Proceedings of the 2014 Industrial and Systems Engineering Research Conference
Y. Guan and H. Liao, eds.

Resource Allocation for Selection Decisions with Measurement Uncertainty

Dennis D. Leber
National Institute of Standards and Technology
Gaithersburg, MD 20899, U.S.A.

Jeffrey W. Herrmann
University of Maryland
College Park, MD 20742, U.S.A.

Abstract

Before selecting an alternative based on attribute values that have been determined through a measurement process that has error, a decision-maker can, in some cases, collect additional data to reduce uncertainty. Unlike previous work in the area of ranking and selection, this study considered the problem of allocating limited data collection resources across multiple attributes rather than across multiple alternatives. In this work we assumed the multiple attribute values are measurements (samples) of physical characteristics and have a normally distributed measurement error. We conducted a simulation study to investigate how the sample allocation affects the likelihood that the best alternative will be selected and how this relationship is influenced by the relative importance of the attributes and the amount of attribute value uncertainty. The results suggested allocation rules based on the decision model and the general shape of the frontier. These rules were compared to a default rule that allocated the experimental budget evenly across the attributes. Better allocations increase the likelihood that the best alternative will be selected.

Keywords

Ranking and selection, experimental design, simulation, multicriteria decision making, decision analysis

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 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 technology? 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.

When attribute values are estimated using experimental measurement data, intrinsic variability in the data leads to measurement uncertainty associated with the estimates that is relevant to the selection problem. In particular, this attribute value uncertainty can limit the decision-maker's ability to identify and select the system (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 (sample allocation) across the decision attributes plays an important role in maximizing the probability of selecting the truly best alternative.

In the following section we indicate how the sample allocation problem considered in this paper differs from the extensive work done in the field of ranking and selection. Expanding our previous work in this area which focused on pass-fail testing (Bernoulli trials) [1], in this paper we address the sample allocation problem for a two attribute selection decision in which the measurement error for each attribute is normally distributed with unknown variance; 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 Experimental Design

Finding the best alternative, out 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 – is known as the selection problem. 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 [2, 3, 4]). 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 [4] demonstrated the strengths and weaknesses of these procedures. LaPorte, Branke and Chen [5] 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 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 [6] 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 [7] and Montgomery [8] 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 experimental budget of N observations 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 [9] is an alternative approach 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 [10], the decision problem we consider is one of choice: given a set of alternatives, $\left\{A_1,\ldots,A_m\right\}$, 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 (utility), as determined by $y_j = v\left(x_{j1},\ldots,x_{jk}\right)$. The decision-maker will select the alternative that provides the greatest overall

value. We assume that the corresponding tradeoffs condition is satisfied [11], and hence an additive value function of the form displayed in Equation (1) is valid to model the decision. Let x_i be the value of attribute X_i, let λ_i be the weight of attribute X_i, and let $v_i(x_i)$ 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:

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

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

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 observed value is a random variable that 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 experiments gather more information about an attribute and will reduce the uncertainty of the estimate for the true attribute value.) After the information is gathered, the experimental results are used to model the uncertainty of the estimated attribute values, which is then propagated onto the decision-maker's overall value for the alternative.

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 [12].

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. For this work, we assume that the experimental budget is fixed and all experimentation will occur in a single phase. We will be considering sequential allocation policies in future work.

The information-gathering resource allocation problem can be stated as follows: The overall experimental budget, *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* 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 experimental budget. To test this intuition, we conducted a simulation study to understand how the experimental 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 two attributes, X_1 and X_2, that are measured using two different techniques. The error of each measurement technique is normally distributed with unknown variance. The alternatives, when characterized by their true values of X_1 and X_2, form a concave efficient frontier in R^2 space. The individual value functions $v_1(x_1)$ and $v_2(x_2)$ were both defined to be linear.

4.1 Training Cases and Measurement Error

Based on the results of previous work that considered pass-fail testing to estimate attribute values in terms of probabilities [1], we expected that sample allocation rules might provide the decision-maker with guidance. We first generated a set of 20 training cases (sets of alternatives) and used these to guide the values of the parameters in our sample allocation rules. Each training case consisted of five alternatives described by two 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. We measured two characteristics of 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 two extreme alternatives. The measure of general angle is defined as the acute angle formed by the line segment connecting the two extreme alternatives and the horizontal line that goes through the point with the greatest value on attribute two. The measures of general angle, θ (measured in degrees), and nonlinearity, NL, are related by $0 \le NL \le \frac{1}{4}\sin\left(\frac{\theta\pi}{90}\right)$. The 20 training cases provided efficient frontiers that varied with regards to general angle and nonlinearity. Figure 1 displays the true attribute values of the 20 training cases (left panel) and an illustration of the measures of general angle and nonlinearity (right panel).

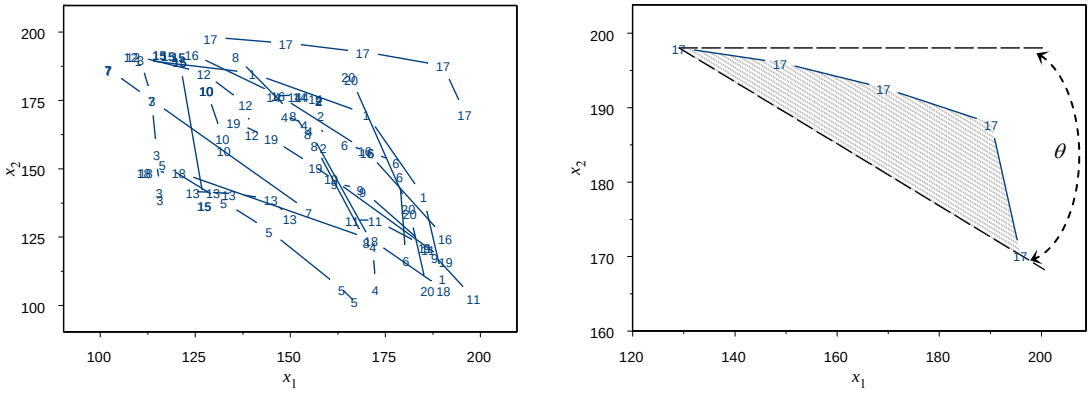

Figure 1: Twenty training cases (left panel) with the case number displayed by the numeric plotting character. The right panel illustrates the measures of nonlinearity (scaled shaded area) and general angle (θ).

For our simulation, it was assumed that each of the two attributes was measured with a different measurement technique and that technique maintained a measurement variability that was consistent across all alternatives measured. We set the actual measurement variance of each attribute (σ_1^2 and σ_2^2) to 1^2, 10^2, 20^2, and 30^2, which created 16 different "measurement error scenarios."

4.2 Evaluating Sample Allocations

An experimental sample generates one (random) measurement of one attribute of one alternative. Given an experimental budget of $N = n_1 + n_2 = 10$ samples for each alternative, the problem is to determine n_1 and n_2, the number of samples of attribute 1 and attribute 2, 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).

We evaluated, using the 20 training cases, all of the possible sample allocations for N = 10 total samples per alternative ($n_1 = 0, 1, \ldots, 10$; $n_2 = N - n_1$) over a range of values of λ_1 and λ_2, the weights in the decision value function. In particular, 19 decision weight pairs $(\lambda_1, \lambda_2) = (0.05, 0.95), (0.10, 0.90), \ldots, (0.95, 0.05)$. To do this, for

each case (20), measurement error scenario (16), and sample allocation (11) – a total of 3520 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 50 measurements, 10 for each of the 5 alternatives, with n_1 measurements observed from attribute 1 and n_2 measurements observed from attribute 2. 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 of the 3520 sets of sample measurements. The uncertain attribute values were modeled, a priori, with a normal distribution with mean of 150 and variance of 35^2. The unknown measurement variability was modeled, a priori, with a scaled inverse-chi-square distribution with parameters 0.01 and 1 (note that this distribution closely resembles that of a non-informative Jeffrey's prior distribution). A Bayesian semi-conjugate prior model for normally distributed data [13] 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 two attributes and calculating the overall decision value of the alternative using each of the 19 decision value functions (as defined by the 19 decision weight pairs). For each decision weight pair, 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. Repeating this selection process over all 1,000 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 11 sample allocations, for each of the 19 decision weights, across the 320 subcases. For each of the 320 subcases and each of the 19 decision weights, there is at least one optimal sample allocation for attribute 1, denoted n_1^*, 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 n_1^* allocation. Within the confines of the problem which include the alternatives' attribute values and the total experimental 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 shaded contour plots of Figure 2 present the *rel fcs* as a function of n_1 and λ_1, ranging from dark (low *rel fcs* values) to light (high, desirable *rel fcs* values) for training case 16. The solid squares within the plots denote n_1^*, the optimal sample allocation for attribute 1 at each λ_1 value. The contour plots of Figure 2 serve to illustrate two general trends observed across the 20 training cases.

First, consider left panel of Figure 2 (subcase 16.30.1) where the measurement variability associated with attribute 1 is large ($\sigma_1^2 = 30^2$) and that associated with attribute 2 is small ($\sigma_2^2 = 1^2$). When attribute 1 is very important to the decision-maker (λ_1 is near 1), then only sample allocations that allocate nearly all 10 samples to attribute 1 (n_1 is near 10) produce favorable *rel fcs* values. When, however, attribute 1 is not important (λ_1 is near 0), nearly all sample allocations produce very favorable results. Although attribute 2 is very important in this situation, because it has low measurement variability, sample allocations with small values of n_2 still produce favorable *rel fcs* values. This phenomenon is seen in reverse in subcase 16.1.30 (right panel, Figure 2). Subcase 16.30.30 (center panel, Figure 2) combines the restrictive observations of the previous two subcases because both attributes have large measurement variability ($\sigma_1^2 = \sigma_2^2 = 30^2$). That is, only sample allocations that allocate nearly all 10 samples to attribute 1 produce favorable *rel fcs* results when attribute 1 is very important, and only sample allocations that allocate nearly all 10 samples to attribute 2 produce favorable *rel fcs* results when attribute 2 is very important. These observations begin to illustrate the complex tradeoff between apportioning samples based on measurement variability and sampling the important attribute.

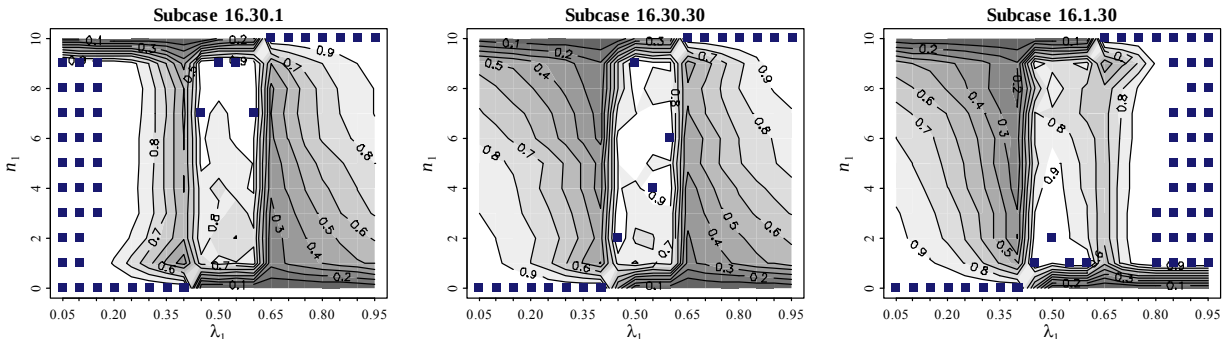

Figure 2: Contour plots displaying *rel fcs* as a function of n_1 and λ_1 for training case 16 under measurement error scenarios (30^2, 1^2), (30^2, 30^2), and (1^2, 30^2). For each λ_1, the solid squares denote n_1^*, the optimal sample allocation for attribute 1.

The second trend illustrated in Figure 2 is that, in general, as the importance of attribute 1 increases (that is, as the weight λ_1 increases from 0 to 1), the n_1 values that generate the most desirable sample allocations increase. This relationship is not strictly linear, but it can be reasonably represented by an "S-curve" such as the logistic function. This can be seen most clearly in subcase 16.30.30 (center panel, Figure 2). Although the location and shape of a representative S-curve varied from case to case, we saw that these parameters depended on the shape of the frontier.

4.3 Creating 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 either attribute and would use a balanced allocation of $n_1 = n_2 = N/2$. 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 values of λ_1 and λ_2 are available, then the decision-maker may choose to assign n_1 and n_2 proportional to λ_1 and λ_2. The observations made from Figure 2 in Section 4.2 showed that the optimal sample allocation generally increased as λ_1 increased. Since n_1 and n_2 must be integer values, rounding is necessary, e.g., $n_1 = \text{round}(\lambda_1 N)$, $n_2 = N - n_1$. We refer to this sample allocation approach as the *proportional allocation rule*.

Although the proportional allocation rule is simple, the evaluations showed that the relationship between λ_1 and n_1^* was usually not linear, but rather, distinctly nonlinear. To approximate this relationship, we used a step-like function that we called the *step rule*. For $0 \le c_1 \le c_2 \le 1$ and $0 < \lambda_1 < 1$, the step rule assigns n_1 and n_2 as follows:

$$n_1 = \begin{cases} 0 & \text{if } \lambda_1 \le c_1 \\ round\left[\left(\dfrac{\lambda_1 - c_1}{c_2 - c_1}\right) N\right] & \text{if } c_1 < c_2 \text{ and } c_1 < \lambda_1 \le c_2 \quad ; \quad n_2 = N - n_1 \\ N & \text{if } \lambda_1 > c_2 \end{cases} \tag{2}$$

We determined, for each training case, the values of the parameters c_1 and c_2 that maximized the average relative frequency of correct selection across all measurement error scenarios. We used those results to generate insights into how these parameter values depend upon the shape of the concave frontier (the values of the general overall angle θ and the nonlinearity measure *NL*). From this, we hypothesized a quadratic relationship between θ and c_1 (and c_2) and found the best fit using restricted least-squares regression:

$$\begin{aligned} c_1 &= f_1(\theta) = 8.47 \times 10^{-4}\theta + 1.14 \times 10^{-4}\theta^2 \\ c_2 &= f_2(\theta) = 2.09 \times 10^{-2}\theta - 1.09 \times 10^{-4}\theta^2 \end{aligned} \tag{3}$$

We also hypothesized a bivariate relationship between θ and *NL* and c_1 (and c_2) and found the best fit using restricted least-squares regression:

$$c_1 = f_1(\theta, NL) = 1.11\times10^{-2}\theta + 0.674NL - 5.94\times10^{-2}\theta\cdot NL$$
$$c_2 = f_2(\theta, NL) = 1.11\times10^{-2}\theta + 4.56NL - 5.69\times10^{-2}\theta\cdot NL \quad (4)$$

(The details of these derivations are omitted due to space constraints.) Thus, if the decision-maker had information about the shape of the concave frontier, a specific step rule could be generated and used to determine the sample allocation. Note that when $\theta = 0$ (a horizontal frontier), then $NL = 0$ and $c_1 = c_2 = 0$. When $\theta = 90$ (a vertical frontier), then $NL = 0$ and $c_1 = c_2 = 1$. The corresponding sample allocations are $(n_1, n_2) = (10, 0)$ and $(n_1, n_2) = (0, 10)$ for any values of λ_1 and λ_2.

An illustration of the sample allocations generated by each allocation rule and their resulting *rel fcs* values over the range of λ_1 for subcase 8.20.20 is provided in Figure 3. As in Figure 2, the shaded contour plots of Figure 3 present the *rel fcs* as a function of n_1 and λ_1, ranging from dark (low *rel fcs* values) to light (high, desirable *rel fcs* values). The solid squares within the plots denote n_1^*, the optimal sample allocation for attribute 1 at each λ_1 value. The solid line in each plot represents the sample allocation generated by the specified allocation rule for the range of decision weight λ_1.

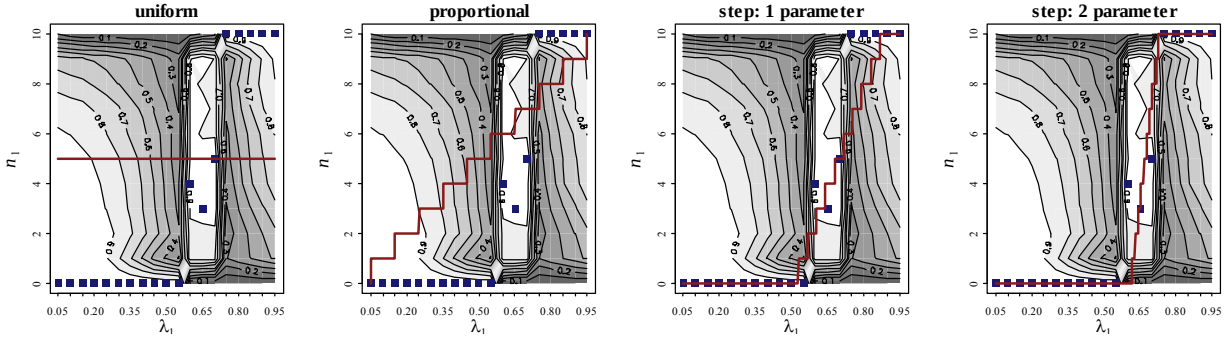

Figure 3: Sample allocations (solid line) generated by each allocation rule for training subcase 8.20.20. The shaded contour plots display the *rel fcs* as a function of n_1 and λ_1. The solid squares denote n_1^*, the optimal sample allocation for attribute 1.

A good allocation rule provides sample allocations that result in high *rel fcs* values for any decision weight (λ_1 value); an ideal rule provides a sample allocation equal to n_1^* at each λ_1 value. While no rule presented in Figure 3 provides an optimal sample allocation at every λ_1 value, the step allocation rules provide sample allocations that provide high – and often optimal – *rel fcs* values for most λ_1 values. Sample allocations provided by the uniform and proportional allocation rules are seen to result in less desirable (darkly shaded) *rel fcs* values over many values of λ_1.

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 instance generation process ensured that the alternatives formed a concave efficient frontier with attribute values restricted to the domain of [100, 200]. We calculated the nonlinearity (*NL*) and general angle (θ) measures for each case.

We tested the sample allocation rules using all 500 testing cases and 19 decision weight pairs in the value function, $(\lambda_1, \lambda_2) = (0.05, 0.95), (0.10, 0.90), \ldots, (0.95, 0.05$. To each of the 500 testing cases, we assigned a pair of measurement variability values, σ_1^2 and σ_2^2, to be associated with the two attributes, X_1 and X_2. The assigned σ_i values ($i = 1, 2$) were independent, random draws from a uniform distribution with parameters *min* = 1 and *max* = 30. Then, for each of the 11 sample allocations, for each of the 19 decision weight pairs, 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 identify a sample allocation. The *rel fcs* value for these allocations 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 assumptions provided by the Central Limit Theorem.

5. Results

The four sample allocation rules studied (*uniform*, *proportional*, *one-parameter step*, and *two-parameter step*) provided larger average *rel fcs* values across the range of decision weights than an arbitrary (random) allocation of experimental samples across alternatives. This underscores the importance of the experimental design process when embarking upon a data collection exercise to support a selection decision. The general ranking of the allocation rules in terms of performance (average *rel fcs*) from best-performing to worst-performing is as follows: two-parameter step, one-parameter step, proportional, uniform. At λ_1 values very near 0 and very near 1 the proportional allocation rule and the step rules provide similar sample allocations ($n_1 = 0$ at $\lambda_1 \approx 0$ and $n_1 = N$ at $\lambda_1 \approx 1$) and thus displayed similar performance at these decision weight values. At λ_1 values near 0.5 the proportional allocation rule and the uniform allocation rule provide similar sample allocations ($n_1 = n_2 = N/2$) and thus displayed similar performance at these decision weight values. Figure 4 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 evaluation cases and the 95 % confidence interval at each λ_1 value.

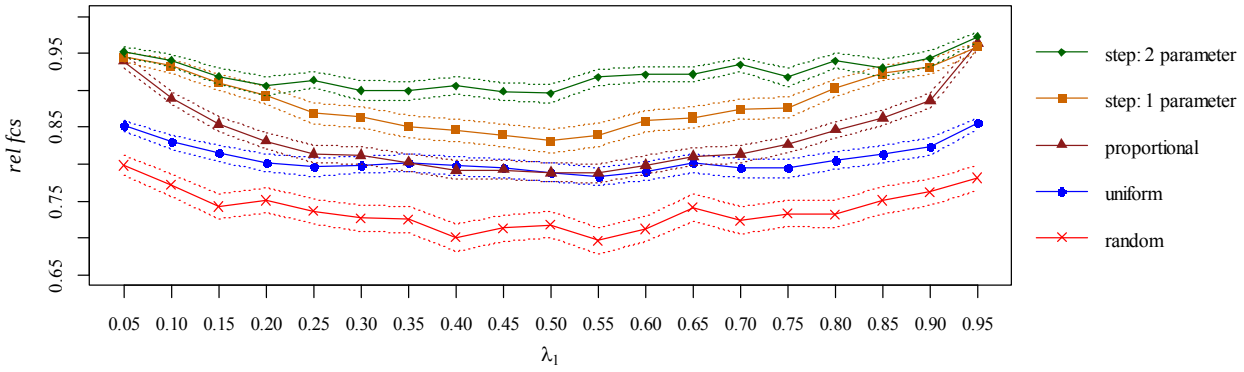

Figure 4: Relative frequency of correct selection averaged across all testing cases for each λ_1 value. The dotted lines represent the 95 % confidence intervals.

As shown in Figure 4, the sample allocations generated by the uniform allocation rule led, on average, to a *rel fcs* near 0.80, and this performance is nearly constant over most values of λ_1. The average *rel fcs* for the proportional allocation rule increases as λ_1 moves away from 0.5 and towards either $\lambda_1 = 0$ or $\lambda_1 = 1$. Overall, the average *rel fcs* for the proportional allocation rule is approximately 0.83. The one-parameter step rule provides a maximum average *rel fcs* value of 0.95 at $\lambda_1 = 0.95$, a minimum average value of 0.83 at $\lambda_1 = 0.5$, and an overall average *rel fcs* of 0.88. The two-parameter step rule provides a maximum average *rel fcs* value of 0.97 at $\lambda_1 = 0.95$, a minimum average value of 0.89 at $\lambda_1 = 0.5$, and an overall average *rel fcs* of 0.92. We thus conclude that, for nearly all values of λ_1, the two-parameter step rule, which leverages the frontier characteristic measures of overall theta and nonlinearity, provides average relative frequency of correct selection values that are statistically distinguishable (non-overlapping confidence intervals) and superior to the other allocation rules.

When λ_1 is near 0.5 (attributes are equally important), sample allocations that are not optimal perform very poorly because they fail to provide enough information about one of the attributes, which reduces the likelihood of selecting the truly best alternative. When one attribute is largely more important than the other (i.e., when λ_1 is near 0 or 1), and the less important attribute also has more uncertainty, we have seen that there is a range of near-optimal sample allocations that have a high *rel fcs*. Thus, it appears that it is easier to generate a near-optimal sample allocation when λ_1 is near 0 or 1, which would perhaps explain why the curves in Figure 4 tend be lower when λ_1 is near 0.5 and higher when λ_1 is near 0 or 1.

6. Summary and Conclusions

The ultimate goal of this work was to provide guidance on allocating a fixed experimental budget (samples) 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 experimental 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. This was clearly illustrated by the contour plots in Figure 2 of Section 4.2 where it was shown that for any given set of decision weights the relative frequency of correct selection can vary considerably based on the implemented sample allocation. We have shown that a sample allocation based upon the decision model weights as well as the characteristics of the set of alternatives (and their associated true attribute values) improves the probability of selecting the true best alternative over a sample allocation that does not consider this information. This emphasizes the importance for projects focused on a selection decision to be managed so that the decision modeling and the experimental 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 two-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, and the *one-* and *two-parameter step rules*. When the experiment is 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 experimental 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, particularly for $\lambda_1 < 0.25$ and $\lambda_1 > 0.75$ (See Figure 4). If, in addition to the decision model, the decision-maker can provide information regarding the shape of the concave frontier formed by the considered alternatives, then the proposed step rules can be leveraged. The one-parameter step rule requires knowledge of the general slope of the efficient frontier and prescribes 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 model weights considered. The two-parameter step rule requires additional information in the form of the nonlinearity measure of the efficient frontier, but the resulting sample allocations provide even further improvement in the frequency of correct selection over the allocations provided by the one-parameter step rule, particularly for $0.20 < \lambda_1 < 0.80$.

We observed these same general trends in the performance of the uniform, proportional, and one-parameter step allocation rules in our previous work that considered attributes evaluated using pass-fail testing. (We did not consider the two-parameter step allocation rule in our previous work.) A notable difference was that the case to case variability in the allocation rule performance was substantially smaller in the pass-fail testing study. We attributed this to the bounded nature of the Bernoulli parameter versus the unbounded positive range of values that the measurement variability can take in the normally distributed measurement error case.

Although the amount of measurement variability associated with each attribute was seen to impact the optimal sample allocations on an individual case basis (see Section 4.2), it was not seen to have an impact on the overall performance of the allocation rules. When one attribute is more important than the other, but the less important attribute has more uncertainty, we have seen that there is a range of near-optimal sample allocations that have a high *rel fcs*.

We expect that these results will hold in cases with more than five alternatives and decision situations with more than two attributes. A nonlinear single-attribute value function 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.

This work on allocating experimental samples across multiple attributes in a selection decision when the attribute values are subject to normally distributed measurement error expands upon our previous work in this area which focused on pass-fail testing (Bernoulli trials). While both of these studies considered only a single-phased experiment with two attributes, our future work will consider sequential allocation policies and will expand beyond two attributes.

References

1. Leber, D.D., and Herrmann, J.W., 2013, "Allocating Attribute-Specific Information-Gathering Resources to Improve Selection Decisions," Proceedings of the Winter Simulation Conference, December 8-11, Washington, DC, 101-112.
2. Bechhofer, R.E., Santer, T.J., and Goldsman, D.M., 1995, Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons, John Wiley and Sons, Inc., New York.
3. Kim, S.-H., & Nelson, B.L., 2006, "Selecting the Best System" appears in Handbook in Operations Research and Managment Science, Vol. 13, Henderson, S.G. and Nelson B.L. (eds.), Elsevier , Oxford, 501-534.
4. Branke, J., Chick, S.E., and Schmidt, C., 2007, "Selecting a Selection Procedure," Management Science, 1916-1932.
5. LaPorte, G.J., Branke, J., and Chen, C.-H., 2012, "Optimal Computing Budget Allocation for Small Computing Budgets," Proceedings of the Winter Simulation Conference, December 9-12, Berlin, Geramany, 3324-3336.
6. Butler, J., Morrice, D.J., and Mullarkey, P.W., 2001, "A Multiple Attribute Utility Theory Approach to Ranking and Selection," Management Science, 800-816.
7. Box, G.E., Hunter, J.S., and Hunter, W.G., 2005, Statistics for Experimenters, Second ed., John Wiley & Sons Inc., Hoboken, NJ.
8. Montgomery, D.C., 2013, Design and Analysis of Experiments, Eighth ed., John Wiley & Sons, New York.
9. Chaloner, K., and Verdinelli, I., 1995, "Bayesian Experimental Design: A Review," Statistical Science, 10(3), 273-304.
10. Roy, B., 2005, "Paradigms and Challenges," appears in Multiple Criteria Decision Analysis State of the Art Surveys, Figueira J., Greco S., and Ehrgott M. (eds.), Springer, New York, 3-24.
11. Keeney, R.L., and Raiffa, H., 1993, Decisions with Multiple Objectives - Preferences and Value Tradeoffs, Second ed., Cambridge University Press, New York.
12. Leber, D.D., and Herrmann, J.W., 2012, "Incorporating Attribute Value Uncertainty into Decision Analysis," Proceedings of the 2012 Industrial and Systems Engineering Research Conference, May 19-23, Orlando, FL.
13. Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B., 2004, Bayesian Data Analysis, Second ed., Chapman & Hall/CRC, New York.