishable (non-overlapping confidence intervals) and superior to the other allocation rules.

It can be observed from the right panel of Figure 4 that there were a number of evaluation cases that displayed nearly identical frontier curve characteristic measures of nonlinearity and general angle. The contour plots of the relative frequency of correct selection as a function of $n_1$ and $\lambda_1$, and ultimately the relative frequency of correct selection values that resulted from the sample allocations dictated by the three allocation rules were very similar when comparing several such pair of evaluation cases near the limits of the characteristic measures. On the other hand, these similarities were not observed when comparing pairs of cases near the middle of the characteristic measures' domain. This observation leads us to believe that there is at least one additional frontier curve characteristic that would help distinguish relative frequency of correct selection performance from case to case, and ultimately aid in identifying the optimal sample allocation. Additional frontier curve characteristics that we intend to explore and serve as a subject for future research include the distance alternatives are from one another within the frontier, as well as the general proximity of the frontier relative to the limits of the attribute values.

## 6    CONCLUSIONS

We have demonstrated that when collecting data to support a selection decision, the allocation of experimental budget across the attributes does impact the probability of correct selection. This was clearly illustrated by the contour plots in Figure 3 of Section 4.4. Further, we have shown that the decision model as well as the set of alternatives (and their associated true attribute values) also influences the probability that the decision-maker will choose the truly best alternative. While it is not uncommon in large projects for the decision modeling and the experimental planning to be done in isolation, we have shown how the decision modeling efforts can help improve the experimental planning, which in turn can improve the overall selection results of the project.

For the two-attribute case studied where the non-dominated decision alternatives form a concave efficient frontier, we evaluated three novel, previously unstudied sample allocation rules: *uniform allocation*, *proportional allocation*, and the *step rule*. The *uniform allocation* rule would be a reasonable approach for allocating the experimental budget when information available about the decision model and the alternative's attribute values is not available. We have displayed, however, that this allocation rule nearly always provides an allocation that is sub-optimal. Defining the decision model prior to the data collection phase allows one to utilize the *proportional allocation* rule. We have shown, for the decision models considered, that the proportional allocation rule provides sample allocations that outperform the sample allocations provided by the naïve uniform allocation rule, particularly for $\lambda_1 < 0.25$ and $\lambda_1 > 0.75$. If, in addition to the decision model, the decision-maker can provide the general slope of the concave frontier formed by the considered alternatives, then the proposed *step rule* can be utilized. The step rule identified sample allocations that provided, on average, a significant improvement in the relative frequency of correct selection over the allocations provided by the uniform allocation and the proportional allocation rules for nearly all the decision models considered.

While the proposed step rule provided sample allocations with favorable correct selection values for most frontiers and decision models, it did not identify the optimal allocation in all cases. There may be additional frontier characterization measures, beyond the general angle, that could be implemented into the calculation of $c_1$ and $c_2$ in the step rule to more often provide sample allocations that are optimal. The step rule was designed as a simplification of the *S*-curves observed to describe the relationship between the optimal allocation and the decision model parameter $\lambda_1$. Though the relative simplicity of the step rule is attractive, a more complicated model might provide better performance.

This study focused on the attribute-specific information-gathering resource allocation problem when considering two attributes that describe the alternative's probability of success in two different tasks with information collected in the form of Bernoulli trials during a single-phased experiment. Our future work will consider sequential allocation policies with more than two attributes where these attributes may be supported by measurements other than Bernoulli results.

## REFERENCES

Box, G. E. P., S. J. Hunter, and W. G. Hunter. 2005. *Statistics for Experimenters*. Second Edition. 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.

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

Keeney, R. L., and H. Raiffa. 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: John Wiley & Sons, Inc.

LaPorte, G. J., J. Branke, and C. Chen. 2012. "Optimal Computing Budget Allocation for Small Computing Budgets." In *Proceeding of the 2012 Winter Simulation Conference*, Edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 3324-3336. Berlin, Germany: Institute of Electrical and Electronics Engineers, Inc.

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 Industial Engineers, Inc.

Montgomery, D. C. 2013. *Design and Analysis of Experiments*. Eigth 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.