ABSTRACT

Title of Dissertation:	OPTIMAL BUDGET-CONSTRAINED SAMPLE ALLOCATION FOR SELECTION DECISIONS WITH MULTIPLE UNCERTAIN ATTRIBUTES
	Dennis D. Leber, Doctor of Philosophy, 2016
Dissertation directed by:	Professor Jeffrey W. Herrmann Department of Mechanical Engineering

A decision-maker, when faced with a limited and fixed budget to collect data in support of a multiple attribute selection decision, must decide how many samples to observe from each alternative and attribute. This allocation decision is of particular importance when the information gained leads to uncertain estimates of the attribute values as with sample data collected from observations such as measurements, experimental evaluations, or simulation runs. For example, when the U.S. Department of Homeland Security must decide upon a radiation detection system to acquire, a number of performance attributes are of interest and must be measured in order to characterize each of the considered systems. We identified and evaluated several approaches to incorporate the uncertainty in the attribute value estimates into a normative model for a multiple attribute selection decision. Assuming an additive multiple attribute value model, we demonstrated the idea of propagating the at-

tribute value uncertainty and describing the decision values for each alternative as probability distributions. These distributions were used to select an alternative. With the goal of maximizing the probability of correct selection we developed and evaluated, under several different sets of assumptions, procedures to allocate the fixed experimental budget across the multiple attributes and alternatives. Through a series of simulation studies, we compared the performance of these allocation procedures to the simple, but common, allocation procedure that distributed the sample budget equally across the alternatives and attributes. We found the allocation procedures that were developed based on the inclusion of decision-maker knowledge, such as knowledge of the decision model, outperformed those that neglected such information. Beginning with general knowledge of the attribute values provided by Bayesian prior distributions, and updating this knowledge with each observed sample, the sequential allocation procedure performed particularly well. These observations demonstrate that managing projects focused on a selection decision so that the decision modeling and the experimental planning are done jointly, rather than in isolation, can improve the overall selection results.

OPTIMAL BUDGET-CONSTRAINED SAMPLE ALLOCATION FOR SELECTION DECISIONS WITH MULTIPLE UNCERTAIN ATTRIBUTES

by

Dennis D. Leber

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Jeffrey W. Herrmann, Chair
Professor Michel Cukier
Professor David J. Lovell
Professor Peter A. Sandborn
Professor Linda C. Schmidt

Acknowledgments

I would like to thank my advisor, Dr. Jeffrey Herrmann, for his continual support, patience, and guidance during this work. I greatly appreciate the sense of partnership that he has provided throughout my studies; this style of teaching is something that I will carry on.

I acknowledge the support of the National Institute of Standards and Technology and my colleagues in the Statistical Engineering Division. They each touched this work in one way or another. I give particular recognition to the guidance and contributions of Dr. James Filliben, Dr. Adam Pintar, and Mr. William Guthrie who are always willing to engage to help further my knowledge and understanding; you have played a significant role in this research and, more importantly, in my continual intellectual growth.

To my parents, David and Carol, for instilling the confidence, drive, and work ethic necessary to complete this monumental personal achievement.

And most importantly, I would like to thank my amazing and loving wife, Maria, and our incredible sons, Chase and Mason, for their understanding and support of me during this selfish endeavor. Your prevailing love, laughter, and encouragement keep me going every day.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Decision-makers are often required to make decisions based on limited knowledge as a consequence of time and resource constraints on their information gathering efforts. These decisions are further complicated when important characteristics about the alternatives are described by uncertain quantities such as estimates based on experimental or measurement results.

We define a decision where a decision-maker must select a single alternative from a finite set of alternatives and each alternative is described by several characteristics that are important to the decision-maker as a *multiple attribute selection decision*. We use the term *attribute value uncertainty* to refer to the uncertainty in the quantities that describe the alternatives' characteristics that are important to the decision-maker. A common normative model of the multiple attribute selection decision entails the decision-maker placing a decision value on each alternative by considering the tradeoffs amongst, and the desirability of, the alternatives' important characteristics and choosing the alternative that provides the greatest decision value (Keeney & Raiffa, 1993; Clemen & Reilly, 2001; Kirkwood, 1997; Dyer, 2005).

We illustrate this model in Figure 1.1, where m alternatives, $a_1,\ldots,a_m$, are each described by k important characteristics (attributes). The values of the attributes, $\mu_{i1},\ldots,\mu_{ik}$, are used by the decision-maker in developing a decision value, ξ_i, for each of the

$i = 1, \ldots, m$ alternatives. The decision-maker then selects the alternative, a_s, that has the largest decision value.

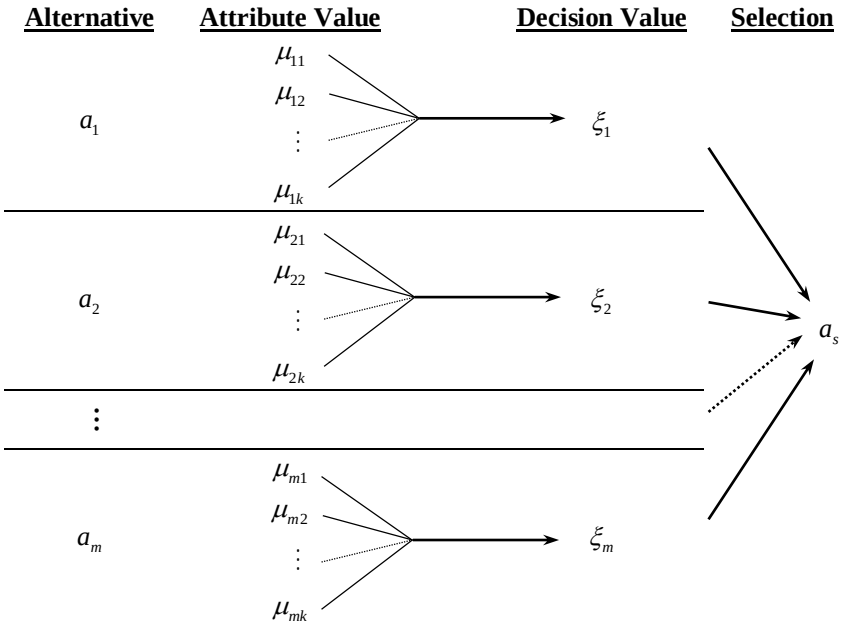

Figure 1.1: Conceptual model of a multiple attribute selection decision.

In this dissertation, we consider the problem of incorporating attribute value uncertainty in a normative decision model and, further, how a decision-maker might optimally use his limited information-gathering budget to identify the best alternative.

1.1 Motivating Examples

This section describes three motivating examples. The first example, regarding the uncertainty in the United States' census values and Congressional reapportionment, illustrates a series of single attribute selection decisions that contain attribute value uncertainty. Selecting a consumer service based on customer reviews provides an example of a multiple at-

tribute selection decision with attribute value uncertainty. And finally, the selection decision faced by the U.S. Department of Homeland Security in choosing a radiation and nuclear detection system to install at U.S. based international arrival airport terminals provides an example of a multiple attribute selection decision with attribute value uncertainty where the allocation of the fixed information-gathering budget may be altered to optimally support the identification of the best system.

1.1.1 Census Uncertainty in U.S. Congressional Reapportionment

Article I, Section 2 of the United States Constitution mandates that, every 10 years, a census of the United States be taken. The census yields data about millions of families that is valuable to government agencies, social scientists and economists, marketing firms, and many others who want to know who lives where. Though large, the resources of the U.S. Census Bureau are limited. Insufficient resources and enumeration errors imply that the census data collection methods cannot determine the precise population of every state, county, and place in the United States. For the 2010 Census, the Census Bureau conducted a post-enumeration study (the 2010 Census Coverage Measurement program, or CCM survey) to estimate the size of the error in the census counts. This study estimated a net overcount of 0.01% (Mule, 2012) (Davis & Mulligan, 2012). Although this was smaller than the 2000 Census net undercount (0.49%), the 2010 study showed more variability in the state-level errors for the 2010 Census (some states had a larger undercount or overcount than in 2000). In another attempt to understand the uncertainty in the population estimates, the Census Bureau used demographic analysis to estimate the population change from demographic data such as births and deaths, and this analysis yielded a range of population

estimates. The resulting low and high estimates for the population of the United States were 305 million and 312 million (U.S. Census Bureau , 2010).

The primary reason for the decennial census is to assign seats in the House of Representatives. How to assign this integer number of representatives in a fair manner to the States of the Union has been an often debated question dating back to the Constitutional Convention of 1787. Over the years, a number of methods proposed by prominent politicians and mathematicians have been employed in the reapportionment of Congress. Balinski and Young (1982) provide a comprehensive historical note of these methods.

The total number of seats in the House of Representatives currently equals 435, and each state will be allocated at least one seat. The decision of which Congressional apportionment realization to implement is the responsibility of the United States Congress and is currently computed using the method of equal proportions. This method has been implemented computationally as follows (Burnett, 2011): The first 50 seats are allocated with each state receiving one seat, which satisfies the minimum requirement. Each subsequent seat is assigned to the state with the largest priority value. (That is, the state with the largest priority value is selected for that seat.) Each state's priority value depends upon its population and the number of seats already assigned.

Because uncertainty exists in the estimate of the states' populations, uncertainty is also prevalent in the priority value, and thus each selection decision has attribute value uncertainty.

1.1.2 Selecting Consumer Services Based on Ratings

Consumers today can access personal opinions regarding just about anything with only a few clicks of a mouse. While some opinions – fashion, entertainment, political – may be appealing only in the eye of the beholder, others can be very useful for consumers.

Although individual qualitative reviews are certainly useful, more quantitative summaries of consumers' opinions are available from non-profit organizations such as Consumers Union and the Center for the Study of Services. Such surveys attempts to assess the true values of performance rating criteria for products and services. Because any survey is based on limited data, the summary statistics used for the performance rating criterion value are estimates of their true values and contains uncertainties, i.e., attribute value uncertainty. When selecting a product or service provider, the decision-maker should consider how this uncertainty affects the relative desirability of the alternatives.

For example, the Washington Consumers' Checkbook (Center for the Study of Services, 2011) includes ten performance rating criteria for 94 roofing firms. The results of the review were obtained through a survey of the organization's members. For six of the criteria, the measure provided is the proportion of customers surveyed who rated the firm as "superior." The review's performance criteria values are uncertain estimates of the true attribute values that describe each roofer. Consider a consumer who wishes to choose a roofer for a job and is using the survey results to inform this decision. This consumer faces a multiple attribute selection decision with attribute value uncertainty.

1.1.3 Selecting a Radiological and Nuclear Detection System

In 2008 the Domestic Nuclear Detection Office (DNDO) of the U.S. Department of Homeland Security was congressionally mandated to work with the U.S. Customs and Border Protection (CBP) on evaluating and improving the radiation detection systems in U.S. based international airports. As a result of this mandate, the DNDO PaxBag pilot program was initiated to identify the best possible system design for detecting, identifying and localizing illicit radiological or nuclear (rad/nuc) material entering the United States through international passenger and baggage screening. A major objective of the PaxBag program was to identify a system that displayed strong potential for improved capability over currently deployed technology to put forth for an operational demonstration.

Passengers arriving at U.S. based international airport terminals are subject to two mandatory encounters with CBP agents. As illustrated in Figure 1.2, passengers exit the aircraft and choose one of several queues prior to the first CBP encounter at the passport control booths. Upon exiting the passport control booths, passengers proceed to the luggage carousel to claim any luggage they may have. With luggage in tow, passengers proceed to the second CBP encounter at one of several customs declaration booths before exiting the international arrival terminal for the main terminal of the airport. During each of these encounters, or at any other time within the international arrival terminal, passengers may be subject to further interrogation as deemed necessary.

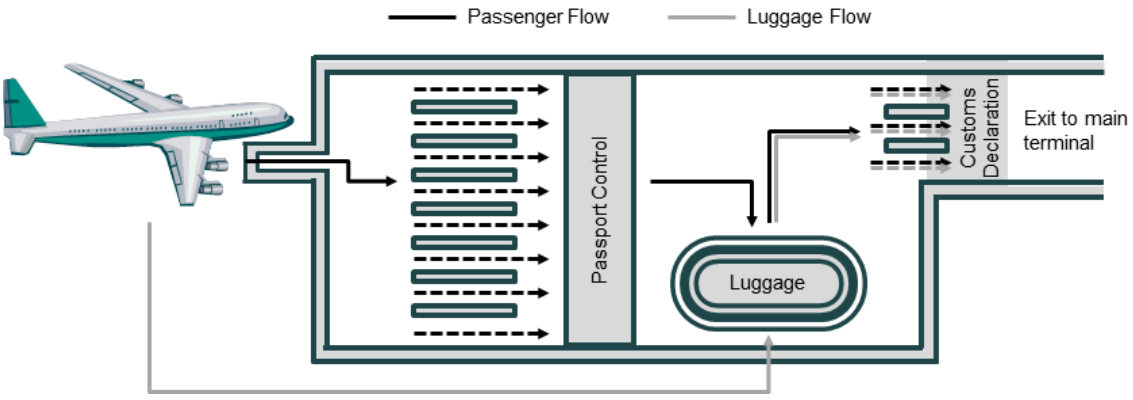

Figure 1.2: Generic passenger and luggage flow at a U.S. based international arrival terminal.

With respect to screening, a detection system has two subsystems: passport control and customs declaration. Each subsystem will have one of three detection sensor technologies. In each subsystem, the sensor technology will be placed either at the booths containing the CBP agents or in the queues prior to the booths and used in one of four operational modes. Thus, there are $3 \times 2 \times 4 = 24$ combinations of sensor technologies, sensor locations, and operational modes for each subsystem, which yields a total of $24^2 = 576$ possible pairs of subsystems. The detection system alternatives in this multiple attribute selection decision are these 576 possible combinations. The goal of the decision analysis is to select the detection system alternative that maximizes the value function as defined by the DNDO and CBP preference structure.

Three main objectives (Figure 1.3) were considered in developing the decision model: (1) maximize material interdiction, (2) minimize operational impact, and (3) minimize cost. Sub-objectives and attributes were developed to support the main objectives.

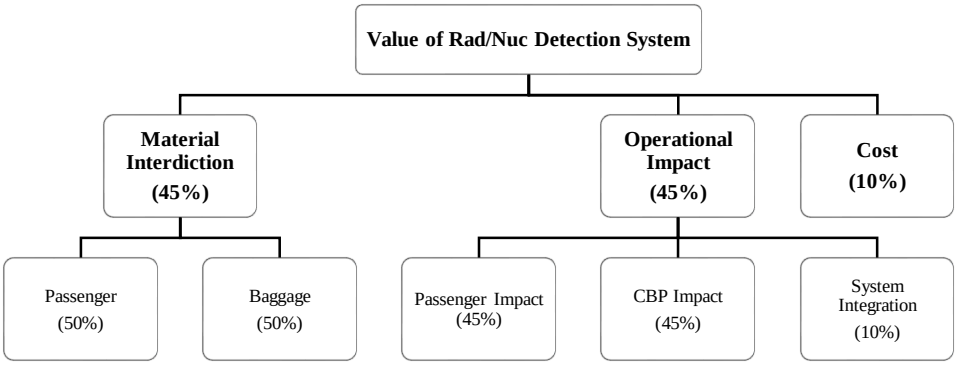

Figure 1.3: Objectives and associated weights for the selection of a radiation detection system.

Due to the uncertainty in their assessment, those attributes of direct interest to this research were the detection system performance measures that supported the material interdiction objective (interdiction performance attributes). These attributes were based on the detection systems' ability to detect and identify eleven radiological and nuclear sources of specific interest to DNDO which included Special Nuclear Material and radioisotopes used in industry and medicine. All other attributes (those that supported the second and third objectives) were considered to contain no uncertainty in this analysis.

To provide a value for each of the interdiction performance attributes, the detection and identification performance of the three detection sensor technologies against the eleven rad/nuc sources under the four different operational modes was experimentally evaluated in a laboratory setting. The results of these limited number of experimental evaluations were expressed as probability estimates with associated uncertainties and provided estimates of the sensor's true capabilities. By accounting for the sensor technologies and operational modes of each subsystem, these uncertain probability estimates were used in the evaluation of the interdiction performance attributes for each detection system alternative.

These uncertain attribute values were combined with the remaining attribute values using a multiple attribute decision model.

To select the detection system to put forth for the operational demonstration, DNDO and CBP formulated a multiple attribute selection decision model and developed a laboratory experimental plan to support the estimation of the attribute values. The development of the experimental plan 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 selection decision processes where the values of multiple attributes are estimated based upon experimental evaluations.

1.2 Research Questions

A decision-maker faced with a limited and fixed budget for collecting information about multiple attributes of the alternatives in a selection decision must decide how much of his information-gathering budget to allocate to each attribute of each alternative. This allocation decision is important when the information gained leads to uncertain estimates of the attribute values as with sample data collected from observations such as measurements, experimental evaluations, or simulation runs. These uncertain attribute values lead to uncertainty in selecting the true best alternative. When more information is gathered about one attribute of one alternative, its value becomes more certain, but this comes at the expense of less information (more uncertainty) about the remaining attributes and alternatives.

To understand how the allocation decision impacts the decision-maker's ability to select the true best alternative (correct selection), we must first understand how the uncertainties in the attribute values may be reflected in the selection decision model. In this dissertation we address the following research questions as they pertain to this described situation:

1. What approaches can a decision-maker use to incorporate and propagate uncertainty and identify the best alternative when faced with a multiple attribute selection decision with attribute value uncertainty? What are the advantages and disadvantages of these approaches?
2. What approaches can be used to inform the allocation of a limited and fixed budget for collecting observations across multiple attributes of multiple alternatives in support of a multiple attribute selection decision with attribute value uncertainty? How well do each of these approaches perform?

1.3 Notation

As a convenience to the reader, here we provide a list of notation used throughout this dissertation.

a_i	An individual alternative, *alternative i*; throughout, alternatives are indexed by *i*
$\{a_1,\ldots,a_m\}$	Set of *m* alternatives
Attribute *j*	An individual attribute, *attribute j*; throughout, attributes are indexed by *j*
μ_{ij}	The true value of attribute *j* of alternative a_i

X_{ij}	A random variable that represents the outcome of an evaluation of attribute *j* of alternative a_i
x_{ij}	An observed realization (sample observation) of an evaluation of attribute *j* of alternative a_i
n_{ij}	An integer number of observations for alternative a_i, attribute *j*
$\mathbf{x}_{ij}$	Set of n_{ij} observations from an evaluation of attribute *j* of alternative a_i, i.e., $\left(x_{ij1},\ldots,x_{ijn_{ij}}\right)$.
$\mathbf{x}_i$	The set of all observations collected to support alternative a_i, i.e., $\left(\mathbf{x}_{i1},\ldots,\mathbf{x}_{ik}\right)$
$\mathbf{X}$	The set of all observations collected to support the selection decision, i.e., $\left(\mathbf{x}_1,\ldots,\mathbf{x}_m\right)$
λ_j	The decision weight associated with attribute *j*
ξ_i	Decision value for alternative a_i

1.4 Dissertation Overview

The remainder of this dissertation is organized as follows: Chapter 2 is a review of the literature that provides a foundation for the work in this dissertation. Chapter 3 addresses the problem of incorporating attribute value uncertainty into a selection decision. We provide assumptions and present several approaches for selecting an alternative when the attribute value uncertainty has been propagated through the decision model and onto the decision value. Using the methods presented in Chapter 3, Chapter 4 develops and compares several methods for allocating a constrained information-gathering budget across the multiple attributes and alternatives of a multiple attribute selection decision. We consider

both Gaussian and Bernoulli error distributions and present optimal one-shot and sequential allocation procedures. Chapter 5 closes this dissertation with conclusions and proposals for future work.

Chapter 2 Literature Review

The problems of formulating a model to describe a decision amongst alternatives with multiple characteristics important to the decision-maker, deciding how to best allocate experimental effort when gathering information, and selecting a single alternative from a set of alternatives that are described by uncertain or random metrics span several disciplines which include decision analysis, statistical design of experiments, and ranking and selection. In this chapter, we review works from these disciplines and discuss their limitations in addressing the research questions of this dissertation.

2.1 Decision Analysis

At the core of the problem studied in this dissertation, we have a decision problem. More specifically, the decision-maker faces a multiple attribute decision problem with uncertainty in the attribute values. Ron Howard first coined the term *decision analysis* in a 1966 conference talk (Howard, 1966) where he provided a formal procedure for the modeling and analysis of decision problems. Active work in this field had been taking place for more than a decade prior to Howard's introduction of the terminology. Notable contributions during this time include works from von Neumann and Morgenstern (1944), Savage (1954), and Luce and Raiffa (1957). These works provided the foundation to formally address, through analytical methods, the decision problem for which the consequence of the action cannot be realized until some uncertain event is resolved, i.e., decisions with risk.

The method of expected utility theory first formalized by von Neumann and Morgenstern and later put into practical terms for multiple attribute decision analysis in the text of Keeney and Raiffa (1993) provides a structured approach to decision analysis when uncertain events exist through the consideration of the probability distributions over the potential outcomes of the uncertain event.

A decision-maker's preference structure is fundamental to all decisions. Models to describe a decision-maker's preference include ordinal value functions, measureable value functions, and utility functions. Dyer (2005) provides a comprehensive overview of these models, their applications, underlying assumptions, and assessment methods. In brief, ordinal value functions are applicable in decisions under certainty. They lead to a rank ordering of the decision alternatives, but they do not indicate the magnitude of preference among the alternatives. Measureable value functions, also applicable in decisions under certainty, provide an interval scale of measurement; that is, the decision-maker's strength of preference amongst the alternatives is captured. Finally, utility functions are applicable in decisions with risk. The utility model of one's preference structure not only considers the decision-maker's values of the potential consequences but also incorporates his psychological reactions to taking risks. Dyer and Sarin (1979), von Winterfeldt and Edwards (1986), Farquhar and Keller (1989), Keeney and Raiffa (1993), and Kirkwood (1997) provide in-depth discussions of these preference structure models as they apply to both single and multiple attribute decisions.

Multiple attribute decision methods require the decision-maker to consider tradeoffs amongst the multiple criteria being addressed in the decision analysis. The tradeoffs are quantified and modeled by the decision-maker's preference, or attribute, weights. Pandey,

Nikolaidis, and Mourelatos (2011) studied the uncertainty in defining the general structure of the decision-maker's preferences, while others, e.g., (Kahn & Meyer, 1991; Mustajoki et al., 2005; Chambal et al., 2011) have considered the difficulty faced by a decision-maker, particularly a group of decision-makers, in precisely defining attribute weights. The stochastic multiobjective acceptability analysis (SMAA) methods (Lahdelma et al., 1998) were developed to help multiple decision-makers explore the impact of uncertain or unspecified attribute weights. Rather than producing a ranking of alternatives based on a precisely defined preference structure, the SMAA methods describe the ranking of alternatives though several measures across the entire parameter space of preference weights by computing multidimensional integrals. In practice, these multidimensional integrals are estimated using Monte Carlo simulation. The original SMAA method included three descriptive measures. The SMAA-2 method (Lahdelma & Salminen, 2001) expands the descriptive measures to five. Other versions of the SMAA method, including methods to consider ordinal measures, are reviewed in the survey paper of Tervonen and Figueira (2008).

Beyond the uncertainties in the decision-maker's preference structure and about which outcome will occur in a risky decision, a decision-maker may face decision ambiguity. First defined in the 1960s by Daniel Ellsberg (1961), who is best known in the decision analysis community for his now infamous Ellsberg Paradox (see (Einhorn & Hogarth, 1986) for a well described presentation), the term decision ambiguity in the decision analysis context has since been generalized and elaborated on by many. Frisch and Baron (1988) gave the following definition: "Ambiguity is uncertainty about probability, created by missing information that is relevant and could be known." In decision analysis, ambiguity is specific to the probabilities used to describe an uncertain event in a risky decision.

Some have used this idea to challenge the validity of utility theory, particularly as a descriptive theory, though most proponents of utility theory argue that the theory was meant only as a normative one (Raiffa, 1961; Frisch & Baron, 1988; Dyer, 2005). Others have attempted to expand utility theory to include ambiguity (see (Srivastava, 1997) as an example). Herrmann (2015) used the idea of decision ambiguity in the discussion of the value of information. In some cases, a decision-maker may be able to collect additional information to reduce his uncertainty about the probability of a future uncertain event. In short, decision ambiguity refers to the uncertainty in describing the probability profile of a risky decision.

To summarize, uncertainty in decision making is not a new concept. The theory of expected utility addresses the decision problem for which an uncertain future event stands between the decision at hand and the realized consequences. Decision ambiguity considers the uncertainty involved in describing the probability profile of the uncertain event in a risky decision. Other examples of uncertainty in decision making include uncertain decision-maker preference structures and uncertainty in attribute weights. Although these theories and methods encompass many aspects of uncertainty in decision making, they all presume that the consequences (described by the attribute values) are precisely defined and neglect any uncertainty that may exist in their assessment. The PROMETHEE outranking technique allows for some degree of attribute value uncertainty to enter the decision model through the decision-maker's selection of a generalized criterion function (Hyde et al., 2003; Zhang et al., 2010). But we are unaware of any published work that allows for an explicit representation of the attribute value uncertainty to enter the decision model.

2.2 Experiment Design

The statistical design of experiments 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 et al. (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 generically represented by $y = f\left(l_1,\ldots,l_k\right)$, where y is the response variable of interest, there are k experimental factors that each have multiple levels, 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 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 & 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.

Although the research questions in this dissertation are also concerned with sample allocation, the principles and techniques of statistical design of experiments provide guidance on allocating samples across a design space defined by experimental factors and levels, which is different than allocating samples across multiple attributes of multiple alternatives in a selection decision. Using the statistical design of experiments terminology, the measured values for each of the attributes may be viewed as a response variable dependent on the single factor, alternative, with the number of levels equal to the number of alternatives considered. Thus the sample allocation problem of this dissertation is to allocate a sample budget across these multiple responses, each evaluated by a different experiment and whose importance differs according to the attribute weights, with the goal of maximizing the decision-maker's ability to select the most preferred alternative according to his preference model. To our knowledge, the work of the statistical design of experiments discipline does not address these needs.

2.3 Ranking and Selection

Most closely related to the topic of this dissertation are the methods of ranking and selection. Ranking and selection methods are used to compare a finite number of alternatives whose performance measures are generated by a stochastic process, e.g., experimentation via physical measurements or computer simulation. The study of ranking and selection first gained traction in the 1950s in the statistics community with the noteworthy publication of Bechhofer (1954). Gupta and Panchapakesan (1979) published the first modern text on the

subject, and Bechhofer et al. (1995) is a more recent text. During this time, the field of computer simulation, and in particular discrete event simulation, began advancing the work of ranking and selection and now accounts for much of the research in the area.

Kim and Nelson (2006) provided an extensive overview of the recent developments in ranking and selection with a focus on the indifferent zone (IZ) allocation procedure for selecting the alternative with the largest expected value. The IZ procedure, which has no fixed limit on the number of observations, determines how often each alternative is observed (sampled) while guaranteeing a specified probability of correct selection provided that the true performance of the "best" alternative exceeds that of its closest competitor by an amount the experimenter wishes to detect. Butler et al. (2001) applied the IZ procedure to a multiple attribute decision problem using a multiple attribute value model.

Kim and Nelson described four classes of comparisons as they relate to ranking and selection problems: selecting the alternative with the largest or smallest expected performance measure (selection of the best), comparing all alternatives against a standard (comparison with a standard), selecting the alternative with the largest probability of actually being the best performer (multinomial selection), and selecting the system with the largest probability of success (Bernoulli selection). In developing an experimental approach for each class, a constraint is imposed on either the probability of correct selection or on the overall experimental budget. That is, some procedures (e.g., indifference zone procedures) attempt to find a desirable alternative with a guarantee on the probability of correct selection with no regards to the experimental budget, and other procedures attempt to maximize the probability of correct selection while adhering to an experimental budget constraint. Computational results presented by Branke et al. (2007) demonstrated the strengths and

weaknesses of several ranking and selection procedures including the indifference zone, the expected value of information procedure, and the optimal computing budget allocation.

Unlike the IZ allocation procedure, the Optimal Computing Budget Allocation (OCBA) procedure derives a sample allocation based on a fixed budget with the goal of maximizing the probability of correct selection subject to this budget constraint. Thus, OCBA is the ranking and selection procedure most similar to our work. Chen and Lee have published many articles on the subject and presented a comprehensive collection of ideas in a recent text (Chen & Lee, 2011). 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. LaPorte et al. (2012) developed a version of OCBA that is useful when the computing budget is extremely small.

Although the ranking and selection methods have some similarities to the problem considered in this dissertation, a main difference is that they all consider the allocation of samples across multiple alternatives with a single performance measure, while our work is focused on the allocation of samples across both the multiple alternatives and the multiple attributes. We used the idea of Butler et al. (2001) and combined the multiple uncertain attribute values using a multiple attribute decision model to provide an alternative's overall performance measure. As the OCBA procedure is a sequential procedure that derives a sample allocation based on a fixed budget we adapted ideas from OCBA in our development of the sequential allocation procedure.

Chapter 3 Decision Uncertainty in Alternative Selection

The Roman scholar Gaius Plinius Secundus, better known as Pliny the Elder (23-79), stated that, "*the only certainty is that nothing is certain*." Uncertainty and its assessment has become a popular topic in recent years. Lindley (2006) suggested that the reason for this popularity is that the rules for assessing and applying uncertainty are now understood and that past tendencies of suppressing uncertainties are no longer necessary.

Many terms and classifications have been defined relating to the concept of uncertainty. Bevington and Robinson (2002) discussed accuracy and precision as they relate to experiments in the physical sciences. The accuracy of an experiment is a measure of how close the results of an experiment are to the true value. The idea of accuracy is closely related to the statistical term bias, where the bias of an estimator is the difference between the expected value of the estimator and the true value of the parameter being estimated (DeGroot, 1989). Precision is a measure of the reproducibility of the result, irrespective of the agreement with the true value. In general, when the uncertainty of an experiment is noted, it is the precision of the experiment that is being referenced. It is often the case in metrology that the true value of the measurand is unknown, thus the accuracy cannot be determined and only the precision of the measurement can be quantified.

Many have attempted to categorize the sources of uncertainty in the decision analysis context. For example, French (1995) listed ten sources of uncertainty as it relates to decision analysis. He classified these sources into three broad groups: uncertainties expressed

during modeling (or problem structuring), uncertainties expressed during exploration of the models, and uncertainty expressed during interpretation. Stewart (2005) further classified these sources as internal and external uncertainties. Under this classification, internal uncertainties refer to uncertainties due to the problem formulation, structure of the model adopted, and the judgmental inputs required by the model. External uncertainties refer to the lack of knowledge about the consequences of a particular choice.

Stewart's external uncertainty class can be further detailed through the consideration of another common classification of uncertainty: aleatory versus epistemic uncertainty. Aleatory uncertainty is the uncertainty due to a random process; uncertainty as a result of natural, unpredictable variation in the performance of the system under study. Epistemic uncertainty is the uncertainty due to a lack of knowledge about the system under study. The focus of this chapter is incorporating attribute value uncertainties, and more specifically epistemic uncertainties, into a decision analytic model of a selection decision. The propagation of attribute value uncertainty results in an uncertain decision value for which the selection decision is based. We discuss four approaches to selecting an alternative from a set of alternatives described by uncertain attribute values and hence uncertain decision values. We conclude the chapter by applying these methods to the examples introduced in Chapter 1.

The discussions and examples found in this chapter are largely taken from the works of Leber and Herrmann (2012, 2013b).

3.1 Uncertainty in Decision Making

Decision problems can be classified on several dimensions. First, the decision-maker can be either an individual or a group. Second, the number of attributes used to describe the set of consequences can be a single attribute or can consist of multiple attributes. And finally, a decision problem may be classified under conditions of certainty, risk, or uncertainty. These conditions may be defined as follows (Luce & Raiffa, 1957):

1. *Decisions under certainty*: Each alternative is known to lead invariably to a specific outcome.
2. *Decisions with risk*: Each alternative leads to one of a set of possible outcomes, where each outcome occurs with a probability assumed to be known by the decision-maker. These outcomes may be the result of an uncertain future event, for example.
3. *Decisions under strict uncertainty*: Each alternative leads to one of a set of possible outcomes, though nothing is known or can be stated about the probability of the occurrence of each outcome.

The consequence associated with any decision is the result of the selected alternative and the outcome of relevant external factors that are outside the control of the decision-maker (e.g. uncertain future events). To illustrate this perspective, a decision may be represented as a decision table (Table 3.1). The *m* decision alternatives, $a_1,\ldots,a_m$, are the rows in the table. The columns in the table correspond to $s_1,\ldots,s_r$, the *r* mutually exclusive and collectively exhaustive possible outcomes of relevant external factors ("states of nature"). Associated with each possible state of nature is $P(s_l)$, the probability that s_l will be the

true state. As shown in each cell of the table, the consequence that ensues when alternative a_i is selected and s_l is the state of nature is described by *k* attributes and their associated attribute values, $\mu_{ijl}, i = 1,\ldots,m,\ j = 1,\ldots,k, l = 1,\ldots,r$.

Table 3.1: General form of a decision table with multiple attributes and multiple states of nature.

		State of Nature			
		s_1	s_2	…	s_r
Decision Alternatives	a_1	$\mu_{111},\ldots,\mu_{1k1}$	$\mu_{112},\ldots,\mu_{1k2}$	…	$\mu_{11r},\ldots,\mu_{1kr}$
	a_2	$\mu_{211},\ldots,\mu_{2k1}$	$\mu_{212},\ldots,\mu_{2k2}$	…	$\mu_{21r},\ldots,\mu_{2kr}$
	⋮	⋮	⋮	⋮	⋮
	a_m	$\mu_{m11},\ldots,\mu_{mk1}$	$\mu_{m12},\ldots,\mu_{mk2}$	…	$\mu_{m1r},\ldots,\mu_{mkr}$

Table 3.1 clearly displays the components of a decision: the alternatives, the possible states of nature, and the resulting consequences described by attribute values. When the decision components are viewed as displayed in Table 3.1, it becomes evident that any uncertainty in estimating the attribute values (attribute value uncertainty) is essentially unlike uncertainty about which of the set of possible outcomes, $s_1,\ldots,s_r$, will occur (risky decision) and uncertainty in defining the probability of each outcome (decision ambiguity).

While it may be conceivable to model the attribute value uncertainty as an uncertain event in a risky decision, we choose to maintain a decision model that distinguishes the attribute value uncertainty as a unique component of uncertainty. The reason is that a decision-maker can control, to some extent, the amount of uncertainty in an estimate of the true value of an attribute by varying the amount of information observed in its assessment, whereas the outcome of a future uncertain event cannot be controlled in this same manner.

Decision-makers often consider decision alternatives that have consequences that are described by uncertain attribute value estimates. If a decision problem includes attributes whose values are estimated based on a limited set of sample data collected from observations such as measurements, experimental evaluations, or simulation runs, then attribute value uncertainty exists. When these attribute values are provided only as point values, the decision-maker must move forward under the assumptions that the values are accurate and that the level of uncertainty associated with each alternative is equivalent.

Scientists and engineers are trained to quantify and report the uncertainties in assessments, including measured physical quantities such as mass and performance characteristics such as the probability of system failure. These uncertainties may be developed through a variety of techniques including data-based methods and subjective expert opinions.

When faced with attribute value uncertainty, the decision-maker confronts the risk of selecting an alternative that is not the best one, which in fact could be identified if no attribute value uncertainty existed.

3.2 Assumptions

Expanding on the general concepts introduced in Chapter 1, Figure 1.1, the random measured value, X_{ij}, used in estimating attribute *j* from alternative a_i adheres to a probability distribution, denoted $F_j\left(\mu_{ij}, \boldsymbol{\theta}_{ij}\right)$, that depends upon the attribute's true value, μ_{ij}, and other distributional parameters, $\boldsymbol{\theta}_{ij}$, to include the uncertainty associated with the measurement technique. Upon observing the n_{ij} measurements, $x_{ij1},\ldots,x_{ijn_{ik}}$, used to estimate the value of attribute *j* for alternative a_i, a decision-maker may describe the attribute value by

a single point, such as a sample mean, or a distribution, such a Bayesian posterior distribution. We denote this attribute value description as the probability distribution $G_j\left(\mu_{ij}, n_{ij}, \boldsymbol{\varphi}_{ij}\right)$ that depends upon the attribute's true value, μ_{ij}, the number of observed measurements, n_{ij}, and other distributional parameters, $\boldsymbol{\varphi}_{ij}$, to include the uncertainty associated with the measurement technique. The multiple attribute decision model, $f(\bullet)$, is used to combine the attribute values, leading to a decision value for each alternative. The decision-maker may describe the decision values as a single point or a distribution. We denote the decision value description as the probability distribution $H\left(\xi_i, \boldsymbol{\gamma}_i\right)$ that depends upon the alternative's true decision value, ξ_i and other distributional parameters, $\boldsymbol{\gamma}_i$, to include the uncertainty associated with the measurement techniques and the number of observed measurements for each attribute. Based on the decision-maker's description of the decision values, an alternative, a_s, is selected according to a selection rule that takes into account the information that is generated from the measurements. Note that a_s is random because it depends upon the random measurements. The conceptual model of a selection decision presented in Figure 1.1 is expanded in Figure 3.1 to include the measurement processes and distributional descriptions.

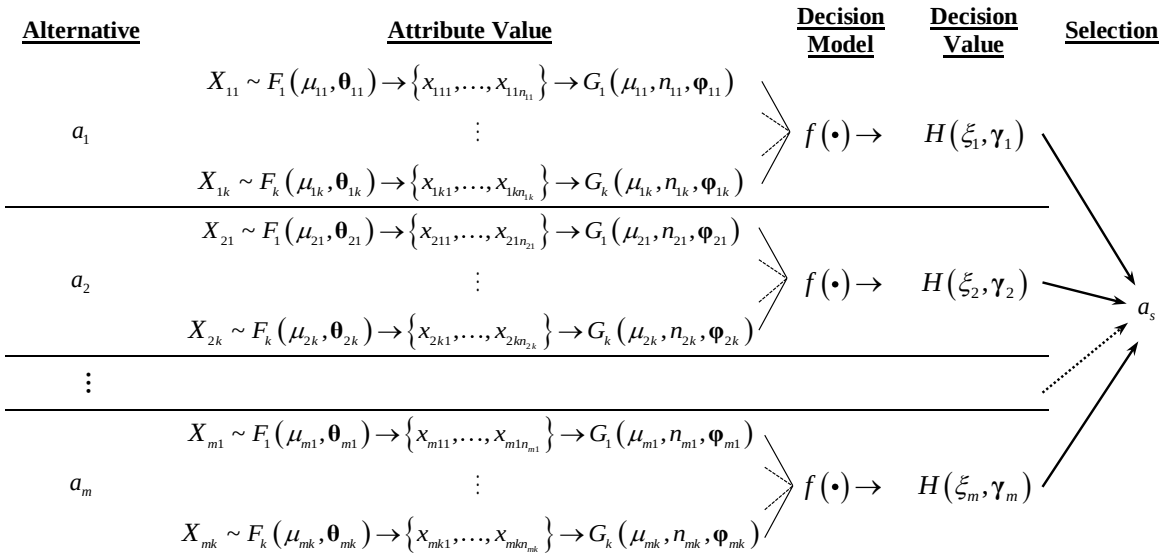

Figure 3.1: Model of a multiple attribute selection decision including measurement processes, value estimations, and distributional representations.

Given this expanded conceptual model of a multiple attribute selection decision with attribute value uncertainty, we make the following assumptions in this chapter.

1. The set of *m* distinct alternatives, $\{a_1, \ldots, a_m\}$, is provided, where *m* is a finite positive integer such that all alternatives can be assessed.
2. Each alternative is described by $k \geq 2$ attributes. The decision-maker's knowledge of the value of attribute *j* of alternative a_i is uncertain and is described by a probability distribution which depends on the attributes true value, μ_{ij}.
3. Also provided is a decision model, $\xi_i = f(\mu_{i1}, \ldots, \mu_{ik}) = \sum_{j=1}^{k} \lambda_j v_j(\mu_{ij})$, that reflects the decision-maker's preference structure and combines the multiple attribute values as provided by the *k* individual value functions, $v_j(\mu_{ij})$, to produce a decision value, ξ_i, for each alternative a_i. The attribute decision weights, λ_j, are defined such that $\sum_{j=1}^{k} \lambda_j = 1$.

Note that the probability distribution used to describe the attribute values (Assumption 2) may be a parametric probability distribution, such as a Gaussian or Binomial probability distribution, or it may be a nonparametric distribution, such as the empirical distribution of the collected set of sample observations.

3.3 Propagating Attribute Value Uncertainty

The ideal attribute value input to the decision model is the true, but often unknown, value. As previously discussed, when attribute values are obtained based on sample data collected from observations such as measurements, experimental evaluations, or simulation runs, the values acquired are merely estimates of the true attribute values. We denote the set of observed samples as $\mathbf{x}_{ij} = \left(x_{ij1}, \ldots, x_{ijn_{ij}} \right)$, where n_{ij} is the number of measurements obtained for attribute *j* of alternative a_i, and the total sample data observed in support of alternative a_i are $\mathbf{x}_i = \left(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{ik} \right)$. While the uncertainty in the attribute value estimates is fundamental to this work, it is the uncertainty in the decision values, ξ_i, that directly impacts the decision-maker's ability to make a correct selection. Describing the uncertain attribute value estimates with probability distributions allows for the attribute value uncertainties to be propagated through the decision model to the decision value. This section describes techniques for propagating the attribute value uncertainty in order to describe the decision value uncertainty; these will be essential parts of the sample allocation methods discussed in Chapter 4.

Under some models of attribute value uncertainty, it is possible to propagate the uncertainty to the decision value using an analytical approach. This is the case when the attribute

values are described by Gaussian (normal) probability distributions. Consider when the decision-maker's knowledge of the attribute values is described by the normally distributed Bayesian posterior distributions in Equation (3.1) where μ_{0ij} and τ_{0ij} are the parameters of the conjugate normal prior distribution and σ_j^2 is the variance of the measurement process used to measure attribute *j*.

$$\mu_{ij} \mid x_{ij1},\ldots,x_{ijn_{ij}} \sim N\left(\frac{\sigma_j^2\mu_{0ij} + n_{ij}\tau_{0ij}^2\bar{X}_{ij}}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}, \frac{\sigma_j^2\tau_{0ij}^2}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}\right) \; i=1,\ldots,m,\; j=1,\ldots,k \qquad (3.1)$$

As the attribute values for each alternative a_i are combined according to the linear decision model, $\xi_i = \sum_{j=1}^{k}\lambda_j\mu_{ij}$ (Assumption 3), it follows from the properties of the sum of normally distributed random variables (Casella & Berger, 2002) that the distribution of the decision-maker's knowledge of the true decision value, ξ_i, can be described by the posterior distribution in Equation (3.2).

$$\xi_i \mid \mathbf{x}_i \sim N\left(\sum_{j=i}^{k}\lambda_j\frac{\sigma_j^2\mu_{0ij} + n_{ij}\tau_{0ij}^2\bar{x}_{ij}}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}, \sum_{j=1}^{k}\lambda_j^2\frac{\sigma_j^2\tau_{0ij}^2}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}\right) \qquad (3.2)$$

When no closed form expression exists for the distribution of the decision values, e.g., when the decision value is a linear combination of Binomial random variables with differing success probabilities, then Monte Carlo simulation provides a method to propagate the attribute value uncertainty. Generally speaking, a Monte Carlo simulation is a computer experiment that consists of drawing a random sample (a number) from a specified distribution, performing some operation involving the drawn sample, and recording the result. This process is repeated numerous times building a distribution of results. Fishman (1996)

discussed error and sample size considerations as well as sampling techniques and variance-reducing techniques.

Monte Carlo simulation can be leveraged to describe the probability distribution of a function of multiple random variables. As the decision value in the multiple attribute selection decision model is a function of multiple attribute values that may be described by probability distributions, the application of Monte Carlo simulation to propagate the attribute value uncertainties to the decision values is a useful tool.

To summarize, when possible, the attribute value uncertainty can be propagated through the decision model using an analytical approach. Monte Carlo simulation offers a general method to propagate the attribute value uncertainty and is particularly useful when no closed form expression exists for the distributional description of the decision value.

3.4 Selection Approaches

Traditional decision analysis approaches clearly identify the most desirable alternative. This property should not be lost when expanding the model to be more comprehensive by including attribute value uncertainty. The result of propagating attribute value uncertainty is a set of uncertain decision value estimates that may be described by distributions. Thus, the problem of selecting an alternative changes from a simple ordering exercise to a comparison of distributions. This section discusses four approaches to compare the resulting decision value distributions: expected value, multinomial selection, stochastic dominance, and majority judgment.

3.4.1 Expected Value

Kim and Nelson (2006) use the term *selection of the best* to describe selecting the alternative with the largest or smallest expected performance measure. Based on this idea, we define the *expected value selection procedure* as selecting the alternative that has the largest expected decision value. For example, when the decision-maker's knowledge of ξ_i is described by the posterior probability distribution of ξ_i provided by Equation (3.2), we select the alternative a_s where $s = \arg\max_i \left(\sum_{j=i}^{k} \lambda_j \frac{\sigma_j^2 \mu_{0ij} + n_{ij} \tau_{0ij}^2 \bar{x}_{ij}}{\sigma_j^2 + n_{ij} \tau_{0ij}^2} \right)$. That is, the selected alternative is chosen by selecting the alternative with the largest posterior mean.

In the multiple attribute selection setting, the expected value selection procedure has the desirable trait of simplicity which likely leads to its common use in practice, even in the presence of attribute value uncertainty. This simplicity can be seen when the popular frequentist method of maximum likelihood is used to provide estimates of the attributes' true values using the maximum likelihood estimator $\hat{\mu}_{ij} = \bar{X}_{ij} = \frac{1}{n_{ij}} \sum_{l=1}^{n_{ij}} x_{ijl}$. From the invariant property of maximum likelihood estimators (Casella & Berger, 2002) it follows that the maximum likelihood estimators of the alternatives' true decision values are $\bar{\xi}_i = \sum_{j=1}^{k} \lambda_j \bar{X}_{ij}$. Thus the alternative a_s is selected where $s = \arg\max_i \bar{\xi}_i$. This selection is made without consideration for the uncertainty in the estimation of the attribute values.

Because the expected value procedure fails, in most cases, to consider the uncertainty in the decision values when making a selection, it may fail to select the alternative that maximizes the decision-maker's satisfaction.

3.4.2 Multinomial Selection

Multinomial selection approaches were originally designed for experiments with a categorical response, for example, which among five soft drinks will a subject say they prefer (Kim & Nelson, 2006). Goldsman (1984a, 1984b) suggested a more general perspective for the field of computer simulation. Given *m* competing alternatives, it is assumed that there is an unknown probability vector $\mathbf{p} = (p_1, \ldots, p_m)$ such that $0 \le p_i \le 1$ and $\sum_{i=1}^{m} p_i = 1$. The p_i are the probabilities that alternative a_i "wins" on any given trial, where winning is the observation of a most desirable criteria of goodness (e.g., the largest decision value). $\mathbf{p}$ thus defines an *m*-nomial probability distribution for winning over the set of alternatives. The goal of a multinomial selection procedure is to identify the alternative with the largest p_i.

We used the ideas of Goldsman to define the *multinomial selection procedure* which selects the alternative that has the greatest probability of having the largest decision value. That is, we define $p_i = P(\xi_i > \xi_r, \forall r = 1, \ldots, m, r \ne i)$ and select the alternative a_s where $s = \arg\max_i p_i$.

Because intractable multiple integrals are often involved in calculating the p_i, we implemented the multinomial selection procedure using Monte Carlo simulation to estimate the p_i. From each of the distributions for $\xi_i, i = 1, \ldots, m$, we draw a single realization and note the alternative with the largest realized value among the *m* values. We repeat this process a large number of times and tabulate the relative frequency, $\hat{p}_i$, that alternative a_i provided the largest realized value. We select the alternative a_s where $s = \arg\max_i \hat{p}_i$.

Though we did not use the techniques, we note that Miller et al. (1998) provided an efficient computational approach for implementing the multinomial selection procedure using resampling of the Monte Carlo samples. And recently Tollesfson et al. (2014) provided optimal algorithms for the multinomial selection procedure.

Due to the intractable multiple integrals often involved in calculating the p_i in the multinomial selection procedure, one must either spend the resources to calculate these integrals or use Monte Carlo simulation to estimate them, which is faster but may (will) be wrong. Using Monte Carlo simulation to obtain accurate estimations of the integrals may also become computationally expensive to implement. However, the approach clearly considers the uncertainty in the attribute value estimates. And further, when attribute and decision values are described using Bayesian posterior distributions, the multinomial selection procedure has the desirable property of directly maximizing the probability of correct selection (see Section 4.7.1). For these reasons, we use Bayesian posterior distributions in describing attribute and decision values and the multinomial selection procedure as a basis for much of the sample allocation work presented in Chapter 4.

3.4.3 Stochastic Dominance

Our third selection approach builds upon the concept of stochastic dominance for comparing distributions. In the following discussion ξ_1 and ξ_2 represent the uncertain decision values for alternatives a_1 and a_2 respectively. Hadar and Russell (1969) discuss stochastic dominance as an approach to predicting a decision-maker's choice between two uncertain events without knowledge of the decision-maker's utility function. They define two types

of stochastic dominance: first-degree stochastic dominance and second-degree stochastic dominance.

ξ_1 stochastically dominates ξ_2 in the *first degree* if and only if

$$P[\xi_1 \le y] \le P[\xi_2 \le y] \quad \forall y \tag{3.3}$$

That is, the value of the cumulative distribution for ξ_1 never exceeds that of ξ_2 for all $y \in \xi$.

When the support of ξ_1 and ξ_2 is contained within the closed interval [a, b], ξ_1 stochastically dominates ξ_2 in the *second degree* if and only if

$$\int_a^t P[\xi_1 \le y]dy \le \int_a^t P[\xi_2 \le y]dy \quad \forall t \in [a,b] \tag{3.4}$$

That is, the area under the cumulative distribution for ξ_1 is less than or equal to that of ξ_2 for $a \le \xi \le t, \forall t \in [a,b]$.

If ξ_1 is found to stochastically dominate ξ_2 in the first degree, then the decision-maker will prefer alternative a_1 to alternative a_2 as long as his utility function is monotonic. If ξ_1 is found to stochastically dominate ξ_2 in the second degree, then the decision-maker will prefer alternative a_1 to alternative a_2 as long as his utility function is concave (which implies that the decision-maker is risk-averse). Under these restrictions, if ξ_1 is found to stochastically dominate ξ_2 in either the first or second degree then the decision-maker will prefer alternative a_1 to alternative a_2 because alternative 1 will have a greater expected utility.

If there exists a single ξ_i that stochastically dominates $\xi_r, \forall r, i \neq r$ (first- or second-degree), and, in at least one case, the inequality in Equation (3.3) or (3.4) is found to be a strict inequality, then alternative *i* can be selected with few underlying assumptions (Hadar & Russell, 1969). We thus define the *stochastic dominance selection procedure* to select alternative a_s such that the distribution that describes the uncertain decision value, ξ_s, stochastically dominates $\xi_r, \forall r, s \neq r$ (first- or second-degree), and, in at least one case, the inequality in Equation (3.3) or (3.4) is found to be a strict inequality.

When the distributions of ξ_1 and ξ_2 are provided as empirical distributions, as in the case where *R* Monte Carlo simulation replicates are used to propagate the attribute value uncertainty, alternative a_1 dominates alternative a_2 based upon the ideas of first-degree stochastic dominance if, for all values *y*, the number of observed values of ξ_1 that are not greater than *y* is less than or equal to the number of observed values of ξ_2 that are not greater than *y*. To check for second-degree stochastic dominance, let $Z_i = \{y_{i[1]}, y_{i[2]}, \ldots, y_{i[R]}\}$ be the ordered set of the *R* decision values for alternative a_i where $y_{i[1]} \leq y_{i[2]} \leq \ldots \leq y_{i[R]}$. Let $f_i(y)$ be the number of decision values in Z_i that are less than or equal to *y*. Note that this is a step function that increases at each value in the set Z_i. Let *a* and *b* be the lower and upper bounds on the decision values for all of the alternatives. Alternative a_1 dominates alternative a_2 based upon the ideas of second-degree stochastic dominance if Inequality (3.5) holds.

$$\int_a^t f_1(y)dy \leq \int_a^t f_2(y)dy \quad \forall t \in [a,b] \tag{3.5}$$

Because $f_1(y)$ and $f_2(y)$ are step functions, it is easy to calculate these integrals for any value of *t*, and this condition holds for all $t \in [a,b]$ if it holds for all $t \in Z_1 \cup Z_2$.

Using the idea of stochastic dominance as a decision rule to select an alternative from a set of alternatives characterized by uncertain decision values is attractive because of the minimal restrictions necessary on the utility functions. Because the integrals required to evaluate stochastic dominance are univariate, they are substantially easier to compute than those required of the multinomial selection procedure. Although many alternatives can be easily dismissed from consideration due to dominance, this procedure may not produce a solution, and thus an alternative would not be identified for selection. This phenomenon is illustrated in the example in Section 3.7.

3.4.4 Majority Judgement

Consider an empirical distribution of *R* values that describe the uncertain decision value ξ_1 and another empirical distribution of *R* values that describe ξ_2. By viewing each value from these distributions as a score assigned by an individual judge or voter, the problem of selecting an alternative based on distributions of decision values may be viewed as one of social choice. A consensus value for each alternative that appropriately represents the message of all judges is sought in comparing and selecting the most desirable alternative. While many models of social choice exist, we consider the method of majority judgment.

In an attempt to identify a model of social choice that overcomes the shortcomings displayed by traditional social choice models such as the Borda and Condorcent methods, Balinski and Laraki (2007, 2010) proposed the method of majority judgment. The majority judgment method relies upon the middlemost interval to identify a social grading function

that has desirable functional properties, provides protection against outcome manipulation by individual voters or judges, and overcomes many of the shortcomings of traditional social choice models. When considering the ordered scores for alternative a_i, $y_{i[1]} \leq \cdots \leq y_{i[R]}$, the *majority-grade* is defined to be the median score, $y_{i\lceil (R+1)/2 \rceil}$, when R is odd and the lower bound of the middlemost interval, $y_{i[R/2]}$, when R is even. The majority judgment method identifies the alternative with the largest majority-grade as the most desirable alternative in the social choice context. If multiple alternatives have the same largest majority-grade, then a single majority-grade value is removed from the set of scores for each alternative in the tie, and the majority-grade of the new distributions are calculated. If a tie again occurs, this process is repeated until a single alternative has the largest majority-grade. The majority judgment method extends this concept to provide a complete rank-ordering termed the *majority-ranking*.

We define the *majority judgement selection procedure* to select the alternative with the largest majority-grade. To implement the majority judgment selection procedure, the majority-grade is computed for each alternative. When the decision values are described by probability distributions that are expressed analytically, the majority-grade is simply the median of the probability distribution. When the ξ_i are described by empirical distributions of R values, the majority-grade for each alternative is the median (if R is odd) or the lower bound of the middlemost interval (if R is even) of the distribution of decision values. The alternative with the largest majority-grade is then identified as the most desirable alternative. If a tie exists, the tie-breaking procedure defined by the majority judgment method is used to identify the single most desirable alternative.

As the majority-grade is defined based upon the middle-most interval, it emphasizes the significance of place in order rather than magnitude. That is, it is robust against extreme scores. Further, the majority-ranking provides, for any two alternatives, a ranking that depends upon the grades of only those two alternatives. In other words, the majority-ranking is independent of irrelevant alternatives (Arrow's IIA) (Arrow, 1951). But like the expected value selection procedure, the majority judgment selection procedure focuses on the locations of the decision value distributions and does little to consider the uncertainty aspects of these distributions.

3.5 Application to Congressional Reapportionment

In this section, we expand on the "Census Uncertainty in U.S. Congressional Reapportionment" example introduced in Section 1.1.1. We illustrate the application of methods presented in this chapter for propagating attribute value uncertainty and selecting an alternative.

3.5.1 Apportionment Decision Model

According to the Encarta Dictionary, apportionment is the division and allocation of something among different people or groups. Specific to the United States Congress, apportionment is the process of dividing the total seats in the House of Representatives among the 50 states according to their proportional population. The apportionment population for 2010 is defined to be the resident population of the 50 states plus overseas U.S. military

and federal civilian employees (and their dependents living with them) allocated to their home state, as reported by the employing federal agencies (Burnett, 2011).

For this example, let $\mathbf{P} = (P_1, \ldots, P_{50})$ be the population vector and $\mathbf{S} = (S_1, \ldots, S_{50})$ be the apportionment vector, where P_i represents the population of state *i* and S_i the number of Congressional seats allocated to state *i* for $i = 1, \ldots, 50$. The total number of seats in the House of Representatives to be allocated is currently equal to 435, and each state will be allocated at least one seat. The problem of Congressional reapportionment in the United States is to define an apportionment vector $\mathbf{S}$ that satisfies Article I, Section 2 of the United States Constitution.

The decision of which Congressional apportionment realization to implement is the responsibility of the United States Congress and is currently computed using the method of equal proportions. This method has been implemented computationally as follows (Burnett, 2011):

1. Seats 1 through 50 are allocated with each state receiving one seat, which satisfies the minimum requirement.
2. Seat 51 is assigned to the state with the largest priority value. (That is, the state with the largest priority value is selected for that seat.) The priority value, v_{ij}, for each state $i = 1, \ldots, 50$ and seat assignment $j = 51, \ldots, 435$ is calculated according to Equation (3.6) where s_{ij} is the number of seats currently assigned to state *i*. Thus for the 51st seat assignment where all states are currently assigned a single seat, $s_{i51} = 1, \forall i = 1, \ldots, 50$.

$$v_{ij} = P_i \frac{1}{\sqrt{(s_{ij}+1)s_{ij}}} \tag{3.6}$$

3. Upon assigning the 51st seat, the s_{ij} are updated to reflect the new allocation, Equation (3.6) is used to calculate the revised priority values, and the 52nd seat is assigned to the state with the largest priority value.
4. This process is repeated until all 435 seats have been assigned.

This computational implementation of the method of equal proportions leads naturally to a decision model that consists of a series of 385 decisions of selecting the state to receive the j^{th} seat, $j = 51,\ldots,435$. Each decision consists of 50 alternatives, the 50 states, with a single attribute, the priority value. The criterion for selecting an alternative in any decision is the state that maximizes the priority value.

We model each of these sequential decisions as a single attribute decision under certainty (Keeney & Raiffa, 1993). Because uncertainty exists in the estimate of the states' populations, these decisions contain attribute value uncertainty. Rather than considering approaches for adjusting the population estimates, we used the methods presented in this chapter to propagate the uncertainty in the population estimates and select a state to assign the Congressional seat for each of the 385 selection decisions.

3.5.2 Census Uncertainty

For more than three decades, there has been debate – often fierce – over the inclusion of statistical adjustments to the census enumeration. One basis for the proposed incorporation of statistical adjustment is that the census enumeration is not accurate and precise but is rather an uncertain estimate of the true population that statistical procedures may improve

(Anderson, et al., 2000). The Census Act and a 1999 Supreme Court ruling prohibit the use of statistical sampling for alternate counts in the reapportionment population. It should be noted that statistical sampling procedures are accepted for non-apportionment purposes such as unemployment and housing estimates.

Over the past 50 years the U.S. Census Bureau has conducted programs as part of the decennial census to evaluate the degree of the census coverage error. Programs in the 1990 and 2000 Censuses were designed to produce alternative counts based on the measurement of net coverage area. As a result of the 1999 Supreme Court ruling prohibiting the use of statistical sampling in determining the population for apportionment purposes, the goal of the 2010 program, referred to as the Census Coverage Measurement (CCM) survey, was to focus on improving the process for subsequent Censuses (Committee on National Statistics, 2009). While the names and year-to-year goals may differ, the census coverage programs all aim to capture some measure of uncertainty in the population estimates.

The 2010 state level apportionment population and results are provided in Table 1 of the U.S. Census report Congressional Apportionment (Burnett, 2011). As previously mentioned, the apportionment population includes the resident population of the 50 states plus overseas U.S. federal employees (military and civilian) and their children. No measures of uncertainty are provided with the apportionment population values.

The results of the 2010 Census Coverage Measurement program are provided in a series of twelve reports published by the U.S. Census Bureau. Two of the CCM reports (Mule, 2012; Davis & Mulligan, 2012) provide state level population estimates along with a measure of uncertainty in the form of a root mean squared error. (The root mean squared error uncertainty measures for the states ranged from 0.73 % to 2.43 %.) The CCM reports'

population estimates do not include overseas U.S. federal employees, residents of remote Alaska, or persons living in group quarters. Thus the population estimates are not useful for reapportionment purposes. However, we assumed that the relative uncertainties provided in the CCM reports represented the uncertainties of the apportionment population values and used these to model the uncertainty in the population estimates, and in turn, the uncertainty in the priority values (the decision value used for seat assignment).

Specifically, we used a parametric bootstrap approach (Efron & Tibshirani, 1993) to model the apportionment population uncertainty for each state and each sequential seat assignment decision. For each state, we modeled its population as a normally distributed random variable, with a mean equal to its apportionment population and a standard deviation equal to that state's CCM relative root mean squared error multiplied by the apportionment population. For a given seat assignment decision we generated 10,000 realizations from these apportionment population models (each realization included a population value for every state) and calculated the corresponding priority values as defined in Equation (3.6). For example, the apportionment population of Maryland, the 20th state alphabetically, was modeled by a $N(\mu_{20}, \sigma_{20})$ probability distribution where $\mu_{20} = 5,789,929$ (reported apportionment population) and $\sigma_{20} = 0.0119\mu_{20} = 68,900$ (root mean squared error of 1.19 %). A single realization from this distribution, 5,841,425, provides the priority value, according to Equation (3.6), for the 51st seat assignment where the current number of seats assigned, $s_{20,51} = 1$, of $v_{20,51} = 5,841,425 \Big/ \sqrt{(2)} = 4,130,511$. This sampling process was replicated 10,000 times for each of the 50 states, producing a distribution of 10,000 priority values for each state.

The multinomial selection procedure (Section 3.4.2) was used to select the state to which the Congressional seat would be assigned. That is, the seat was assigned to the state that most often had the largest priority value. This process was repeated 385 times until all 435 House of Representative seats were assigned.

3.5.3 Results

When incorporating census uncertainty into the sequential seat assignment decisions, the uncertainty in the population estimates had little impact on some seat assignment decisions, but the impact was more substantial on others. Consider the distribution of priority values obtained during the assignment of seat 51 (Figure 3.2). For each state, the 10,000 priority values obtained by sampling the apportionment population models are displayed by the mostly overlapped plotting characters. The single vertical hash for each state represents the single priority value obtained based on the published apportionment population value. In this decision there is little question that California has the largest priority value and should be assigned the 51^{st} seat.

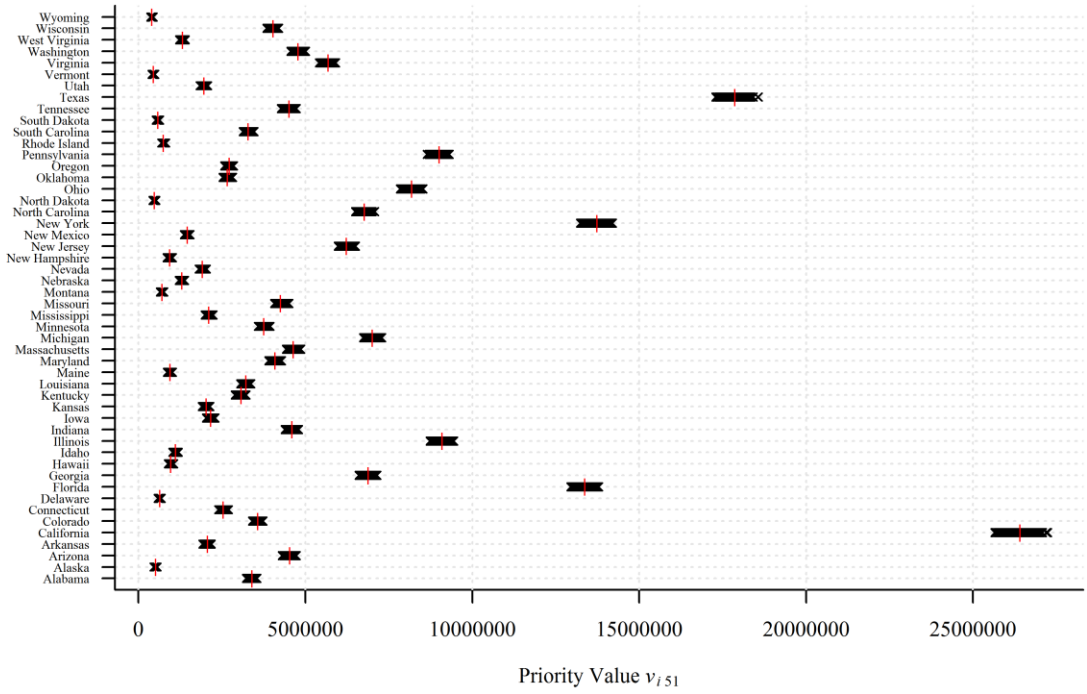

Figure 3.2: Priority values for the assignment of seat 51. The states are arranged in alphabetical order from bottom to top along the y-axis. The distributions of points are a result of the uncertainty propagation. The vertical hashes are the priority values that result from the apportionment population value.

On the other hand, consider the distribution of priority values obtained during the assignment of seat 384 (Figure 3.3). Here the priority values obtained from the apportionment population (vertical hashes) for a number of states are nearly equal. The propagation of the population uncertainty leads to priority value distributions that are difficult to distinguish from one another, and it is not clear which state should be assigned this seat. In this situation a multinomial selection procedure is necessary to aid in the decision.

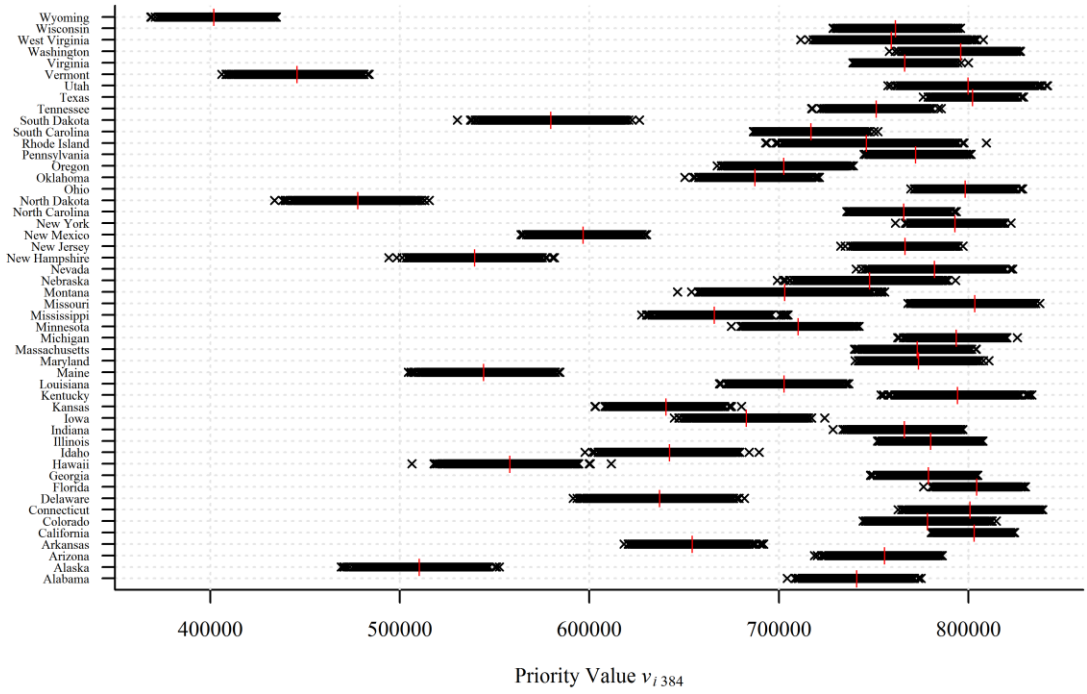

Figure 3.3: Priority values for the assignment of seat 384. The states are arranged in alphabetical order from bottom to top along the y-axis. The distributions of points are a result of the uncertainty propagation. The vertical hashes are the priority values that result from the apportionment population value.

In each randomly generated realization, the multinomial selection procedure identifies the state with the maximum priority value, and each state's multinomial probability is estimated by the proportion of realizations in which it had the maximum priority value. The state with the largest estimated multinomial probability value is selected (assigned the seat). When one state clearly holds the maximum priority value, as California did for seat 51 (Figure 3.2), that state is identified as having the maximum priority value in all of the realizations (that is, its estimated multinomial probability value is 1). When it is less obvious which state holds the maximum priority value, as it was for Seat 384 (Figure 3.3), there may be multiple states with nearly equal multinomial probability values.

The states' estimated multinomial probability values for each of the 385 seat assignment decisions are displayed in the stacked bar chart of Figure 3.4. The bars are sorted by the estimated multinomial probability values. For each seat assignment decision, the height of the bottom (most lightly shaded) bar represents the estimated multinomial probability value for the selected state. The height of the next bar represents the estimated multinomial probability value for the state with the second largest estimated multinomial probability value, and so on.

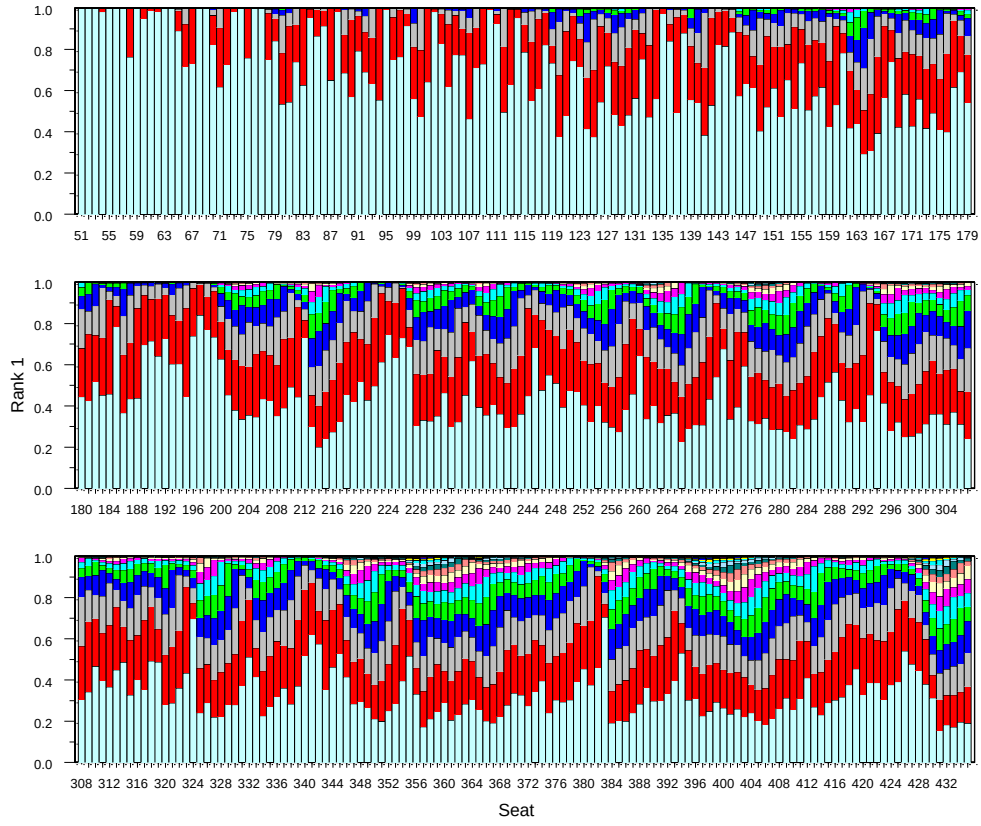

Figure 3.4: Estimated multinomial probability (Rank 1) values for each state in the 385 seat assignment decisions. For each seat decision, the estimated multinomial probability value for the selected state is displayed by the height of the bottom bar. The states with the second and subsequent largest estimated multinomial probability values are displayed by the additional stacked bars.

Two trends are observed in Figure 3.4. First, the estimated multinomial probability values for the selected states are generally smaller and more nearly equal to the other estimated multinomial probability values in the later seat assignment decisions. This reflects a desirable property of apportionment: as more seats are assigned, the differences in the states' representativeness (as measured by the priority values) approaches zero.

Within the general trend of decreasing estimated multinomial probability values for the selected states, we observe the second trend: the estimated multinomial probability values of these selected states occasionally increase dramatically. This often occurs when there are several states that have nearly equal priority values, and there is a significant gap between the priority values of these states and the next-largest priority value. The estimated multinomial probability values for these states will be nearly equal (and much less than 1). The priority value of a state decreases after it is awarded a seat. Thus, after most of these states have been awarded seats (and their priority values decreased), only one of these states remains, and its priority value is significantly larger than any other state, which will lead to a large estimated multinomial probability value for that state.

The complete apportionment (the number of seats assigned to each state) that was generated by the uncertainty propagation approach was exactly the same as the official 2010 Congressional reapportionment that was applied to the 113th Congress. (This also occurred when the approach was applied to the 2000 Census.) The sequence of seat assignments, however, was not the same. For example, in the 2010 apportionment, seat 176 was assigned to New York, and seat 177 was assigned to Arizona. Our approach assigned seat 176 to Arizona and seat 177 to New York.

A total of 77 differences occurred. Of these, 46 differences occurred in 23 pairs like the example. The other discrepancies occurred in longer sequences of seat assignment decisions. The longest had six seats and four differences.

Although the complete apportionments were equal, the observed differences in seat assignment sequences shows that the two approaches could produce differing apportionments. For example, if the House of Representations had only 433 seats, Texas would have a total of 35 seats using the apportionment populations (which assigned seat 433 to Texas). Because the uncertainty propagation approach assigned seat 433 to Minnesota, however, that state would take a seat from Texas (which would then have only 34).

The decision analysis approach incorporated population estimate uncertainty into the apportionment procedure without directly adjusting the apportionment population values. Such an approach could satisfy some critics of the current apportionment process, which prohibits the use of statistical sampling to adjust values in the apportionment population.

3.5.4 Summary

This example of the Congressional reapportionment process illustrated the idea of incorporating attribute value uncertainty (states' uncertain population estimates) in a selection decision. The reapportionment process was modeled as 385 single attribute selection decisions with attribute value uncertainty. The uncertainty model for the attribute values was based on a parametric bootstrap approach with parameter estimates provided by the apportionment population and the CCM program's uncertainty measure. Using Monte Carlo simulation, the population uncertainties were propagated to the priority values (decision

values). Given the relatively simple functional form of the decision value and that the attribute values were modeled as Gaussian random variables, an analytical approach could have been implemented to propagate the uncertainty. We implemented the multinomial selection procedure using Monte Carlo simulation to estimate the multinomial probabilities and select an alternative.

3.6 Application to Consumer Services

In this section, we expand on the "Selecting Consumer Services Based on Ratings" example introduced in Section 1.1.2. We illustrate the application of methods presented in this chapter for propagating attribute value uncertainty and selecting an alternative.

3.6.1 Selecting a Roofer

When homeowners require a repairman or other services, they often seek reviews and recommendations for potential service providers. The Spring/Summer 2011 edition of the Washington Consumers' Checkbook (Center for the Study of Services, 2011) provided an extensive review of roofing firms in the Washington, D.C., metropolitan area. We considered the problem of selecting a roofing firm using the data in the Washington Consumers' Checkbook to illustrate the presence of attribute value uncertainty in a multiple attribute selection decision.

The Washington Consumers' Checkbook review included ten performance rating criteria for 94 roofing firms obtained through a survey of the organization's members. Be-

cause the review is based on only a limited survey sample, the review's performance criteria values are uncertain estimates of the true attribute values that describe each roofer. The consumer who wishes to pick a roofer for a job using these survey results to inform this decision faces a multiple attribute selection decision with attribute value uncertainty. To demonstrate the methods presented in this chapter of incorporating attribute value uncertainty into a multiple attribute selection decision, we formulated a roofer selection decision with the following four performance rating criteria (attributes) that were measured in the survey:

1. Work performed properly on first attempt
2. Began and completed work promptly
3. Neatness of work
4. Percent of $5,000 job the firm allows the customer to pay upon completion

The survey results for the performance rating criteria were used as estimates of the true attribute values. Attributes 1 – 3 are the probability of a "superior" rating with estimated values provided as the proportion of customers surveyed who rated the firm "superior". These attribute value measures are random variables consisting of a collection of Bernoulli trials: the performance criterion was rated by each survey respondent as either superior or not. Thus, for each roofing firm (alternative) $i = 1, \ldots, 87$ (seven firms were removed from the analysis due to incomplete data), the data supporting attribute j = 1, 2, 3 can be modeled by a binomial random variable with parameters μ_{ij} and n_i, where μ_{ij} is the probability of obtaining a rating of "superior" and n_i is the number of survey responses. For these attributes, a larger value is preferred. The fourth attribute is not random; it is provided by the roofing firm and thus for each firm, its value is considered to be a constant. Larger values

are preferred because, if the value is small, the customer must pay more upfront, which increases the customer's financial risk. Summary statistics of the distribution for the four attributes and the number of survey responses across the $m = 87$ firms considered are provided in Table 3.2.

Table 3.2: Summary statistics for the distribution of data across the 87 roofing firms.

	Mean	Std Dev	Min	Median	Max
Number of survey responses	54.17	66.30	10	29	390
Work performed properly	0.74	0.16	0.23	0.79	1.00
Began and completed promptly	0.74	0.16	0.28	0.77	1.00
Neatness of work	0.76	0.16	0.27	0.79	1.00
Percent paid after completion	0.77	0.19	0.33	0.67	1.00

A multiple attribute measurable value function was used to represent the decision-maker's (the author's) preference structure. We assumed the preference structure is such that attributes are mutually preference independent and mutually difference independent. Therefore, the multiple attribute measureable value function can be represented by the sum of single attribute measureable value functions (Dyer & Sarin, 1979), which provides the decision value of each alternative, ξ_i, as displayed in Equation (3.7).

$$\xi_i = v\left(\mu_{i1}, \mu_{i2}, \mu_{i3}, \mu_{i4}\right) = \sum_{j=1}^{4} \lambda_j v_j\left(\mu_{ij}\right) \tag{3.7}$$

Here $\sum_{j=1}^{4} \lambda_j = 1$ and the individual measureable value functions $v_j\left(\mu_{ij}\right)$ are scaled such that, for μ_{ij}^*, the most preferred outcome overall all i for a fixed j, $v_j\left(\mu_{ij}^*\right) = 1$ and, for μ_{ij}^0, the least preferred outcome overall all i for a fixed j, $v_j\left(\mu_{ij}^0\right) = 0$.

3.6.2 Uncertain Attribute and Decision Value Estimates

One may employ the expected performance rating values published as the survey results in conjunction with Equation (3.7) to obtain a decision value for each roofing firm. As described by Keeney and Raiffa (1993), the alternative with the largest resulting decision value would be considered to be the alternative most fitting given the decision-maker's preferences. We refer to such an approach as a *deterministic selection*.

To implement the deterministic selection approach, we developed the individual measureable value functions $v_j(\mu_{ij})$, $j = 1,\ldots,4$ using an augmentation to the midvalue splitting technique that leverages an analytical exponential form (Kirkwood, 1997) based on the attribute value ranges displayed in Table 3.2. The swing weighting procedure (Clemen & Reilly, 2001) was used to obtain the attribute decision weights, $\lambda_1,\ldots,\lambda_4$. The individual measureable value functions and associated weights are provided in Table 3.3.

Table 3.3: Individual measureable value functions and decision weights for deterministic selection.

Attribute	$v_j(\mu_{ij})$	λ_j
Work performed properly	$v_1(\mu_{i1}) = -\frac{1}{3.634}\left(1-e^{(\mu_{i1}-0.23)/0.502}\right)$	0.476
Began and completed promptly	$v_2(\mu_{i2}) = 1-e^{-(\mu_{i2}-0.28)/0.0866}$	0.190
Neatness of work	$v_3(\mu_{i3}) = \mu_{i3}/0.73-0.37$	0.286
Percent paid after completion	$v_4(\mu_{i4}) = \frac{1}{0.909}\left(1-e^{-(\mu_{i4}-0.33)/0.280}\right)$	0.048

Based on these defined individual measureable functions and associated weights, Equation (3.7) was evaluated for each alternative with results displayed graphically in Figure

3.5. Roofing Firm 29, whose decision value equals 0.9837, is identified as the most desirable alternative. This firm is followed closely by Roofing Firm 84, whose decision value equals 0.9835.

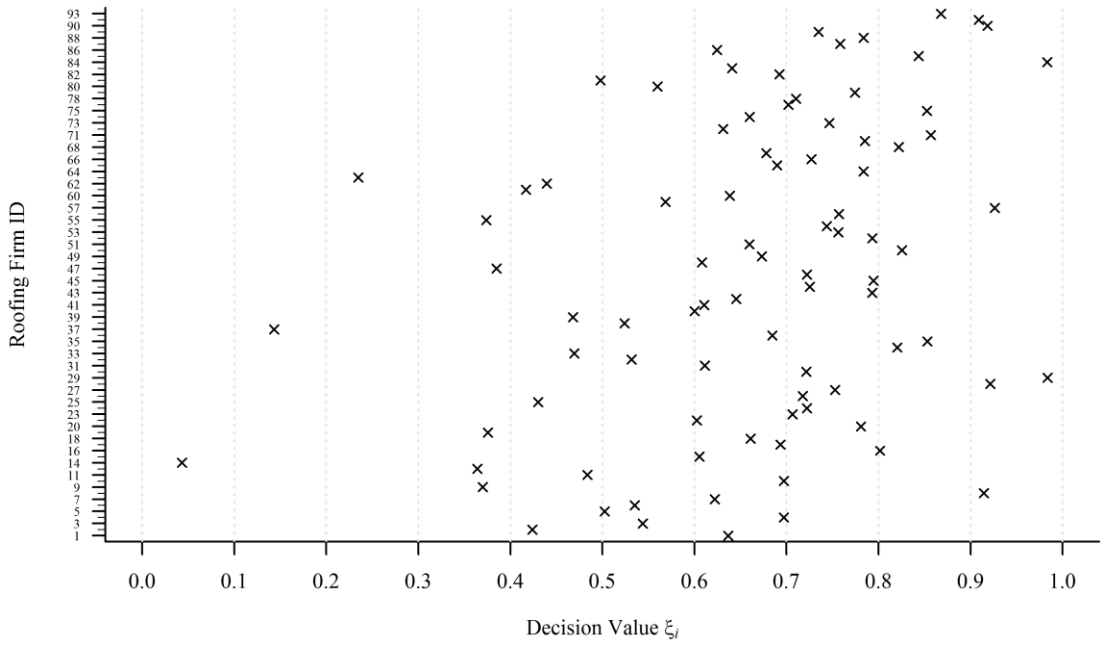

Figure 3.5: Roofing firm decision model results using the deterministic selection approach.

To describe the uncertainty in attributes 1 – 3, we used Bayesian posterior distributions. For each of the 87 alternatives and each attribute j = 1, 2, 3, we began with the assumption (prior knowledge) that the true value of the attribute, μ_{ij}, lies between 0 and 1 with equal likelihood. This is represented by the $Uniform[0,1]$ prior probability distribution, which is equivalent to a $Beta(1,1)$ probability distribution. Observations, i.e., the number of "superior" ratings recorded by the survey (modeled by a $Binomial(n_i, \mu_{ij})$ probability distribution), for each performance criteria j = 1, 2, 3 for each of the 87 roofing firms were used to

update the prior distributions. Because the $Beta(\alpha,\beta)$ distribution is the conjugate prior distribution to the $Binomial(n,\pi)$ distribution, upon observing the survey results, our knowledge of the true attribute value is updated and described by the $Beta\left(1+n_i\hat{p}_{ij},1+n_i\left(1-\hat{p}_{ij}\right)\right)$ posterior distribution where $\hat{p}_{ij}$ is the rating value provided for performance criteria *j* and roofing firm *i* and n_i is the number of survey responses.

Given the posterior distributions for each attribute of each alternative, we drew 1000 random samples from each of these distributions. Based on the distributions of the random realizations for each attribute across all alternatives, we redefined the individual measureable value functions $v_j(\mu_{ij}),\ j=1,2,3$ and attribute decision weights, again using an augmentation to the midvalue splitting technique and the swing weighting procedure. Since attribute 4 was considered a constant having no uncertainty, its individual value function as provided by the deterministic selection analysis was unchanged. The updated individual measureable value functions and associated weights are provided in Table 3.4.

Table 3.4: Individual measureable value functions and decision weights when considering attribute value uncertainty.

Attribute	$v_j(\mu_{ij})$	λ_j
Work performed properly	$v_1(\mu_{i1})=-\frac{1}{4.53}\left(1-e^{(\mu_{i1}-0.027)/0.570}\right)$	0.526
Began and completed promptly	$v_2(\mu_{i2})=\frac{1}{0.954}\left(1-e^{-(\mu_{i2}-0.050)/0.309}\right)$	0.158
Neatness of work	$v_3(\mu_{i3})=\mu_{i3}/0.951-0.05$	0.263
Percent paid after completion	$v_4(\mu_{i4})=\frac{1}{0.909}\left(1-e^{-(\mu_{i4}-0.33)/0.280}\right)$	0.053

Based on these updated individual measureable functions and associated decision weights, Equation (3.7) was evaluated for each alternative for each of the 1000 random

realizations (and constant value of attribute 4). The result was a distribution of 1000 decision values for each roofing firm. We found the minimum value of each of these 87 decision values distributions and identified the greatest of these 87 minimum values. Fifty-nine alternatives were observed to provide a maximum decision value that was less than the greatest minimum value and were thus determined to be dominated. The decision value distributions for the 28 non-dominated alternatives are displayed in Figure 3.6.

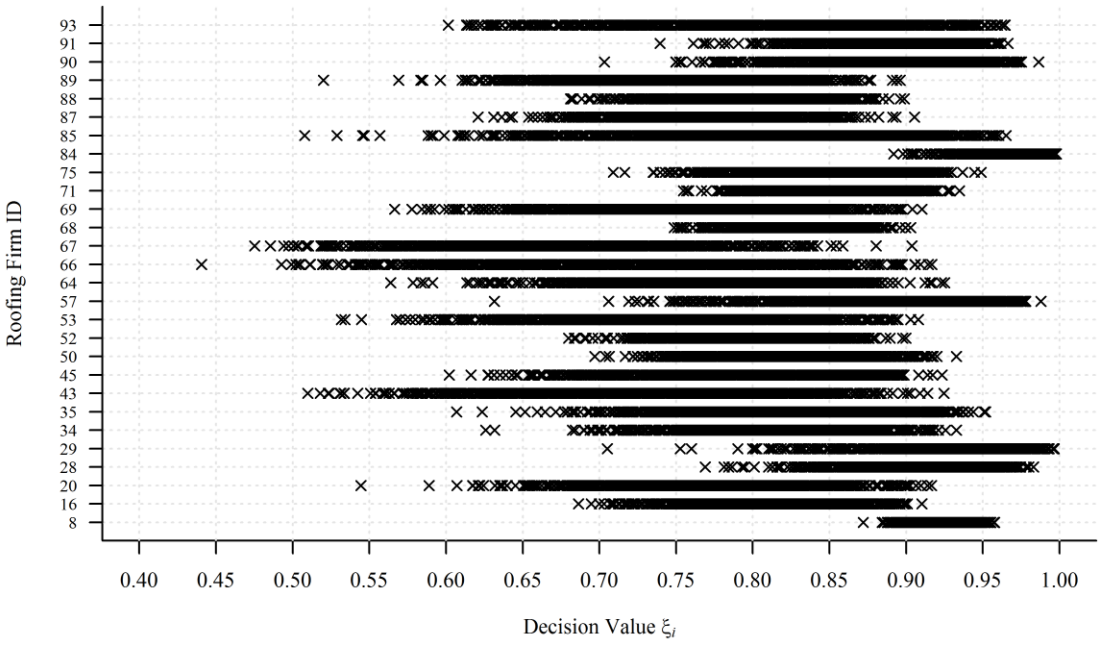

Figure 3.6: Decision value distributions for the 28 non-dominated alternatives.

3.6.3 Results

We applied the multinomial, stochastic dominance, and majority judgment selection procedures to the distributions of decision values for the non-dominated roofing firms. Using

the multinomial and majority judgment procedures, we identified the alternative to be selected first. By removing that alternative from consideration, we were able to identify the alternative to be selected second, and so on. This enabled us to provide a rank order of alternatives. Table 3.5 displays the rank order as provided by the multinomial and majority judgment selection procedures for the firms that had the six largest decision values as provided by the deterministic selection approach. The results for the stochastic dominance decision procedure are best displayed graphically as empirical cumulative distribution curves, which are displayed in Figure 3.7 for the top roofing firms.

Table 3.5: The rank of the top six roofing firms as provided by the deterministic, multinomial, and majority judgement selection procedures.

Roofing Firm	*n*	Deterministic	Multinomial	Majority Judgement
Firm 29	24	1	2	2
Firm 84	82	2	1	1
Firm 57	23	3	3	5
Firm 28	54	4	4	4
Firm 90	36	5	5	6
Firm 8	347	6	6	3

From Table 3.5 we see that the multinomial and majority judgement selection procedures, which incorporate the attribute value uncertainty, identify Roofing Firm 84 as the most desirable alternative with Roofing Firm 29 identified as the second most desirable alternative. The deterministic selection approach flips this order by identifying Roofing Firm 29 as the most desirable alternative (deterministic decision value of 0.9837) and Roofing Firm 84 as the second most desirable (deterministic decision value of 0.9835). In 1000 Monte Carlo simulation realizations, Roofing Firm 84 was identified as the most desirable option 74 % of the time, and Roofing Firm 29 was the most desirable option only 17.5 %

of the time (these values are interpreted as estimates of the p_i in the multinomial selection procedure). The majority-grade in the majority judgment selection procedure for Roofing Firm 84 was 0.9724, while the majority-grade for Roofing Firm 29 was 0.9426.

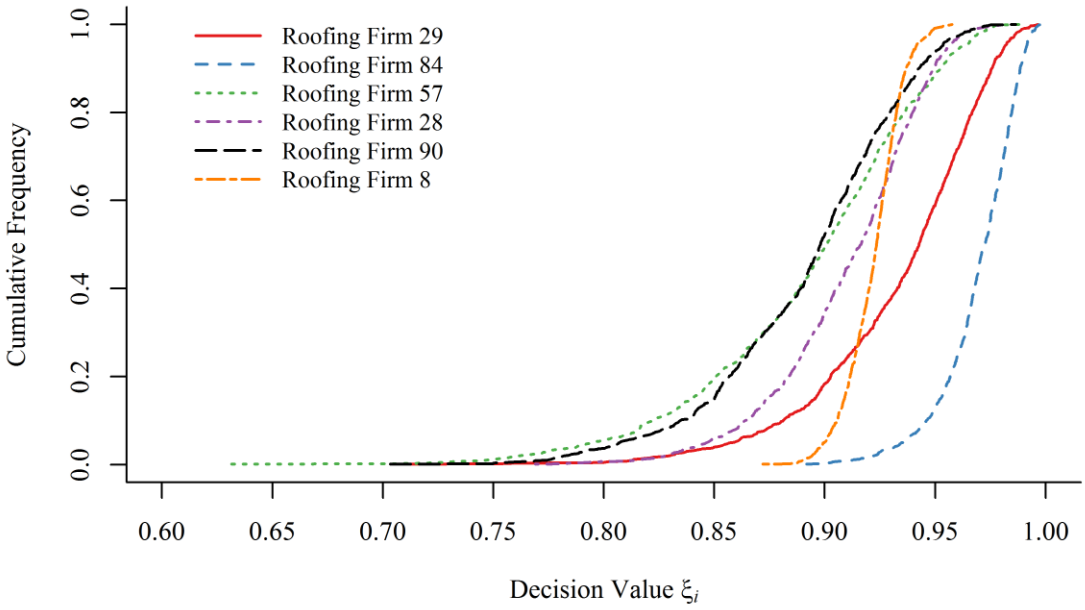

Figure 3.7: Empirical cumulative distribution curves for the top six roofing firms.

As shown in Figure 3.7, the empirical cumulative distribution curve for Roofing Firm 84 never exceeds that of any other alternative, so Roofing Firm 84 stochastically dominates all other alternatives in the first degree. Thus, Roofing Firm 84 is deemed to be the most desirable alternative using the stochastic dominance decision procedure. This result is consistent with the results obtained by the other decision procedures that consider the attribute value uncertainty. Further, the fact that we found one alternative that stochastically dominates all of the other alternatives in the first degree is an extremely powerful result, as

Hadar and Russell (1969) have shown that this is the decision-maker's most preferred alternative regardless of his underlying utility function.

3.6.4 Summary

This example of selecting a roofing firm based on performance criteria estimates obtained through a customer survey illustrated how attribute value uncertainty may be incorporated into a multiple attribute selection decision model and how an alternative may be selected using the selection approaches discussed in Section 3.4. The multiple uncertain attributes were viewed as success probabilities and modeled using Bayesian posterior probability distributions. Monte Carlo simulation was used to propagate the attribute value uncertainty to the decision value. We applied the deterministic, multinomial, majority judgement, and stochastic dominance selection procedures to produce a rank ordering of the roofing firms. We observed that the selection procedures that explicitly consider the attribute value uncertainty (multinomial, majority judgement, and stochastic dominance) all identified Roofing Firm 84 as the most preferred alternative. The deterministic selection procedure, which does not consider the uncertainty in the estimation of the true attribute values, identified Roofing Firm 24 as the most preferred. While we are unable to definitively state which roofing firm is truly the most preferred, this observation illustrates that ignoring attribute value uncertainty in a selection decision does indeed have an impact.

3.7 Application to Radiation Detection

In this section, we expand on the "Selecting a Radiological and Nuclear Detection System" example introduced in Section 1.1.3. We illustrate the application of methods presented in this chapter for propagating attribute value uncertainty and selecting an alternative.

3.7.1 Radiation and Nuclear Detection System Selection Model

To identify a radiation and nuclear detection system from a set of 576 candidate systems to put forth for an operational demonstration in the PaxBag pilot program, the Domestic Nuclear Detection Office (DNDO) and the U.S. Customs and Border Protection (CBP) agencies formulated a multiple attribute decision model consisting of $k = 26$ attributes (Equation (3.8)) to determine the decision values, ξ_i, of each system $i = 1, \ldots, 576$.

$$\xi_i = v(\mu_{i1}, \ldots, \mu_{i26}) = \sum_{j=1}^{26} \lambda_j v_j(\mu_{ij}) \tag{3.8}$$

In Equation (3.8), μ_{ij} is the true value for attribute *j* of alternative a_i. $\sum_{j=1}^{26} \lambda_j = 1$ and the individual value functions $v_j(\mu_{ij})$ are scaled such that, for μ_{ij}^*, the most preferred outcome overall all *i* for a fixed *j*, $v_j(\mu_{ij}^*) = 1$ and, for μ_{ij}^0, the least preferred outcome overall all *i* for a fixed *j*, $v_j(\mu_{ij}^0) = 0$. The individual value functions were defined to be linear, i.e., $v_j(\mu_{ij}) = (\mu_{ij} - \mu_{ij}^0) / (\mu_{ij}^* - \mu_{ij}^0)$.

3.7.2 Uncertain Attribute and Decision Value Estimates

The 26 attributes and their decision weights, λ_j, are provided in Table 3.6. Only the interdiction performance attributes, attributes 1 – 22, were studied in this research because their values were success probabilities that were estimated using the results of laboratory evaluations. We note from Table 3.6 that the attribute decision weights differ across the interdiction performance attributes. This reflects the different level of importance that DNDO and CBP place on the ability of a system to interdict the different sources.

Table 3.6: Decision model attributes and decision weights for the selection of a rad/nuc detection system.

Attribute	λ_j	Attribute	λ_j
1. Inderdict Source 1 - Passenger	0.022500	14. Inderdict Source 3 - Baggage	0.022500
2. Inderdict Source 2 - Passenger	0.022500	15. Inderdict Source 4 - Baggage	0.022500
3. Inderdict Source 3 - Passenger	0.022500	16. Inderdict Source 5 - Baggage	0.022500
4. Inderdict Source 4 - Passenger	0.022500	17. Inderdict Source 6 - Baggage	0.022500
5. Inderdict Source 5 - Passenger	0.022500	18. Inderdict Source 7 - Baggage	0.022500
6. Inderdict Source 6 - Passenger	0.022500	19. Inderdict Source 8 - Baggage	0.016875
7. Inderdict Source 7 - Passenger	0.022500	20. Inderdict Source 9 - Baggage	0.016875
8. Inderdict Source 8 - Passenger	0.016875	21. Inderdict Source 10 - Baggage	0.016875
9. Inderdict Source 9 - Passenger	0.016875	22. Inderdict Source 11 - Baggage	0.016875
10. Inderdict Source 10 - Passenger	0.016875	23. Passenger Delay	0.202500
11. Inderdict Source 11 - Passenger	0.016875	24. Additional CBP Officers	0.202500
12. Inderdict Source 1 - Baggage	0.022500	25. System Integration	0.045000
13. Inderdict Source 2 - Baggage	0.022500	26. Cost	0.100000

The data collected in the laboratory evaluations were the number of times a sensor technology was successful at interdicting the source of interest under the various operational modes. These data were modeled with a $Binomial\left(n_{cd}, \pi_{cd}\right)$ probability distribution,

where n_{cd} is the number of observations made for sensor technology and operational combination *c* against source *d* and π_{cd} is the probability that combination *c* will successfully interdict source *d*. Based upon this data, attribute value estimates (probabilities) were formulated for each system by taking into account the fact that passengers will go through both detection subsystems, while checked baggage will go through only the second subsystem. As these probabilities were based on experimental evaluations from a limited number of trials, they were only estimates for the detection system's true capabilities.

We used Bayesian posterior distributions to describe the uncertain values of the interdiction performance attributes. We began with the assumption (prior knowledge) that the true probability, π_{cd}, that sensor technology and operational combination *c* will successfully detect and identify source *d* lies between 0 and 1 with equal likelihood. This is represented by the $Uniform[0,1]$ prior probability distribution, which is equivalent to a $Beta(1,1)$ probability distribution. The data from the laboratory evaluations were used to update the prior distributions. Because the $Beta(\alpha,\beta)$ distribution is the conjugate prior distribution to the $Binomial(n,\pi)$ distribution, upon observing the experimental evaluation results, our knowledge of π_{cd} is updated and presented by the $Beta(1+x_{cd}, 1+n_{cd}-x_{cd})$ posterior distribution where x_{cd} is the number of times that sensor technology and operational combination *c* was successful at interdicting source *d*. Due to its sensitive nature, we cannot present the experimental evaluation results.

Using Monte Carlo simulation we propagated the uncertainty in the π_{cd} to the attribute values, μ_{ij}, $j = 1, \ldots, 26$, and ultimately to the decision values ξ_i for each of the 576 candidate radiation and nuclear detection systems. Specifically, we drew 1000 random samples from the posterior distribution for π_{cd}. For each of these 1000 realized samples, we calculated a realization of μ_{ij}, for attributes 1 – 22, using the required attribute formulations. For each of the attribute value realizations, together with the constant values provided for attributes 23 – 26, we computed ξ_i using Equation (3.8). This resulted in a distribution (1000 realizations) of ξ_i values for each of the 576 candidate radiation and nuclear detection systems.

We found the minimum value of each of the 576 decision values distributions and identified the greatest of these 576 minimum values. We then determined that 540 alternatives were dominated by observing that the maximum decision value provided for these 540 alternatives was less than the greatest minimum value. The decision value distributions for the remaining 36 non-dominated alternatives are displayed in Figure 3.8.

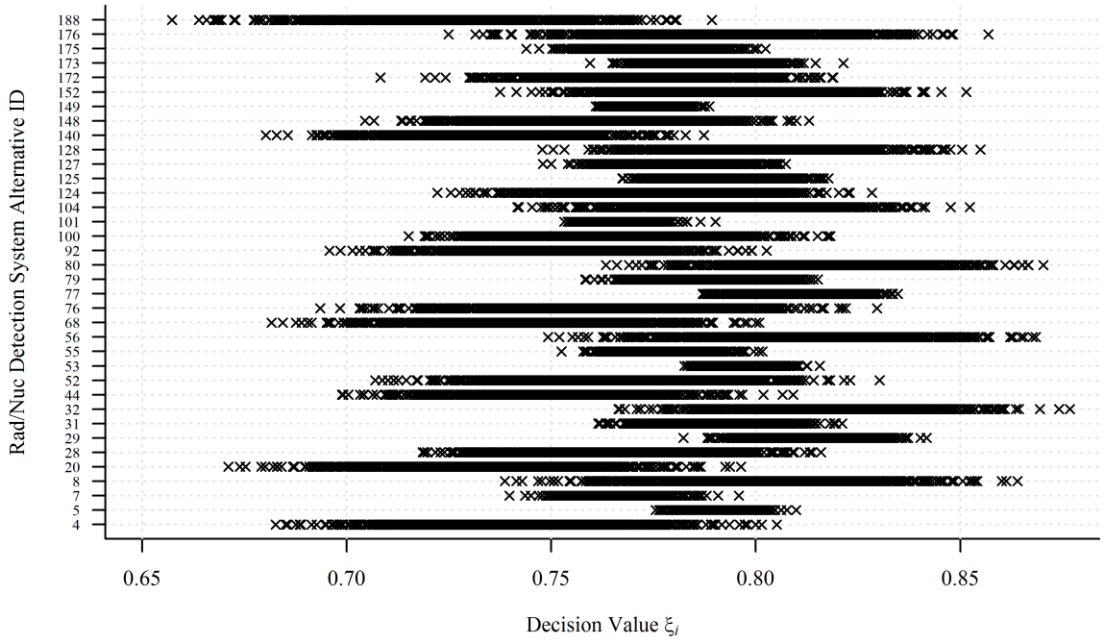

Figure 3.8: Decision value distributions for the 36 non-dominated alternatives.

3.7.3 Results

We applied the multinomial, stochastic dominance, and majority judgment selection procedures to the distributions of decision values for the non-dominated detection systems. Using the multinomial and majority judgment selection procedures, we identified the alternative to be selected first. By removing that alternative from consideration, we were able to identify the alternative to be selected second, and so on. This enabled us to provide a rank order of alternatives. Under the DNDO and CBP preference structure, the order of the top 6 most desirable alternatives provided by both the multinomial and the majority judgment selection procedures are identical. These alternatives are provided in Table 3.7. The empirical cumulative distribution functions of the decision values for these six alternatives

are displayed in Figure 3.9. Among these six, no alternative stochastically dominates all others.

Table 3.7: Six top ranked detection systems according to the multinomial and majority judgment selection procedures.

Rank	Alternative ID	Sensor Tecnology & Placement	Operational Mode
1	80	PMPP 1 in Passport booth	dwell: 30 s
		PMPP 2 in Customs booth	dwell: 30 s
2	32	PMPP 1 in Passport booth	dwell: 2 min
		PMPP 2 in Customs booth	dwell: 30 s
3	29	PMPP 1 in Passport booth	dwell: 2 min
		PMPP 2 in Customs booth	walk: 0.5 m/s
4	77	PMPP 1 in Passport booth	dwell: 30 s
		PMPP 2 in Customs booth	walk: 0.5 m/s
5	56	PMPP 1 in Passport booth	walk: 2.5 m/s
		PMPP 2 in Customs booth	dwell 30 s
6	128	PMPP 2 in Passport booth	dwell: 2 min
		PMPP 2 in Customs booth	dwell: 30 s

Both the multinomial and majority judgment selection procedures identify alternative 80 as the most desirable followed by alternative 32. In the 1000 Monte Carlo simulation realizations, alternative 80 was identified as the most desirable option 33 % of the time, and alternative 32 was the most desirable option 27 % of the time (these values are interpreted as estimates of p_i in the multinomial selection procedure). The majority-grade in the majority judgment selection procedure for alternative 80 was 0.819, while the majority-grade for alternative 32 was 0.818.

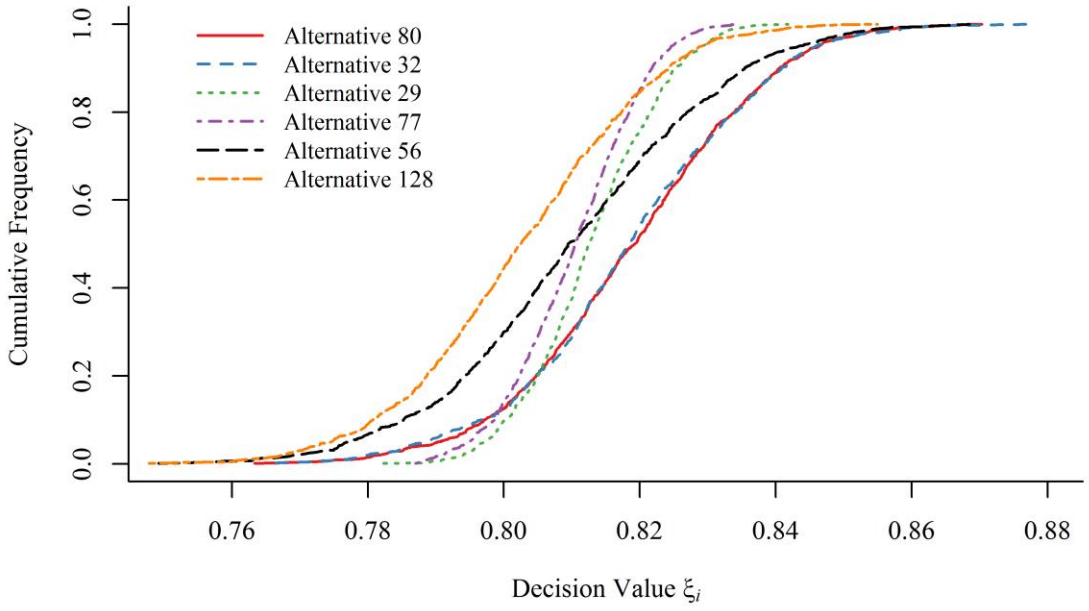

Figure 3.9: Empirical cumulative distributions of the decision values for the six top ranked detection systems according to the multinomial and majority judgment selection procedures.

From the empirical cumulative distributions of the decision values displayed in Figure 3.9, it is seen that the two top ranked alternatives (80 and 32) have nearly identical distributions of decision values. Because they do not stochastically dominate all other alternatives, the decision-makers are faced with making a tradeoff between these alternatives, which have greater uncertainty in their decision values, and alternatives 29 and 77, which have less uncertainty. In this case, the decision value empirical cumulative distribution curves of alternatives 80 and 32 fall largely to the right of all other alternatives.

The DNDO and CBP decision-makers should take comfort in the fact that the top five alternatives displayed in Table 1 utilize the same sensor technologies (PMPP 1 and PMPP 2) and sensor placements (within booths). The only difference between these alternatives is the operational modes. Further, of the 1000 Monte Carlo simulation realizations, one of

these top five alternatives was identified to be the most desirable in the multinomial selection procedure 998 times.

3.7.4 Summary

In this example we illustrated how the DNDO and CBP decision-makers could incorporate uncertain performance criteria estimates obtained through a limited number of laboratory evaluations in the multiple attribute selection decision model for selecting a radiation and nuclear detection system. Of the 26 attributes considered in the decision model, the values of 22 of the attributes were uncertain as they were estimated based on a limited number of laboratory evaluations. Using Bayesian posterior probability distributions and Monte Carlo simulation, we modeled and propagated the uncertainties resulting from the laboratory evaluations to the decision values. We applied the multinomial, majority judgement, and stochastic dominance selection procedures to produce a rank ordering of the candidate systems. The multinomial and majority judgement selection procedures provided identical rankings for the top six alternatives. The stochastic dominance procedure did not provide a selection as no alternative dominated all others.

This example provided the original motivation for this dissertation research because the experiment design used for the laboratory evaluation could have been altered to best support the multiple attribute selection decision. There was a set budget regarding the number of evaluations that could be performed. Provided the multiple attribute selection decision model, how should the limited laboratory experimental budget have been allocated amongst the alternatives and attributes to maximize the probability of the best system being selected? This question of sample allocation is the focus of Chapter 4.

3.8 Summary

This chapter provided general notation and assumptions for the multiple attribute selection decision model considered in this dissertation. We presented two methods, analytical and Monte Carlo simulation, which may be used to propagate attribute value uncertainty through the decision model to the decision value. By identifying and adapting existing methods for comparing multiple random variables, we formulated four selection procedures that may be applied to a multiple attribute selection decision with attribute value uncertainty. The deterministic selection procedure, which is often used in practice, selects the alternative with the largest decision value by evaluating the decision model using the expected attribute values. The multinomial selection procedure selects the alternative that has the greatest probability of having the largest decision value. The stochastic dominance selection procedure selects that alternative whose distribution of the decision value stochastically dominates, in the first- or second-degree, all other alternatives' decision value distributions. The majority judgement selection procedure selects the alternative with the largest majority-grade.

Through demonstrations in three different examples, we modeled the multiple attribute decision and the attribute value uncertainty, propagated the uncertainty to the decision values, and applied one or more of the above mentioned selection procedures to select an alternative. In the Congressional apportionment example, we illustrated how the uncertainty in the states' population estimates could be propagated to the states' priority values and a selection made using the multinomial selection procedure. We used data from a limited survey in the selection of a roofer example to illustrate a multiple attribute selection

decision with attribute value uncertainty. In this example the uncertainty in the performance attributes was due to the limited number of survey responses. We used the deterministic, multinomial, majority judgement, and stochastic dominance selections procedures. A single alternative was found to stochastically dominate all other in the first-degree. This was the same alternative selected by the multinomial and majority judgement selection procedures, but the deterministic selection procedure chose a different alternative. Our final example presented the decision to select a radiation and nuclear detection system to consider for installation in U.S. based international arrival airport terminals. In this example, the 22 uncertain performance measures, provided as success probabilities, for 576 alternative systems, were modeled using Bayesian posterior probability distributions and propagated to the decision value using Monte Carlo simulation. Here, the stochastic dominance selection procedure failed to identify an alternative for selection while the multinomial and majority judgment selection procedures agreed on the top six alternatives.

The selection of a radiation and nuclear detection system example was of particular relevance to the overall work of this dissertation because it presented a situation where the attribute values were estimated based on a limited number of laboratory evaluations. While the number of laboratory evaluations was equivalent for all attributes and alternatives, the allocation of the fixed experimental budget could have been allocated differently to each of the attributes and alternatives. The development and evaluation of sample allocation procedures across multiple attributes of multiple alternatives in a selection decision is the topic addressed in Chapter 4.

Of the selection procedures identified and illustrated in this chapter, the expected value and the majority-judgement selection procedures account for and propagate the uncertainties in the estimates of the attribute values, but then fail to consider the uncertainty in the decision values when selecting an alternative. The stochastic dominance selection procedure does consider the uncertainty in both the attribute and decision values and maintains other desirable features, however, a selection cannot be guaranteed using this procedure. Though Monte Carlo estimation is often required in its implementation, the multinomial selection procedure does directly consider the uncertainty in the attribute and decision values. Also, as we will later demonstrate, when a Bayesian estimation approach is used, a selection made using the multinomial selection procedure also maximizes the probability of making a correct selection. For these reasons, we relied on the multinomial selection procedure and Bayesian estimation as a basis for much of our sample allocation work in Chapter 4. By understanding how the attribute value uncertainty is modeled and used in the selection decision, we were able to focus our efforts in developing sample allocation procedures that improved the likelihood of selecting the true best alternative.

Chapter 4 Information Gathering

In this chapter we turn to the development and evaluation of budget-constrained sample allocation methods for selection decisions with attribute value uncertainty. When attribute values are estimated using sample data collected from measurements, experimental evaluations, or simulation runs, uncertainty associated with the estimates is present and relevant to the selection decision model. In particular, this attribute value uncertainty can limit the decision-maker's ability to identify the alternative that truly maximizes his value (utility) due to the uncertainty of the decision values. Since the decision-maker can reduce the amount of uncertainty associated with each attribute value by increasing the amount of information used in its assessment, the allocation of experimental effort (sample allocation) across the decision alternatives and attributes plays an important role in maximizing the probability of selecting the truly best alternative (the alternative that would have been selected had the true attribute and decision values been known).

In general, the uncertainty in the true attribute values can be reduced by increasing the number of samples used in their estimation, which, in turn, reduces the uncertainty in the decision values and conceivably increases the likelihood of making a correct selection. When the decision-maker is provided a fixed sample budget that must not be exceeded, where a sample measures only one attribute for one alternative at a time, the challenge becomes how to allocate this budget across the alternatives and across the attributes to provide the greatest probability of making the correct selection. If the sample 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 limited.

Section 4.1 expands on the Assumptions provided in Chapter 3 as necessary to formulate the budget-constrained sample allocation problem for selection decisions with multiple uncertain attributes. In Section 4.2 we introduce a simulation approach common to all of the studies presented in this chapter. The work presented in Section 4.3 applies to the Bernoulli measurement model where the sample observations are the results of a series of Bernoulli trials. More generally, the attribute values are success probabilities that are evaluated using pass-fail testing. From empirical studies we formulated and compared allocation procedures for single-stage experimental planning. In Section 4.4 we present work related to the Gaussian measurement model. Here, the attribute values are results from measurement processes that are assumed to contain Gaussian measurement error. This model is useful when the attribute values are physical quantities measured on a continuous scale such as the weight of an object. We formulated and evaluated allocation procedures for single-stage experimental planning from empirical studies (Sections 4.4 & 4.5) and using analytical methods (Section 4.6). The formulation and evaluation of a sequential sample allocation approach for the Gaussian measurement model is presented in Section 4.7. Chapter 4 closes with a summary of the allocation approaches developed in this dissertation in Section 4.8.

The discussions found in this chapter are largely taken from the works of Leber and Herrmann (2013a, 2014a, 2014b, 2015).

4.1 Problem Definition

We assume in the multiple attribute selection decision that the best alternative is the one that has the greatest decision value, which is a function of its true attribute values, and that it is the decision-maker's goal to select this best alternative. While true values for the multiple attributes exist for each alternative, they are unknown to the decision-maker and will be estimated through a series of experiments (measurements, evaluations, or simulation runs). 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 and ultimately the estimated decision values in supporting the selection decision. Provided this uncertainty, the decision-maker seeks to maximize the probability of selecting the most preferred alternative.

The information-gathering resource allocation problem can be stated as follows: The overall experimental budget in terms of sample observations, denoted B, shall not be exceeded and the cost of each observation is equivalent. Thus, B is the upper bound on the number of observations that can be collected. B is fixed and will be divided among the k

attributes of the *m* alternatives. Information is gained about each attribute through experimental (measurement) processes. The experimental processes are independent, and, for a given attribute, the same process is used to assess all alternatives. Let X_{ijl} be the independent random variables that describes the l^{th} outcome from the experimental processes used in estimating the value of the j^{th} attribute of alternative a_i. Realized sample observations are denoted $x_{ij1},\ldots,x_{ijn_{ij}}$, where n_{ij} is the number of measurements obtained for attribute *j* of alternative a_i, and the sample data observed in support of alternative a_i are $\mathbf{x}_i = x_{i11},\ldots x_{i1n_{i1}}, x_{i21},\ldots x_{i2n_{i2}},\ldots,x_{ik1},\ldots x_{ikn_{ik}}$. The problem is to find values n_{ij} that maximize the probability that the decision-maker will choose the truly best alternative (the probability of correct selection), given the decision-maker's preferences as modeled in the selection decision, such that $\sum_{i=1}^{m}\sum_{j=1}^{k} n_{ij} \le B$.

Provided the formulation of the information-gathering resource allocation problem, we restate and amend the assumptions provided in Section 3.2 as the basis for the development of the allocation procedures presented in this chapter.

1. The set of *m* distinct alternatives, $\{a_1,\ldots,a_m\}$, is provided, where *m* is a finite positive integer such that all alternatives can be assessed.
2. Each alternative is described by $k \ge 2$ attributes. The decision-maker's knowledge of the value of attribute *j* of alternative a_i is uncertain and is described by a probability distribution which depends on the attributes true value, μ_{ij}.
3. Also provided is a decision model, $\xi_i = f(\mu_{i1},\ldots,\mu_{ik}) = \sum_{j=1}^{k} \lambda_j v_j(\mu_{ij})$, that reflects the decision-maker's preference structure and combines the multiple attribute

values to produce a decision value, ξ_i, for each alternative a_i. The decision model is a multiple attribute linear value model with linear individual value functions, $v_j(\mu_{ij}) = \mu_{ij}$. The attribute decision weights, λ_j, are defined such that $\sum_{j=1}^{k} \lambda_j = 1$.

4. The total fixed experimental budget in terms of sample observations, denoted *B*, shall not be exceeded and the cost of each measurement is equivalent. Thus, *B* is the upper bound on the number of measurements that can be performed.
5. The decision-maker seeks to maximize the probability of selecting the true best alternative (*PCS*).

A Bayesian posterior distribution representation of the uncertain attribute values and uncertain decision values is the basis for many of the sample allocation procedures developed and evaluated in this chapter. The Bayesian approach to modeling uncertainty provides a natural avenue to incorporate prior knowledge. This feature was advantageous in our development of a sequential allocation approach where the decision-maker's knowledge is updated with each new observation. Further, the ability to recognize prior knowledge allows for a selection decision to be made based on an allocation where no additional information is collected for one or more attributes of one or more alternatives. We also assume in much of the work of this chapter that, when faced with uncertain decision values, the decision-maker prefers (and will select) the alternative that has the greatest probability of being the best (largest decision value) among the given set of alternatives, i.e., he uses the multinomial selection procedure.

This problem definition and set of assumptions hold in general for the work in this chapter with additional definitions and assumptions provided as needed in the sections to

follow. We close this section with a quick look overview (Table 4.1) of the decision scenarios for which the allocation procedures in each section of this chapter are developed.

Table 4.1: Quick look overview of allocation procedures developed in Chapter 4.

Section	**4.3**	**4.4**	**4.5**	**4.6a**	**4.6b**	**4.7**
Sampling Approach	single-stage	single-stage	single-stage	single-stage	single-stage	sequential
Attribute Type	success probability	continuous measurand	continuous measurand	continuous measurand	continuous measurand	continuous measurand
Number of Attributes	2	2	3	$k \geq 2$	$k \geq 2$	$k \geq 2$
Measurement Model	Bernoulli	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Measurement Uncertainty	NA	unknown	known	known	known	known
Estimation Approach	Bayesian	Bayesian	Bayesian	ML	Bayesian	Bayesian
Selection Procedure	multinomial	multinomial	multinomial	expected value	expected value	multinomial
Allocation Procedure Comparison	▪ uniform ▪ proportion ▪ step	▪ uniform ▪ proportion ▪ 1-param step ▪ 2-param step	▪ uniform ▪ proportion ▪ 3-zone ▪ 4-zone	▪ uniform ▪ MLE EV	▪ uniform ▪ Bayes EV	▪ uniform (1-stage) ▪ proportion (1-stage) ▪ sequential

4.2 A Simulation Approach

We relied on simulation studies to provide insights and evaluations of the allocation approaches. Unlike the examples presented in Chapter 3, by using simulated data with known truths we are able to determine when the alternative with the largest true decision value had been selected. This allowed for us to estimate the probability of correct selection as a performance and evaluation criteria of each of the developed allocation procedures.

We developed our allocation procedures with the focus on decision cases with $m \geq 2$ alternatives and $k \geq 2$ attributes whose true attribute values formed concave efficient frontiers (Keeney & Raiffa, 1993). Such a set of non-dominated decision alternatives provided the most challenging decision situation in terms of identifying the true best alternative under the assumed linear multiple attribute decision model (Assumption 3). While some of the allocation procedures are generalizable, most of the evaluations in this chapter are based on decision cases with $m = 5$ alternatives and $k = 2$ attributes. These values were chosen to create decision cases that were both non-trivial and capable of being evaluated with reasonable computational effort. In the Gaussian measurement model cases, the error associated with the measurement processes used to collect information for each attribute was also defined as part of the decision case. The description of the algorithms used to generate these simulated decision cases is described in the following sections.

Given a simulated decision case, which consisted of k true attribute values for each of m alternatives, and sometimes k measurement errors, we simulated results of the experimental evaluation process under a given sample allocation for each attribute of each alternative. That is, for each attribute j of each alternative a_i, we randomly selected n_{ij} observations from a probability distribution whose functional form is defined by the assumed measurement model and whose distributional parameters are equal to the true attribute value (and measurement error, if applicable) as defined by the decision case. Provided these simulated observations, $x_{ij1},\ldots,x_{ijn_{ij}}$, we modeled the value and uncertainty of attribute j of alternative a_i according to the defined estimation approach for all i and j. Based upon these attribute value models, we then, for each alternative a_i, evaluated the given multiple attribute decision model (with specific attribute weights) and propagated the attribute value

uncertainty using either the analytical approach or the Monte Carlo simulation approach presented in Section 3.3. The result of this evaluation was a description, i.e., distribution, of the decision value, ξ_i, for each of the $i = 1, \ldots, m$ alternatives. The defined selection procedure was applied to this set of *m* uncertain decision values, and an alternative was identified for selection.

Because the decision case provides the true attribute values, the true best alternative can be identified by evaluating the given multiple attribute decision model (with specific attribute weights), for each alternative, using its true attribute values. The result is a set of *m* true decision values, and the true best alternative is the one that provides the largest true decision value.

We compared the alternative that was identified for selection as a result of the simulated experimental evaluation to the true best alternative to conclude whether a correct selection had been made. By repeating this process many times, either for the same decision case or over multiple decision cases (or both), we are able to define the *frequency of correct selection* (*fcs*) measure as the proportion of times that a correct selection was made. As an estimate of the probability of correct selection (*PCS*), the *fcs* provides an evaluation metric for the allocation performance under the specific multiple attribute decision model. While details may vary, we used this general simulation and evaluation approach in all of the studies presented in this chapter.

4.3 Allocation for Two Attribute Bernoulli Measurement Model

In this section we describe a single-stage allocation method for the Bernoulli measurement model that was developed using results from an empirical study. In a single-stage allocation, the sample allocation (set of n_{ij} values) is determined prior to, and is unaltered throughout, the data collection process. We use the term *Bernoulli measurement model* to refer to the situation where the attributes are success probabilities that are evaluated using a series of Bernoulli trials (e.g., pass-fail testing).

We describe the outcomes of the evaluation process using the independent random variables $X_{ijl} \sim Bernoulli\left(\pi_{ij}\right)$ with the sample observations $x_{ij1},\ldots,x_{ijn_{ij}}$. In this notation, π_{ij} is the true attribute value, a success probability, for attribute *j* of alternative a_i and the notation of the decision model from Assumption 3 of Section 4.1 becomes as displayed in Equation (4.1) with linear individual value functions, $v_j\left(\pi_{ij}\right)=\pi_{ij}$.

$$\xi_i = f\left(\pi_{i1},\ldots,\pi_{ik}\right)=\sum_{j=1}^{k}\lambda_j v_j\left(\pi_{ij}\right) \tag{4.1}$$

We assume that the decision-maker's knowledge of the unknown true attribute value, π_{ij}, prior to observing any sample evaluations, can be described by a $Beta\left(\alpha_{ij},\beta_{ij}\right)$ probability distribution with parameters $\alpha_{ij}=\beta_{ij}=1,\forall i,j$. This $Beta(1,1)$ probability distribution is equivalent to a $Uniform[0,1]$ probability distribution and is interpreted as the decision-maker's a priori knowledge is that the true attribute value may be any value within the interval $(0,1)$ with equal likelihood. Let the random variable $Y_{ij}=\sum_{l=1}^{n_{ij}}X_{ijl}$ be the number of successful evaluations observed out of n_{ij} total evaluations, for attribute *j* of

alternative a_i, then $Y_{ij} \sim Binomial\left(n_{ij}, \pi_{ij}\right)$. Because the $Beta\left(\alpha, \beta\right)$ distribution is the conjugate prior distribution to the $Binomial\left(n, \pi\right)$ distribution, upon observing realizations $y_{ij} = \sum_{l=1}^{n_{ij}} x_{ijl}$, the decision-maker's knowledge of π_{ij} is updated and presented by the posterior distribution (Gelman et al., 2004) described in Equation (4.2). This is the probability distribution referenced in Assumption 2 of Section 4.1 for this Bernoulli measurement model.

$$\pi_{ij} \mid y_{ij} \sim Beta\left(1+y_{ij}, 1+n_{ij}-y_{ij}\right) \quad i=1,\ldots,m,\ j=1,\ldots,k \tag{4.2}$$

From the description of the decision-maker's knowledge of the true attribute values after observing the results of n_{ij} experimental evaluations for each attribute of each alternative, we used Monte Carlo simulation to propagate the attribute value uncertainty to the decision values, ξ_i. The multinomial selection procedure described in Section 3.4.2 was then used to select an alternative based upon then uncertain decision values.

It was further assumed that the overall fixed experimental budget (Assumption 4 of Section 4.1), $B = cmk$, where c is a positive integer, was divided equally among the m alternatives. The budget for each alternative was further divided among the k attributes. In general, the budgets for different alternatives could have been divided differently, but we made the simplifying assumption that the allocation is the same for all alternatives. And thus, $n_{1j} = \cdots = n_{mj}, \forall j$. This assumption is reasonable when there is no informative prior information about the alternatives that would suggest treating them differently.

4.3.1 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. We considered the situation where each of $m = 5$ alternatives is described by $k = 2$ attributes, and the two attributes are the alternative's probability of success in two different, independent tasks. The alternatives, when characterized by their true attribute values form a concave efficient frontier in R^2 space. The overall experimental budget of $B = 50$ Bernoulli trials was divided equally among the five alternatives. The problem was to determine the number of trials to be allocated to attribute 1, n_{i1}, and the number of trials to be allocated to attribute 2, n_{i2}, (where $n_{i1} + n_{i2} = 10$) to maximize the probability of correct selection.

To form the training decision cases for our initial simulation experimentation, we developed 20 concave efficient frontiers (decision cases) where each decision 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 using a construction algorithm (see Appendix A for details). The left panel of Figure 4.1 displays the attribute values of the 20 decision cases considered in our initial experimentation.

Given the random nature of the 20 decision cases, many aspects of possible sets of alternatives 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, *NL*, and a measure of general angle, θ. Generally speaking, the nonlinearity measure is the area of the polygon formed by the attribute values scaled by the distance between the extreme attribute pairs. The general angle is the acute angle formed by the line segment connecting the extreme attribute pairs and the horizontal line that passes through the maximum attribute 2 value (see Appendix B for details). The right panel of Figure 4.1 provides an illustration of the nonlinearity and general angle measures.

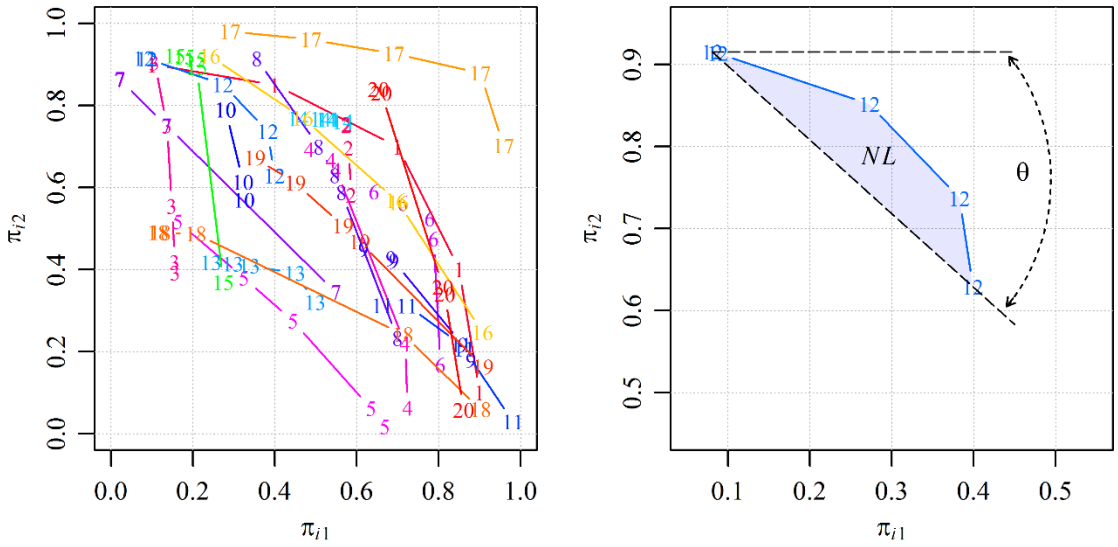

Figure 4.1: Twenty training cases (left panel) with the case number displayed by the numeric plotting character. Right panel illustrates the measures of nonlinearity (scaled shaded area) and general angle, θ.

Under the provided assumptions for this Bernoulli measurement model, we simulated, as described in Section 4.2, the experimental evaluation process for each case using all 11 possible sample allocations defined as (n_{i1}, n_{i2}) pairs: $(0,10),(1,9),\ldots,(10,0)$. The parameters of the Bayesian posterior distributions for the attribute values (Equation (4.2)) were calculated based on the outcome of each simulated evaluation. We defined 19 unique decision models by applying the 19 unique (λ_1, λ_2) pair of attribute decision weights:

$(0.05, 0.95), (0.10, 0.90), \ldots, (0.95, 0.05)$ to Equation (4.1). Using 1000 Monte Carlo simulation replicates, the attribute value uncertainty was propagated to the decision values under each decision model, and a selection was made using the multinomial selection procedure. For each decision case, we repeated this simulation a total of 1000 times, comparing each resulting selection to the true best alternative. The final result was 4180 *fcs* values: one for each of the 11 allocations, 19 decision models, and 20 decision cases.

4.3.2 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. Using the set of five alternatives with the true 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 for the attribute values.

As with our general problem, we allocated a fixed number of samples (in this case 250) across the five alternatives and two probability of success attributes as n_{i1} and $n_{i2} = 50 - n_{i1}$ where $n_{1j} = \cdots = n_{5j}, \forall j = 1, 2$. For each alternative and attribute, 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_{ij} equals π_{ij}, the true attribute value. That is, the observed results $y_{ij} = \pi_{ij} n_{ij}$ and the expected Bayesian posterior distribution is $Beta\left(1 + \pi_{ij} n_{ij}, 1 + n_{ij}\left(1 - \pi_{ij}\right)\right)$. The 95 % 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 % credible intervals for each attribute in R^2 space. Because the attributes are independent, the probability that the alternative's true attribute values are believed to be contained within this credible box is $0.95^2 = 0.9025$. Figure 4.2 displays graphically the expected credible boxes for each of the five alternatives for three different sample allocations.

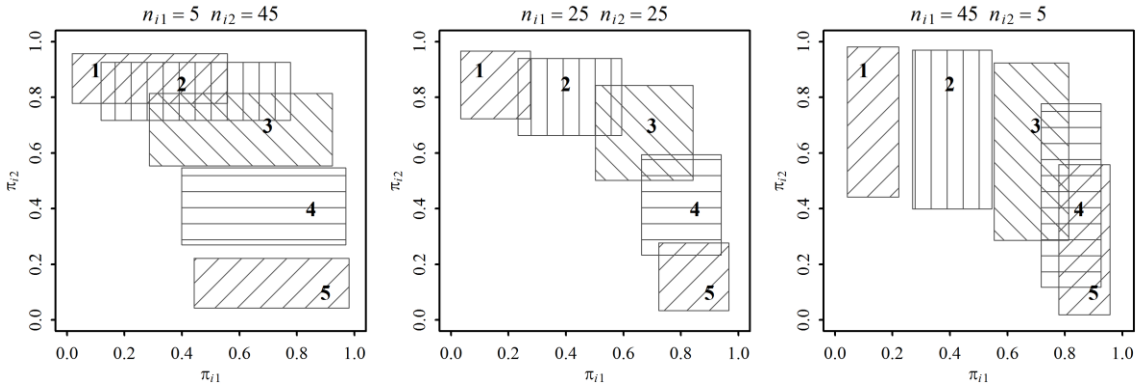

Figure 4.2: Expected credible boxes for three different sample allocations when the proportion of successes observed equals the true attribute value. The position of the plotting character (1 to 5) indicates the true attribute values for each alternative.

When moving from the left panel (sample allocation $n_{i1} = 5, n_{i2} = 45$) to the center panel ($n_{i1} = 25, n_{i2} = 25$) and to the right panel ($n_{i1} = 45, n_{i2} = 5$) of Figure 4.2, 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 λ_2 value). Whereby the opposite is suggested for allocations placing more samples with attribute 1 (better suited for decision models with large λ_1 value).

4.3.3 Frequency of Correct Selection Results

For each of the 20 decision cases and each of the 19 decision models considered in the decision simulation, there is at least one sample allocation value of n_{i1}, denoted n_{i1}^*, that produced the maximum frequency of correct selection (in some cases, for some values of λ_1, there were multiple values of n_{i1}^*). This optimal sample allocation should maximize the probability of choosing the true best alternative. Since $\lambda_2 = 1 - \lambda_1$ and $n_{i2} = 10 - n_{i1}$, it was sufficient to consider only λ_1 and n_{i1} when exploring the *fcs* results. The relationship between n_{i1}^* and λ_1 was seen to follow a general trend in which n_{i1}^* increased as λ_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 θ, the frontier characteristic measure of overall angle.

For each of the 11 sample allocations for each decision case and decision model, we defined the *relative frequency of correct selection* (*rel fcs*) as the ratio of the *fcs* for that sample allocation to the *fcs* that results from the optimal sample allocation. This measure provided a continuous response variable, as a function of n_{i1} and λ_1, that indicates the relative quality of each sample allocation versus the optimal sample allocation. The *rel fcs* 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_{i1}, λ_1, and the *rel fcs* as a contour plot for each case, further insights were gained.

The shaded contour plots of Figure 4.3 present the relative frequency of correct selection as a function of n_{i1} and λ_1 ranging from dark (low values) to light (high, desirable values). The solid squares within the plots denote n_{i1}^*, the sample allocation that produced the maximum frequency of correct selection for each λ_1 considered in the simulation study. The contour plots of Figure 4.3 serve to illustrate three general trends observed across the 20 training cases.

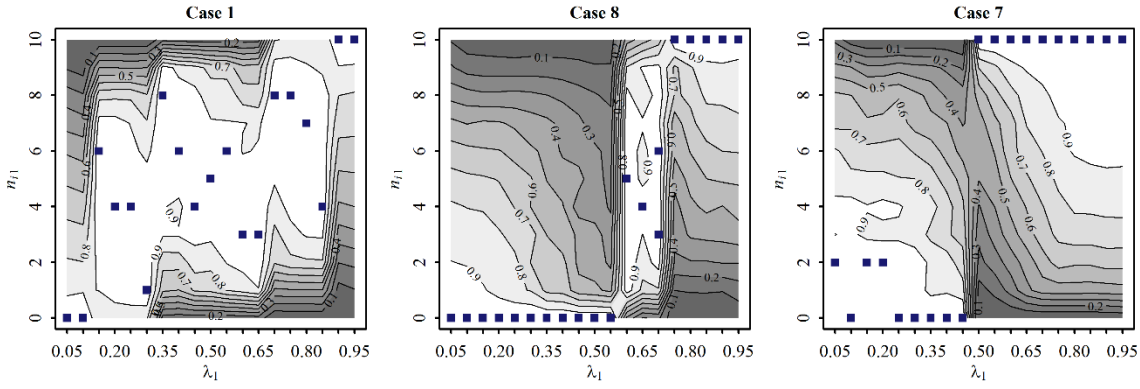

Figure 4.3: Contour plots for the *rel fcs* as a function of n_{i1} and λ_1 for three selected training cases. The solid squares denote n_{i1}^* for the λ_1 values evaluated.

The left panel in Figure 4.3 displays the contour plot as a result of decision case 1. The general increasing trend of n_{i1}^* as a function of λ_1 is observed, but further, it is observed that favorable *rel fcs* values 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 λ_1. The center panel in Figure 4.3 displays the contour plot as a result of decision case 8. Again the general increasing trend of n_{i1}^* as a function of λ_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 4.3 is the contour plot as a result of decision case 7. Here, favorable *rel fcs* values at low values of λ_1 correspond to low values of n_{i1} (lower left region of graphic) and at high values of λ_1 favorable relative frequency of correct selection values correspond to high values of n_{i1} (upper right region of graphic). But unlike the first two cases, these two regions are disconnected. We suspect that had we evaluated several decision models with $0.45 < \lambda_1 < 0.50$ in decision case 7 that a very steep and defined optimal "path" from $\lambda_1 = 0$ to $\lambda_1 = 1$ would have been presented.

4.3.4 Allocation Procedures

In general, the optimal allocation procedure 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 $n_{i1} = n_{i2} = \frac{B}{2m}$ would be implemented as the experimental plan to collect data to support the estimation of each attribute value of each alternative and ultimately the selection decision. We refer to this sample allocation as the *uniform allocation procedure*. 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 λ_1 and λ_2, is available, the decision-maker may choose to implement an allocation procedure that assigns n_{i1} and

n_{i2} proportional to λ_1 and λ_2. Such an approach is supported by the general insights discussed in Section 4.3.2 as well as the observations made is Section 4.3.3 that the optimal allocation n_{i1}^* increased as λ_1 increased. Since n_{i1} and n_{i2} must be integer values, rounding is necessary; we chose the rounding rule $n_{i1} = round\left(\lambda_1 \frac{B}{m}\right)$, $n_{i2} = \frac{B}{m} - n_{i1}$. We refer to this sample allocation approach as the *proportional allocation procedure*.

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, θ, a case-specific allocation procedure may be utilized. It was observed in the work described in Section 4.3.3 that as a function of λ_1, the optimal sample allocation, n_{ij}^*, 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, for $0 \le c_1 \le c_2 \le 1$ and $0 < \lambda_1 < 1$, a *step allocation procedure* with n_{i1} defined by Equation (4.3) and $n_{i2} = \frac{B}{m} - n_{i1}$.

$$n_{i1} = \begin{cases} 0 & \text{if } \lambda_1 \le c_1 \\ round\left[\left(\frac{\lambda_1 - c_1}{c_2 - c_1}\right)\frac{B}{m}\right] & \text{if } c_1 < c_2 \text{ and } c_1 < \lambda_1 \le c_2 \\ \frac{B}{m} & \text{if } \lambda_1 > c_2 \end{cases} \tag{4.3}$$

For each of our 20 training cases, we determined the values of c_1 and c_2 in Equation (4.3) that maximized the average relative frequency of correct selection across the 19 evaluated values of λ_1. From this, we hypothesized a quadratic relationship between θ and c_1

(and c_2) and found the best fit (Equation (4.4)) using restricted least-squares regression that required the modeled responses $c_1 = c_2 = 0$ when $\theta = 0$, and $c_1 = c_2 = 1$ when $\theta = 90$.

$$c_1(\theta) = \tfrac{1}{225}\theta + \tfrac{1}{13500}\theta^2$$
$$c_2(\theta) = \tfrac{7}{450}\theta - \tfrac{1}{20250}\theta^2 \qquad (4.4)$$

Note that for a horizontal frontier ($\theta = 0$ and $c_1 = c_2 = 0$) and a vertical frontier ($\theta = 90$ and $c_1 = c_2 = 1$) this step procedure provides the sample allocations $(n_{i1} = 10, n_{i2} = 0)$ and $(n_{i1} = 0, n_{i2} = 10)$, respectively, for any decision model values of λ_1 and λ_2.

An illustration of the sample allocations generated by each allocation procedures and their resulting *rel fcs* values over the range of λ_1 for training case 8 is provided in Figure 4.4. As in Figure 4.3, the shaded contour plots of Figure 4.4 present the *rel fcs* as a function of n_{i1} 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_{i1}^*, the optimal sample allocation at each λ_1 value. The solid red line in each plot represents the sample allocation generated by the specified allocation procedure for the attribute decision weight, λ_1.

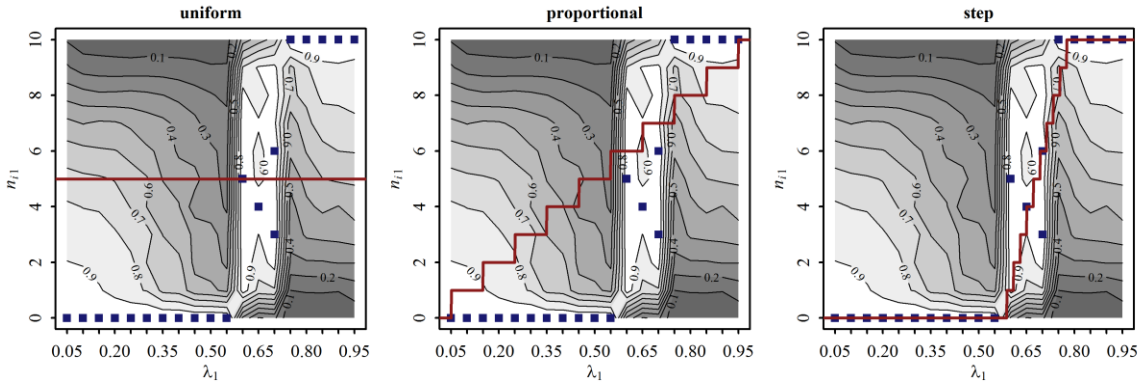

Figure 4.4: Sample allocations (solid red line) generated by each allocation procedures for training case 8. The shaded contour plots display the *rel fcs* as a function of n_{i1} and λ_1. The solid squares denote n_{i1}^*.

A good allocation procedure provides sample allocations that result in high *rel fcs* values for any attribute decision weight (λ_1 value); an ideal procedure provides a sample allocation equal to n_{i1}^* at each λ_1 value. While no procedure presented in Figure 4.4 provides an optimal sample allocation at every λ_1 value, the step allocation procedure offers 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 procedures are seen to result in less desirable (darkly shaded) *rel fcs* values over many values of λ_1.

4.3.5 Comparison of Allocation Procedures

The second part of our study compared the performance of the allocation procedures. One hundred new concave frontiers (evaluation cases), each containing 5 alternatives described by 2 attributes were randomly generated using the construction algorithm described in Section 4.3.1. These evaluation cases were used to evaluate and compare the three sample allocation procedures described in Section 4.3.4 (uniform allocation, proportional allocation, and step procedure). The attribute values of the 100 evaluation cases and their nonlinearity and general angle measure values are displayed in Figure 4.5.

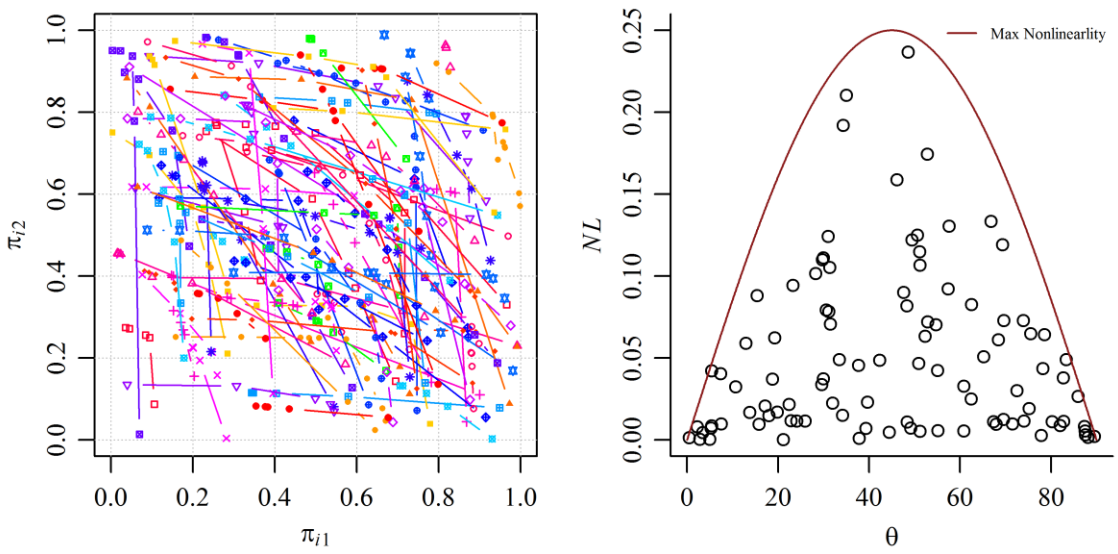

Figure 4.5: Attribute values (left panel) and nonlinearity and general angle measure values (right panel) for the 100 concave frontiers in the evaluation set.

The sample allocations directed by each of the three allocation procedures were determined for each of the 100 evaluation cases. Based on these sample allocations, we used the decision simulation described in Section 4.2, again with 1000 simulation replicates and 1000 Monte Carlo simulation replicates, to evaluate the *fcs* (and *rel fcs*) for each evaluation case and each of the 19 decision models (λ_1, λ_2 pair). We then examined, on a case-by-case basis, the relative frequency of correct selection that resulted from the simulation, as a function of λ_1. At each λ_1 value, we calculated the average *rel fcs* across the 100 evaluation cases. The uncertainties in the average *rel fcs* were expressed as 95 % pointwise confidence intervals based on the normality assumptions provided by the Central Limit Theorem.

While it varied from case to case and across λ_1 values, in general, the step allocation procedure provided the largest relative frequency of correct selection values. The proportional allocation procedure displayed similar performance to the step procedure with λ_1

values near 0 and near 1, but provided lower *rel fcs* values elsewhere. The uniform allocation procedure nearly always produced the smallest *rel fcs* values. Figure 4.6 illustrates these general conclusions by displaying, for each of the three allocation procedures, the *rel fcs* averaged across all evaluation cases at each λ_1 value.

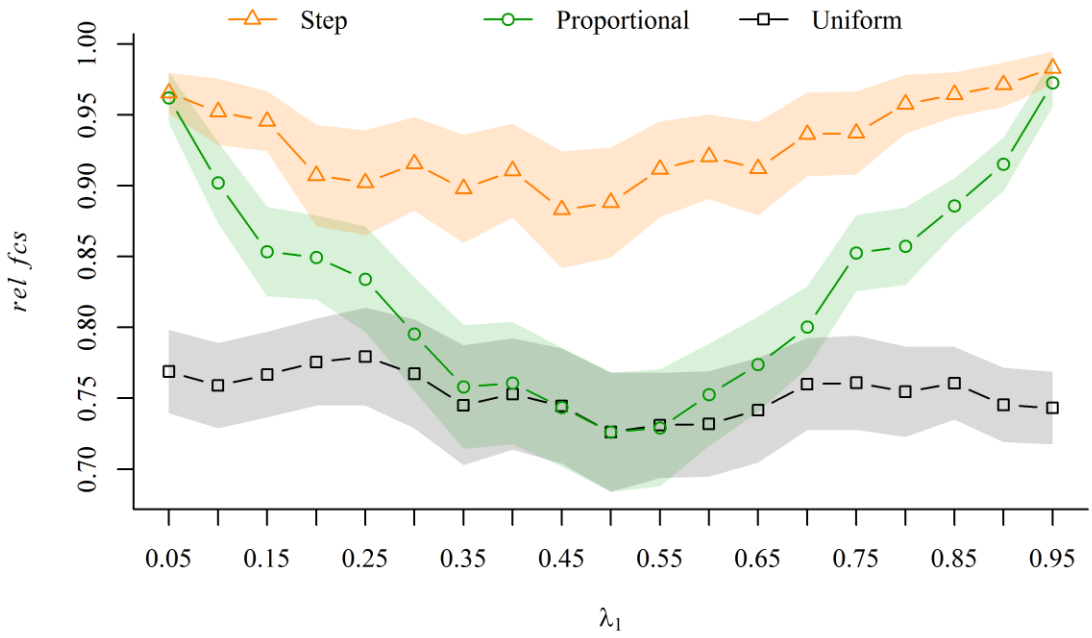

Figure 4.6: *rel fcs* values averaged across all evaluation cases for the step allocation for each λ_1 value; shaded area represents 95 % pointwise confidence bounds for the true relative *PCS*.

It is observed from Figure 4.6 that sample allocations provided by the uniform allocation procedure lead to, on average, a relative frequency of correct selection near 0.75, and this performance is nearly constant over all values of λ_1. The sample allocations provided by the proportional allocation procedure near $\lambda_1 = 0.5$ are identical to the uniform allocation, and hence the performance of the proportional and uniform allocation procedures are comparable in this region. The relative frequency of correct selection for the proportional

allocation procedure increases as λ_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 procedure is approximately 0.83. The step procedure provided 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 λ_1, the step procedure provides average relative frequency of correct selection values that are statistically distinguishable (non-overlapping confidence intervals) and superior to the other allocation procedures.

It can be observed from the right panel of Figure 4.5 that there were a number of evaluation cases that displayed nearly identical frontier curve characteristic measures of non-linearity and general angle. The contour plots of the *rel fcs* as a function of n_{i1} and λ_1, and ultimately the *rel fcs* values that resulted from the sample allocations dictated by the three allocation procedures were very similar when comparing several such pair of evaluation cases near the limits of the characteristic measures ($\theta = 0$ or $\theta = 90$). 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.

4.4 Allocation for Two Attribute Gaussian Measurement Model

In this section we describe a single-stage allocation method for the Gaussian measurement model that was developed using results from an empirical study. Again, in a single-stage allocation, the sample allocation (set of n_{ij} values) is determined prior to, and is unaltered throughout, the data collection process. We use the term *Gaussian measurement model* to refer to the situation where the attributes are continuous measurands and the error in the process used to obtain the measurements is modeled as a normal probability distribution (e.g., measured physical quantities such as distance).

In the Gaussian measurement model we use the random variable $X_{ijl} = \mu_{ij} + \varepsilon_{ijl}$, with realized sample observations denoted $x_{ij1},\ldots,x_{ijn_{ij}}$, to describe the l^{th} outcome from the measurement process used in estimating the value of the j^{th} attribute of alternative a_i. Here μ_{ij} is the true attribute value, ε_{ijl} is the random measurement error, and n_{ij} is the number of measurements obtained for attribute *j* of alternative a_i. This notation is the basis for the decision model as written in Assumption 3 of Section 4.1 and repeated here as Equation (4.5) with linear individual value functions $v_j\left(\mu_{ij}\right) = \mu_{ij}$.

$$\xi_i = f\left(\mu_{i1},\ldots,\mu_{ik}\right) = \sum_{j=1}^{k} \lambda_j v_j\left(\mu_{ij}\right) \tag{4.5}$$

An experimental sample (measurement process) generates one random measurement of one attribute of one alternative. Separate and independent measurement processes are used in obtaining n_{ij} measurement data (samples) for each attribute, and for a given attribute *j*, the measurement process is the same for all $i = 1,\ldots,m$ alternatives. Further, it is

assumed that the measurement errors associated with these measurement processes are independent and identically distributed (i.i.d.) and can be modeled by a normal probability distribution centered at zero, $\varepsilon_{ijl} \overset{i.i.d.}{\sim} N\left(0, \sigma_j^2\right), i = 1, \ldots . m; l = 1, \ldots, n_{ij}$. It follows that the measurement distribution for $X_{ijl} \sim N\left(\mu_{ij}, \sigma_j^2\right)$, and in this section we assume that σ_j^2 is unknown. It is common in the field of metrology for the error associated with a continuous measurand to be modeled with a Gaussian distribution (Joint Committee for Guides in Metrology, 2008).

We assume that the decision-maker's prior knowledge of the unknown true parameters for this Gaussian measurement model, μ_{ij} and σ_j^2, can be described by the probability distributions provided in Equation (4.6).

$$\begin{aligned} \mu_{ij} &\sim N\left(\mu_{0ij}, \tau_{0ij}^2\right) \\ \sigma_j^2 &\sim \text{Inv-}\chi^2\left(\nu_{0j}, \sigma_{0j}^2\right) \end{aligned} \tag{4.6}$$

While it was preferable to minimize the assumption on prior knowledge by using a non-informative pair of prior distributions such as a Jeffrey's prior on σ_j^2 and a uniform distribution over $\left(-\infty, \infty\right)$ on μ_{ij}, these priors are improper priors and do not integrate to one. When updated based on observed data and Bayes rule, the resulting posterior is indeed proper. However, in this work there are situations where no data is provided, i.e., no sample allocated to the attribute, thus the prior distributions will be relied upon as a surrogate for the posterior, and since the non-informative prior distributions are improper, they are not useful in the analysis. Also, this non-informative uniform prior on μ_{ij} allows for negative values to be observed with non-negligible probabilities, which is problematic in a measurement model where measurand values are often assumed to be non-negative.

The selected prior distributions of Equation (4.6) are attractive as they do maintain some non-informative traits. The selected prior on σ_j^2 approaches a Jeffrey's prior in form when $\sigma_{0j}^2 = 1$ and ν_{0j} is very small (e.g., 0.001), but remains a proper prior. The selected prior on μ_{ij} can approach uniform distribution in form when τ_{0ij}^2 is selected to be very large. To best meet these goals, we used the prior parameters $\mu_{0ij} = 150$ and $\tau_{0j}^2 = 35^2$, and $\nu_{0j} = 0.01$ and $\sigma_{0j}^2 = 1$, for all *i* and *j*. The prior parameters for μ_{ij} were chosen to be centered on the domain of the true attribute values as defined in the simulation study described below in Section 4.4.1, with a variance that provided as large a distributional spread as possible without creating non-negligible probabilities over negative values of μ_{ij}. The prior parameters for σ_j^2 were chosen so that the prior distribution closely resembles a non-informative Jeffrey's prior distribution.

This Bayesian semi-conjugate prior model for normally distributed data (Gelman et al., 2004) provided the useful conditional posterior distribution for μ_{ij} given by Equation (3.7).

$$\mu_{ij} \mid \sigma_j^2, x_{ij1}, \ldots, x_{ijn_{ij}} \sim N\left(\frac{\sigma_j^2 \mu_{0ij} + n_{ij} \tau_{0ij}^2 \bar{x}_{ij}}{\sigma_j^2 + n_{ij} \tau_{0ij}^2}, \frac{\sigma_j^2 \tau_{0ij}^2}{\sigma_j^2 + n_{ij} \tau_{0ij}^2} \right) \tag{4.7}$$

To make use of Equation (4.7), a Monte Carlo simulation approach must be used that first samples from the marginal posterior distribution of σ_j^2, $p\left(\sigma_j^2 \mid x_{ij1}, \ldots, x_{ijn_{ij}}\right)$, and then samples from the conditional posterior distribution for μ_{ij}. In this case, however, the marginal posterior distribution of σ_j^2 does not follow a standard parametric form, and thus must be computed numerically (see pp. 82 (Gelman et al., 2004)). Fortunately, the LearnBayes package (Albert, 2014) of the R software environment for statistical computing and

graphics (R Core Team, 2015) provides a function (normpostsim) that produces simulated samples from the joint posterior distribution of μ_{ij} and σ_j^2 using efficient Gibbs sampling.

The simulated samples of μ_{ij} provided empirical distributions for the decision-maker's knowledge of the true attribute values after observing n_{ij} sample measurements for each attribute of each alternative. From these distributions, we used Monte Carlo simulation to propagate the attribute value uncertainty to the decision values, ξ_i. The multinomial selection procedure described in Section 3.4.2 was then used to select an alternative based upon the uncertain decision values.

We again assumed that there is no informative prior information about the alternatives that would suggest treating them differently and thus the overall fixed experimental budget allocation was the same for all alternatives. Specifically, the overall fixed experimental budget, $B = cmk$, where c is a positive integer, was divided equally among the m alternatives and further divided among the k attributes such that $n_{1j} = \cdots = n_{mj}, \forall j$.

4.4.1 Simulation Study

We conducted a simulation study to understand how the experimental sample allocation affects the probability of correct selection and to test our general intuition that 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. We considered the situation where each of $m = 5$ alternatives is described by $k = 2$ attributes that are measured using two different and independent measurement processes. The error of each measurement technique is normally distributed with unknown

variance. The alternatives, when characterized by their true attribute values form a concave efficient frontier in R^2 space. The overall experimental budget of $B = 50$ sample measurements was divided equally among the five alternatives. The problem was to determine the number of samples to be allocated to attribute 1, n_{i1}, and the number of samples to be allocated to attribute 2, n_{i2}, (where $n_{i1} + n_{i2} = 10$) to maximize the probability of correct selection.

Based on the results of Section 4.3, which considered pass-fail testing to estimate attribute values in terms of probabilities, we expected that sample allocation procedures might provide the decision-maker with guidance. We generated a set of 20 training cases and used these to guide the values of the parameters in our sample allocation procedures. Like the Bernoulli measurement model study, our decision cases consisted of five alternatives described by two attributes. To generate these training cases, we transformed the true values of the attributes of the random training cases used in the Bernoulli measurement model study to the domain [100, 200] and again measured the nonlinearity and general angle characteristics of each case. Figure 1 displays the true attribute values of the 20 training cases.

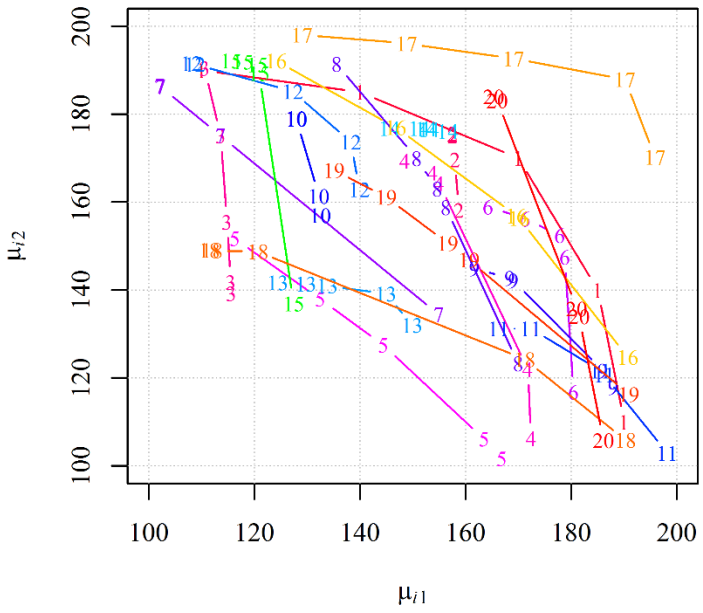

Figure 4.7: Twenty training cases for the Gaussian measurement model with the case number displayed by the numeric plotting character.

In the simulation study for the Gaussian measurement model, we also needed to provide the true but unknown variance of the measurement error, σ_j^2, associated with the measurement process used to measure the value of each attribute *j*. For each training cases, we considered the 16 different *measurement error scenarios* described in Table 4.2 which defined the variance of the measurement errors, σ_1^2 and σ_2^2, for the measurement processes used to measure attribute 1 and attribute 2. We referred to a decision case under a particular measurement error scenario as a "subcase" and used the notation 16.30.1, for example, to refer to decision case 16 with a measurement error variance for attribute 1 of 30^2 and a measurement error variance for attribute 2 of 1^2. In this initial simulation experiment for

the Gaussian measurement model, we considered 320 subcases: each of the 20 decision cases evaluated under each of the 16 measurement error scenarios.

Table 4.2: Measurement error scenarios for the Gaussian measurement model simulation study with two attributes.

Scenario	σ_1^2	σ_2^2
1	1^2	1^2
2	1^2	10^2
3	1^2	20^2
4	1^2	30^2
5	10^2	1^2
6	10^2	10^2
7	10^2	20^2
8	10^2	30^2
9	20^2	1^2
10	20^2	10^2
11	20^2	20^2
12	20^2	30^2
13	30^2	1^2
14	30^2	10^2
15	30^2	20^2
16	30^2	30^2

Under the provided assumptions for this Gaussian measurement model, we simulated, as described in Section 4.2, the experimental evaluation process for each subcase using all 11 possible sample allocations defined as the (n_{i1}, n_{i2}) pairs: $(0,10),(1,9),\ldots,(10,0)$. We defined 19 unique decision models by applying the 19 unique (λ_1, λ_2) pair of attribute decision weights: $(0.05,0.95),(0.10,0.90),\ldots,(0.95,0.05)$ to Equation (4.5). Based on the outcome of the simulated evaluation process for each subcase under each allocation, we

generated 1000 simulated samples from the conditional posterior distribution of μ_{ij} (Equation (4.7)) for each attribute of each alternative and calculated the decision values, ξ_i, as defined in Equation (4.5) for each of the 19 decision models. Using these empirical distributions of ξ_i, an alternative was selected using the multinomial selection procedure and checked to determine whether this alternative was the true best. We repeated this experimental evaluation simulation a total of 1000 times. The final result was 66,880 *fcs* values: one for each of the 11 allocations, 19 decision models, and 320 subcases.

4.4.2 Frequency of Correct Selection Results

We identified n_{i1}^* and defined the *rel fcs* for each of the 320 subcases and each of the 19 decision models. Recall that for a given decision model, n_{i1}^* is the optimal sample allocation for attribute 1 that produced the maximum frequency of correct selection and the *relative frequency of correct selection* (*rel fcs*) for each sample allocations is the ratio of the frequency of correct selection for that sample allocation to the frequency of correct selection for the n_{i1}^* allocation. Though complicated by the measurement error scenario, similar observations were made regarding the relationship between n_{i1}^* and λ_1 as were observed in Section 4.3.3.

The shaded contour plots of Figure 4.8 present the *rel fcs* as a function of n_{i1} 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_{i1}^*, the optimal sample allocation for

attribute 1 at each λ_1 value. The contour plots of Figure 4.8 serve to illustrate two general trends observed across the 320 subcases.

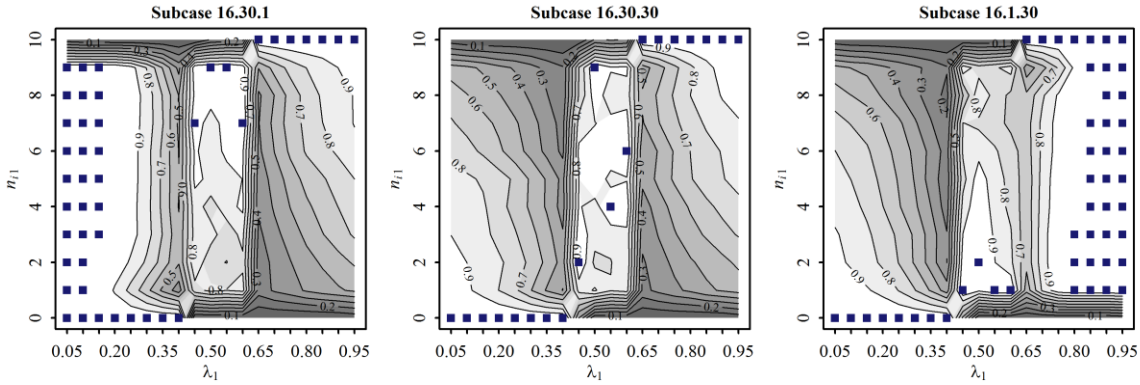

Figure 4.8: Contour plots displaying *rel fcs* as a function of n_{i1} and λ_1 for training case 16 under measurement error scenarios $\left(30^2,1^2\right)$, $\left(30^2,30^2\right)$, and $\left(1^2,30^2\right)$. For each λ_1, the squares denote n_{i1}^*.

First, consider the left panel of Figure 4.8 (subcase 16.30.1) where the measurement error variance 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_{i1} is near 10) produce favorable *rel fcs* values. When, however, attribute 1 is not important (when 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, non-zero values of n_{i2} still produce favorable *rel fcs* values. This phenomenon is seen in reverse in subcase 16.1.30 (right panel, Figure 4.8). Subcase 16.30.30 (center panel, Figure 4.8) combines the restrictive observations of the previous two subcases because both attributes have large measurement error variance $\left(\sigma_1^2 = \sigma_2^2 = 30^2\right)$. 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.

Like we observed in Section 4.3.3, we see in Figure 4.8 that, in general, as the importance of attribute 1 increases (that is, as the weight λ_1 increases from 0 to 1), the n_{i1} 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 4.8). 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.

4.4.3 Allocation Procedures

As in the Bernoulli study of Section 4.3, we again consider the uniform allocation procedure, a balanced allocation of $n_{i1} = n_{i2} = \frac{B}{2m}$, and the proportional allocation procedure, where the decision-maker assigns n_{i1} and n_{i2} proportional to λ_1 and λ_2 by $n_{i1} = round\left(\lambda_1 \frac{B}{m}\right)$ and $n_{i2} = \frac{B}{m} - n_{i1}$. Although the proportional allocation procedure is simple, the evaluations showed that the relationship between λ_1 and n_{i1}^* was usually not linear, but rather, distinctly nonlinear. To approximate this relationship, we used a step-like function that we called a *step allocation procedure*. For $0 \le c_1 \le c_2 \le 1$ and $0 < \lambda_1 < 1$, the step procedure assigns n_{i1} as defined by Equation (4.8) and $n_{i2} = \frac{B}{m} - n_{i1}$.

$$n_{i1} = \begin{cases} 0 & \text{if } \lambda_1 \le c_1 \\ round\left[\left(\dfrac{\lambda_1 - c_1}{c_2 - c_1}\right)\dfrac{B}{m}\right] & \text{if } c_1 < c_2 \text{ and } c_1 < \lambda_1 \le c_2 \\ \dfrac{B}{m} & \text{if } \lambda_1 > c_2 \end{cases} \tag{4.8}$$

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 (Equation (4.9)) using restricted least-squares regression that required the modeled responses $c_1 = c_2 = 0$ when $\theta = 0$ (a horizontal frontier), and $c_1 = c_2 = 1$ when $\theta = 90$ (a vertical frontier). We refer to the step allocation procedure as a *1-parameter step allocation procedure* when c_1 and c_2 in the step function of Equation (4.8) are defined according to the univariate models of Equation (4.9).

$$\begin{aligned} c_1 &= f_1(\theta) = 8.47\times10^{-4}\theta + 1.14\times10^{-4}\theta^2 \\ c_2 &= f_2(\theta) = 2.09\times10^{-2}\theta - 1.09\times10^{-4}\theta^2 \end{aligned} \tag{4.9}$$

We also hypothesized a bivariate relationship between θ and *NL* and c_1 (and c_2) and found the best fit (Equation (4.10)) using restricted least-squares regression that required the modeled responses $c_1 = c_2 = 0$ when $\theta = 0$ and $NL = 0$; $c_1 = c_2 = 1$ when $\theta = 90$ and $NL = 0$; and, $c_1 = 0, c_2 = 1$ when $\theta = 45$ and $NL = 0.25$. We refer to the step allocation

procedure as the *2-parameter step allocation procedure* when c_1 and c_2 in the step function of Equation (4.8) are defined according to the bivariate models of Equation (4.10).

$$\begin{aligned} c_1 &= f_1(\theta, NL) = 1.11\times10^{-2}\theta + 0.674NL - 5.94\times10^{-2}\theta\cdot NL \\ c_2 &= f_2(\theta, NL) = 1.11\times10^{-2}\theta + 4.56NL - 5.69\times10^{-2}\theta\cdot NL \end{aligned} \tag{4.10}$$

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 for a horizontal frontier ($\theta = 0$, $NL = 0$, and $c_1 = c_2 = 0$) and a vertical frontier ($\theta = 90$, $NL = 0$, and $c_1 = c_2 = 1$), both the 1-parameter and the 2-parameter step procedures provide the sample allocations $(n_{i1} = 10, n_{i2} = 0)$ and $(n_{i1} = 0, n_{i2} = 10)$, respectively, for any decision model values of λ_1 and λ_2.

An illustration of the sample allocations generated by each allocation procedure and their resulting *rel fcs* values over the range of λ_1 for subcase 8.20.20 is provided in Figure 4.9. As in Figure 4.8, the shaded contour plots of Figure 4.9 present the *rel fcs* as a function of n_{i1} 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_{i1}^*, 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 procedure for the range of attribute decision weight, λ_1.

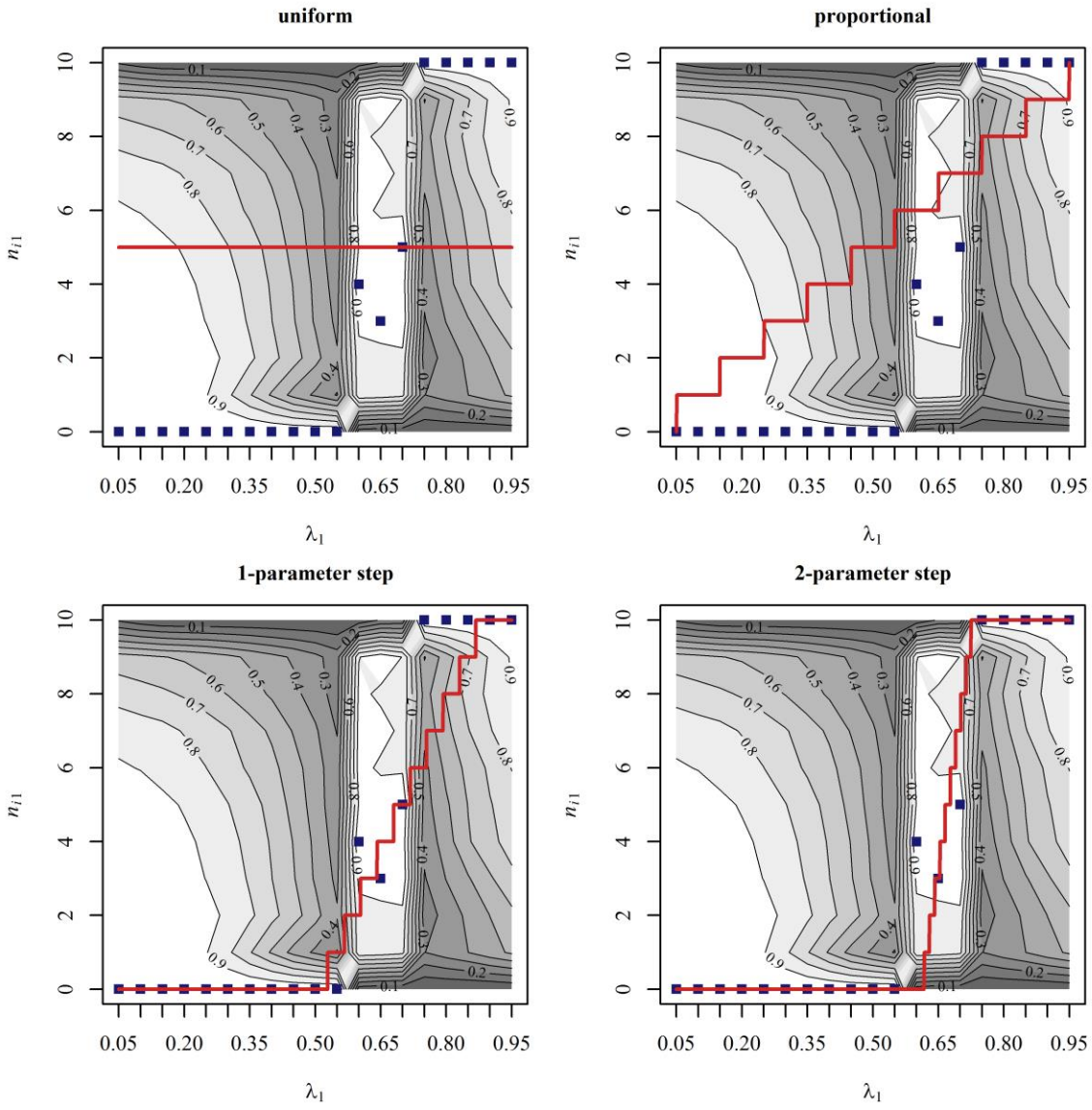

Figure 4.9: Sample allocations (red line) generated by each allocation procedure for training subcase 8.20.20. The shaded contour plots display the *rel fcs* as a function of n_{i1} and λ_1 with n_{i1}^* denoted as solid squares.

A good allocation procedure provides sample allocations that result in high *rel fcs* values for any attribute decision weight (λ_1 value); an ideal procedure provides a sample allocation equal to n_{i1}^* at each λ_1 value. While no procedure presented in Figure 4.9 provides an optimal sample allocation at every λ_1 value, the step allocation procedures offer 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 procedures are seen to result in less desirable (darkly shaded) *rel fcs* values over many values of λ_1.

4.4.4 Comparison of Allocation Procedures

To test the sample allocation procedures, we generated 500 new concave frontiers (evaluation cases) using the construction algorithm. Each evaluation case was a set of 5 randomly generated alternatives described by 2 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]. We calculated the nonlinearity (*NL*) and general angle (θ) measures for each evaluation case.

We tested each of the sample allocation procedures developed in Section 4.4.3 using all 500 evaluation cases and 19 decision models. To each of the 500 evaluation cases, we assigned a pair of measurement error variance values, σ_1^2 and σ_2^2. The assigned σ_j values $(j = 1, 2)$ were independent, random draws from a uniform distribution with parameters min = 1 and max = 30. Then, for each of the 11 possible sample allocations, for each of the 19 attribute models, across the 500 evaluation cases, we obtained a *rel fcs* value using the same simulation process used for the training cases as described in Section 4.4.1. For each evaluation case and decision model, we used each of the sample allocation procedures to produce a sample allocation. The *rel fcs* value from the results of the simulation for these allocations were identified. The performance of an allocation procedure, for each attribute decision weight, was defined to be the average *rel fcs* of its sample allocation across the

500 evaluation cases. The uncertainties in the average *rel fcs* were expressed as 95 % pointwise confidence intervals based on the normality assumptions provided by the Central Limit Theorem.

The general ranking of the allocation procedures in terms of performance (average *rel fcs*) from best-performing to worst-performing is as follows: 2-parameter step, 1-parameter step, proportional, uniform. At λ_1 values very near 0 and very near 1 the proportional allocation procedure and the step procedures provide similar sample allocations ($n_{i1} = 0$ at $\lambda_1 \approx 0$ and $n_{i1} = \frac{B}{m}$ at $\lambda_1 \approx 1$) and thus displayed similar performance at these attribute decision weight values. At λ_1 values near 0.5 the proportional allocation procedure and the uniform allocation procedure provide similar sample allocations ($n_{i1} = n_{i2} = \frac{B}{2m}$) and thus displayed similar performance at these attribute decision weight values. Figure 4.10 illustrates these general conclusions by displaying, for each of the four allocation procedures studied, the relative frequency of correct selection averaged across all evaluation cases at each λ_1 value and the 95 % pointwise confidence interval for the true relative *PCS* value.

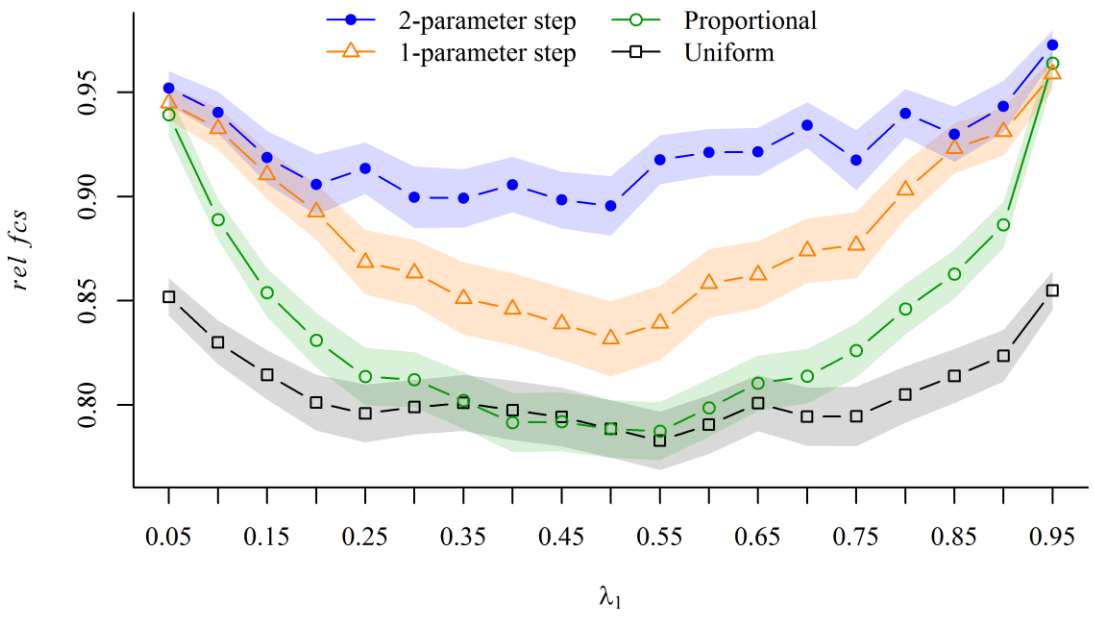

Figure 4.10: *rel fcs* values averaged across all testing cases for the 2-parameter step allocation for each λ_1 value; shaded area represents 95 % pointwise confidence bounds for the true relative *PCS*.

As shown in Figure 4.10, the sample allocations generated by the uniform allocation procedure 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 procedure 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 procedure is approximately 0.84. The one-parameter step procedure provides a maximum average *rel fcs* value of 0.96 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 procedure provides a maximum average *rel fcs* value of 0.97 at $\lambda_1 = 0.95$, a minimum average value of 0.90 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 procedure, 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 procedures.

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 measurement 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 evaluation curves in Figure 4.10 tend be lower when λ_1 is near 0.5 and higher when λ_1 is near 0 or 1.

4.5 Allocation for Three Attribute Gaussian Measurement Model

In this section we continue our study of the single-stage allocation method for the Gaussian measurement model. We develop allocation methods for a fixed experimental sample budget across three attributes based on insights gained from an empirical study.

As a continuation of the Gaussian measurement model work presented in Section 4.4, the notation regarding the measurement process and the decision model also hold in this $k = 3$ attribute study. We again assume that the measurement errors associated with the measurement processes, $\varepsilon_{ijl} \overset{i.i.d.}{\sim} N\left(0, \sigma_j^2\right), i = 1, \ldots . m, l = 1, \ldots, n_{ij}$ and the measurement dis-

tribution $X_{ijl} \sim N\left(\mu_{ij}, \sigma_j^2\right)$. Unlike Section 4.4, here we assume that σ_j^2 is known. In metrology, it is not uncommon for the variance of the measuring process to be well characterized and assumed known (Lee P. M., 1997).

This knowledge of σ_j^2 allows for us to use a conjugate normal prior distribution, $N\left(\mu_{0ij}, \tau_{0j}^2\right)$, to describe the decision-maker's prior knowledge of the single unknown true parameters for this Gaussian measurement model, μ_{ij} (the true attribute value). While this conjugate prior model simplified computations over the semi-conjugate model used in the two-attribute study, we needed to continue to be cautious of defining a prior distribution that allowed for negative values to be observed with non-negligible probabilities and thus we used the prior parameters $\mu_{0ij} = 150$ and $\tau_{0j}^2 = 35^2$. These prior parameters for μ_{ij} were chosen to be centered on the domain of the true attribute values as defined in the simulation study described below in Section 4.5.1, with a variance that provided as large a distributional spread as possible without creating non-negligible probabilities over negative values of μ_{ij}. Upon observing n_{ij} measurement samples for attribute *j* of alternative a_i, the decision-maker's knowledge of μ_{ij} is updated and presented by the normally distributed posterior distribution (Gelman et al., 2004) displayed in Equation (4.11).

$$\mu_{ij} \mid x_{ij1}, \ldots, x_{ijn_{ij}} \sim N\left(\frac{\sigma_j^2 \mu_{0ij} + n_{ij}\tau_{0ij}^2 \bar{x}_{ij}}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}, \frac{\sigma_j^2 \tau_{0ij}^2}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}\right) \tag{4.11}$$

As discussed in Section 3.3, since the decision model (Equation (4.5)) is linear, we are able to analytically propagate the uncertainty in the decision-maker's knowledge of μ_{ij} to

the decision value ξ_i. That is, given $\mathbf{x}_i$, the total sample data observed in support of alternative a_i, the decision-maker's knowledge of the true decision value, ξ_i, can be described by the posterior distribution in Equation (4.12).

$$\xi_i \mid \mathbf{x}_i \sim N\left(\sum_{j=i}^{k} \lambda_j \frac{\sigma_j^2 \mu_{0ij} + n_{ij}\tau_{0ij}^2 \bar{x}_{ij}}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}, \sum_{j=1}^{k} \lambda_j^2 \frac{\sigma_j^2 \tau_{0ij}^2}{\sigma_j^2 + n_{ij}\tau_{0ij}^2} \right) \tag{4.12}$$

Based on these uncertain decision values, we used the multinomial selection procedure described in Section 3.4.2 to select an alternative.

Like in the two-attribute study, we assumed that the overall fixed experimental budget, $B = cmk$, where c is a positive integer, was divided equally among the m alternatives and further divided among the k attributes such that $n_{1j} = \cdots = n_{mj}, \forall j$ (i.e., the allocation is the same for all alternatives).

4.5.1 Simulation Study

We conducted a simulation study similar to that used in studying the two attribute Gaussian measurement model to understand how the experimental sample allocation affects the probability of correct selection. However, in this case, we considered the situation where each of $m = 5$ alternatives is described by $k = 3$ attributes. Again, each attribute is measured using a different measurement technique, and an experimental sample (measurement process) generates one random measurement of one attribute of one alternative. The error of each measurement technique is normally distributed with known variance. The alternatives, when characterized by their true attribute values form a concave efficient frontier in R^3 space. The overall experimental budget of $B = 45$ sample measurements was divided equally among the five alternatives. The problem was to determine the sample allocation

triplet, (n_{i1}, n_{i2}, n_{i3}) such that $n_{i1} + n_{i2} + n_{i3} = 9$ to maximize the probability of correct selection.

We generated a set of 20 training cases, evaluated every possible sample allocation, and used the results to generate insights for developing sample allocation procedures. 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 construction algorithm used to generate the training cases in R^3 space is provided in Appendix C.

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 measurement error variance of the three attributes, σ_1^2, σ_2^2, and σ_3^2, to one of 10^2 or 30^2, which created the $2^3 = 8$ different "measurement error scenarios" described in Table 4.3. Again, we referred to a decision case under a particular measurement error scenario as a "subcase" and we use the notation analogous to that used in two attribute study, e.g., 7.30.10.10, to refer to decision case 7 with measurement error variances $\sigma_1^2 = 30^2$, $\sigma_2^2 = 10^2$, and $\sigma_3^2 = 10^2$. In this initial simulation experiment for the three attribute Gaussian measurement model, we considered 160 subcases: each of the 20 decision cases evaluated under each of the 8 measurement error scenarios.

Table 4.3: Measurement error scenarios for the Gaussian measurement model simulation study with three attributes.

Scenario	σ_1^2	σ_2^2	σ_3^2
1	10^2	10^2	10^2
2	10^2	10^2	30^2
3	10^2	30^2	10^2
4	10^2	30^2	30^2
5	30^2	10^2	10^2
6	30^2	10^2	30^2
7	30^2	30^2	10^2
8	30^2	30^2	30^2

Under the provided assumptions for this three attribute Gaussian measurement model, we simulated, as described in Section 4.2, the experimental evaluation process for each subcase using all 55 possible sample allocations defined as the (n_{i1}, n_{i2}, n_{i3}) triplets: $(0,0,9),(0,1,8),\ldots,(9,0,0)$. We defined 39 decision models by applying the 39 $(\lambda_1, \lambda_2, \lambda_3)$ triplets of attribute decision weights: $(0.1,0.1,0.8),(0.1,0.2,0.7),\ldots,(0.8,0.1,0.1)$ and $(0.05,0.05,0.9),(0.05,0.9,0.05),(0.9,0.05,0.05)$ to Equation (4.5). For each decision model, the parameters of the Bayesian posterior distributions for the decision values (Equation (4.12)) were calculated based on the outcome of each simulated evaluation. Using these posterior distributions of ξ_i, an alternative was selected using the multinomial selection procedure and checked to determine whether this alternative was the true best. We repeated this experimental evaluation simulation a total of 1000 times. The final result was 343,200 *fcs* values: one for each of the 55 allocations, 39 decision models, and 160 subcases.

4.5.2 Frequency of Correct Selection Results

For each of the 160 subcases and each of the 39 decision models, 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 model, we defined the *relative frequency of correct selection* (*rel fcs*) for each sample allocation as the ratio of the *fcs* for that sample allocation to the *fcs* 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 4.11. Each panel of Figure 4.11 displays the *rel fcs* values for the indicated training subcase under a single decision model defined by the attribute 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_{i1} and n_{i2}, ranging from dark (low *rel fcs* values) to light (high, desirable *rel fcs* values). Note that results are only feasible in the region $n_{i2} \le 9 - n_{i1}$ since the overall sample budget $n_{i1} + n_{i2} + n_{i3} = 9$. 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.

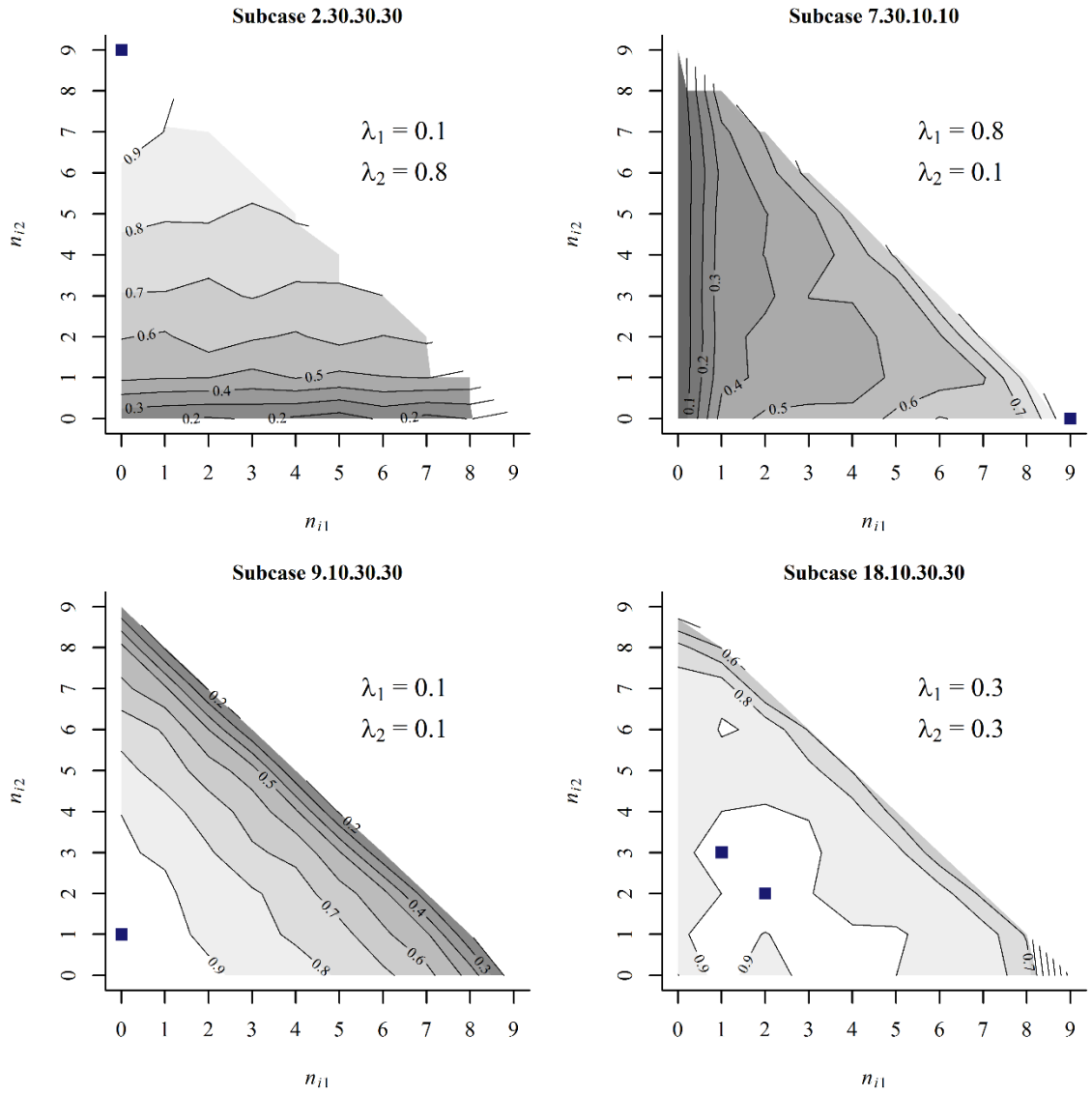

Figure 4.11: Contour plots displaying *rel fcs* as a function of n_{i1} and n_{i2} for training subcase 2.30.30.30 under decision model $(\lambda_1 = 0.1, \lambda_2 = 0.8, \lambda_3 = 0.1)$, subcase 7.30.10.10 under decision model (0.8, 0.1, 0.1), subcase 9.10.30.30 under decision model (0.1, 0.1, 0.8), 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 4.11 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 (top right panel) where a sample allocation of $n_{i1} = 9, n_{i2} = 0, n_{i3} = 0$ is indicated to be the optimal sample allocation.

If a different sample allocation is selected, say $n_{i1}=1, n_{i2}=6, n_{i3}=3$, 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 4.11 is that when the decision models are such that high weight (high λ_j 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 ($\frac{B}{m}$ samples) to the highly weighted attribute. This trend is illustrated by Subcase 9.10.30.30 under decision model (0.1, 0.1, 0.8), (bottom left panel of Figure 4.11) and was seen repeatedly throughout the 160 training subcases.

4.5.3 Allocation Procedures

We again consider the uniform allocation procedure and the proportional allocation procedure. In this three attribute study, the uniform allocation procedure is defined as $n_{i1}=n_{i2}=n_{i3}=\frac{B}{3m}$. The proportional allocation procedure is defined as $n_{i1}=round\left(\lambda_1 \frac{B}{m}\right)$, $n_{i2}=round\left(\lambda_2 \frac{B}{m}\right)$, $n_{i3}=\frac{B}{m}-n_{i1}-n_{i2}$. As an example of the proportional allocation procedure, when the attribute decision weights are $(0.1, 0.5, 0.4)$ and the budget $B/m=9$, 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 attribute 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 the previous work involving two attributes. We thus created two

"zone" allocation approaches that determined the allocation based on the attribute decision weight values of λ_1, λ_2, and λ_3.

The *3-zone allocation procedures* assigns the allocation $(n_{i1}, n_{i2}, n_{i3}) = (9,0,0)$ when the decision model has a λ_1 value near 1, assigns the allocation $(0,9,0)$ when the decision model has a λ_2 value near 1, and assigns the allocation $(0,0,9)$ when the decision model has a λ_3 value near 1. The *4-zone allocation procedure* assigns the same allocation as the three-zone allocation procedure except for decision models in which all of the attribute decision weights are between 0.2 and 0.4; to these triplets the procedure assigns the allocation $(n_{i1}, n_{i2}, n_{i3}) = (3,3,3)$. Figure 4.12 illustrates the sample allocations provided by the 3-zone and 4-zone allocation procedures as a function of decision model.

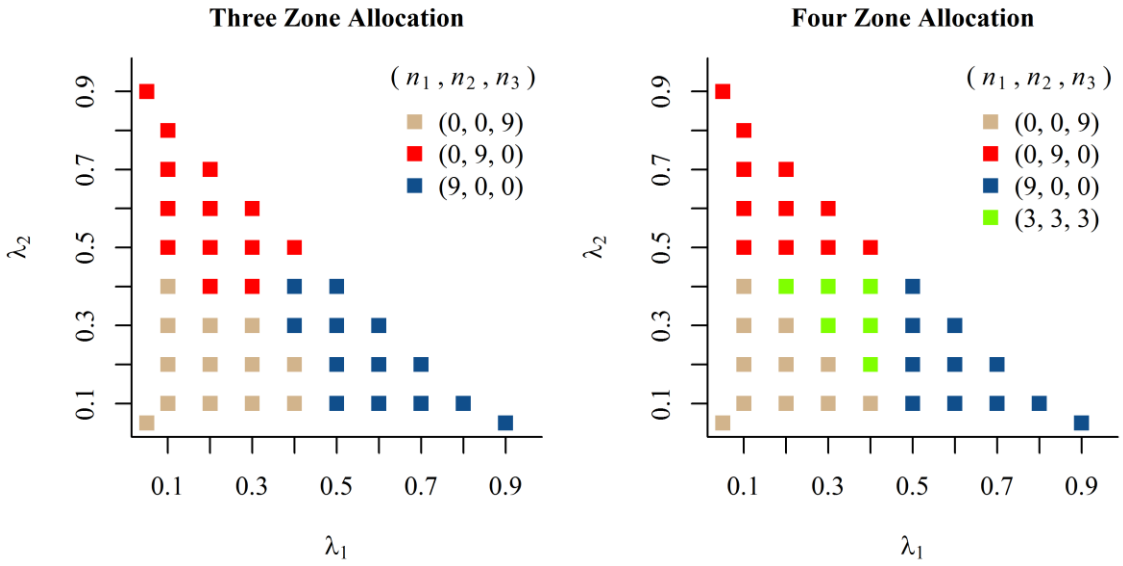

Figure 4.12: Sample allocation definitions for the 3-zone (left) and 4-zone (right) allocation procedures.

4.5.4 Comparison of Allocation Procedures

To test the sample allocation procedures, we generated 500 new concave frontiers (evaluation cases) using the construction algorithm to generate the cases in R^3 space described in Appendix C. Each evaluation case was a set of $m = 5$ randomly generated alternatives described by $k = 3$ attributes. The frontier generation process ensured that the alternatives formed a concave efficient frontier with attribute values restricted to the domain of [100, 200].

We tested the sample allocation procedures developed in Section 4.5.3 using all 500 evaluation cases and the 39 decision models used in exploring the training cases. To each of the 500 evaluation cases, we assigned a triplet of measurement error variance values, $\left(\sigma_1^2, \sigma_2^2, \sigma_3^2\right)$. The assigned σ_j values, $j = 1, 2, 3$, were independent, random draws from a uniform distribution with parameters min = 1 and max = 30. Then, for each of the 55 possible sample allocations, for each of the 39 decision models, across the 500 testing cases, we obtained a *rel fcs* value using the same simulation process used for the training cases as described in Section 4.5.1. For each evaluation case and decision model combination, we used each of the sample allocation procedures to produce a sample allocation. The *rel fcs* value from the results of the simulation for these allocations were computed. The performance of an allocation procedure, for each decision model, 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 % pointwise confidence intervals for the true relative *PCS* value based upon the normality assumption as justified by the Central Limit Theorem.

When $\lambda_1 = \lambda_2 = \lambda_3$ the proportional allocation procedure and the uniform allocation procedure provide the same sample allocation, $n_{i1} = n_{i2} = n_{i3} = \frac{B}{3m}$, and thus the procedures displayed similar performance near these decision weight values. Otherwise, the proportional allocation procedure provided *rel fcs* values that exceeded those provided by the uniform allocation procedure. Figure 4.13 illustrates these general conclusions by displaying, for each of the four allocation procedures studied, the relative frequency of correct selection averaged across all test cases and the 95 % pointwise confidence interval for the true relative *PCS* value at each decision model.

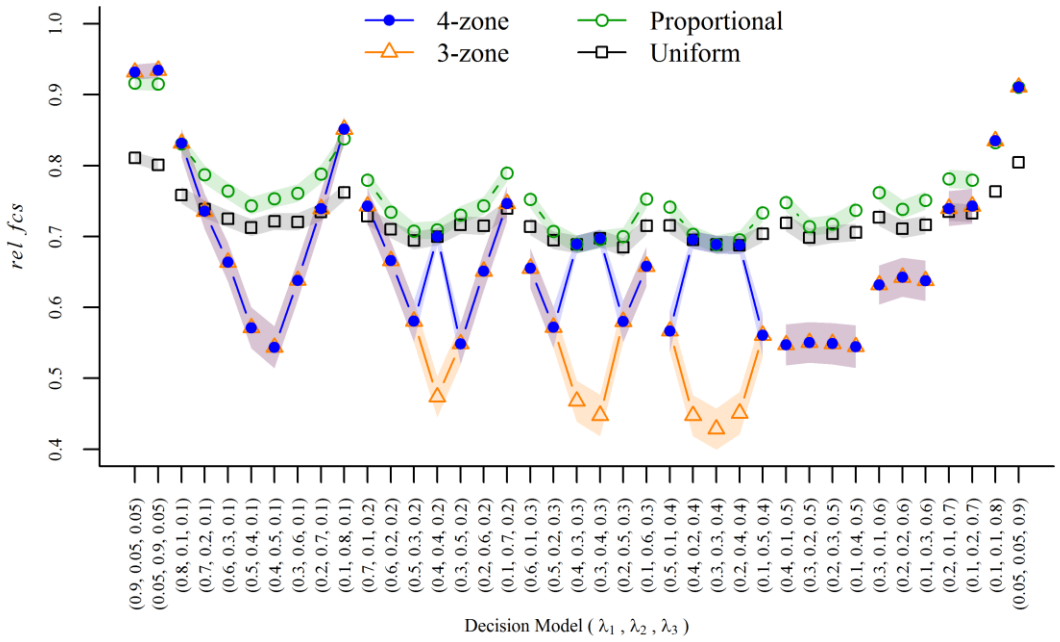

Figure 4.13: *rel fcs* values averaged across all testing cases for the 4-zone allocation for each decision weight value; shaded area represents 95 % pointwise confidence bounds for the true relative *PCS*.

The 3-zone and 4-zone allocation procedure, which leverage extreme sample allocations, provided the largest *rel fcs* value as λ_j approaches 1 for any $j = 1, 2, 3$. However, as

the λ_j move away from 1 and approach equality at $\frac{1}{3}$, the performance of the 3- and 4-zone allocation procedures rapidly decreases. With few exceptions, when $0.3 \le \lambda_j \le 0.6$ for any $j = 1,2,3$, the average *rel fcs* values provided by the 3- and 4-zone allocation procedures either the average *rel fcs* values provided the uniform and proportional allocation procedures. Only when $0.2 \le \lambda_j \le 0.4$ for all $j = 1,2,3$ does the performance of the 4-zone allocation procedure exceed that of the 3-zone allocation procedure. It is within this range of λ_j that the 4-zone allocation utilizes the uniform allocation.

4.6 An Optimal Allocation for Gaussian Measurement Model

In this section we continue our study of single-stage allocation methods for the Gaussian measurement model first introduced in Section 4.4. Here we take an analytical approach in an attempt to develop optimal allocation procedures.

4.6.1 Assumptions

While many of the assumptions are stated elsewhere in this dissertation, we provide here a comprehensive set of assumptions for this analytical study of single-stage allocation methods for the Gaussian measurement model.

1. The set of *m* distinct alternatives, $\{a_1, \ldots, a_m\}$, is provided, where *m* is a finite positive integer such that all alternatives can be assessed.
2. Each alternative is described by $k \ge 2$ attributes. The true but unknown value of attribute *j* of alternative a_i is μ_{ij}. Separate and independent measurement processes

are used in obtaining measurement data (samples) for each attribute, and for a given attribute *j*, the measurement process is the same for all $i = 1,\ldots,m$ alternatives. The measurement data are collected in a single experimental effort (single-stage). n_{ij} is the number of samples observed for attribute *j* of alternative a_i. We represent the outcomes of the measurement process by the random variable $X_{ijl} = \mu_{ij} + \varepsilon_{ijl}$. The random measurement error associated with each observation, $\varepsilon_{ijl} \overset{i.i.d.}{\sim} N\left(0, \sigma_j^2\right)$, $i = 1,\ldots m; l = 1,\ldots,n_{ij}$, and σ_j^2 is known. It follows that $X_{ijl} \sim N\left(\mu_{ij}, \sigma_j^2\right)$ with realizations $x_{ij1},\ldots,x_{ijn_{ij}}$. Note that it is common in the field of metrology for the error associated with a continuous measurand to be modeled with a Gaussian distribution (Joint Committee for Guides in Metrology, 2008) and for the variance of this distribution to be well characterized and assumed known (Lee P. M., 1997).

3. Also provided is a decision model, $\xi_i = f\left(\mu_{i1},\ldots,\mu_{ik}\right) = \sum_{j=1}^{k} \lambda_j v_j\left(\mu_{ij}\right)$, that reflects the decision-maker's preference structure and combines the multiple attribute values to produce a decision value, ξ_i, for each alternative a_i. The decision model is a multiple attribute linear value model with linear individual value functions, $v_j\left(\mu_{ij}\right) = \mu_{ij}$. The decision weights, λ_j, are defined such that $\sum_{j=1}^{k} \lambda_j = 1$.
4. The total fixed experimental budget in terms of sample observations, denoted *B*, shall not be exceeded and the cost of each measurement is equivalent. Thus, *B* is the upper bound on the number of measurements that can be performed. Unlike in our empirical studies, we do not assume that the allocations be equivalent across alternatives.

5. The decision-maker seeks to maximize the probability of selecting the true best alternative (*PCS*).

4.6.2 Expected Value Selection using Maximum Likelihood Estimation

The decision-maker uses the limited number of measurements, $x_{ij1},\ldots,x_{ijn_{ij}}$, to estimate the true value, μ_{ij}, of attribute *j* of alternative a_i in support of the selection decision. Under the Gaussian measurement model where $X_{ij} \sim N\left(\mu_{ij},\sigma_j^2\right)$ and σ_j^2 known, the maximum likelihood (ML) estimator for the true attribute value μ_{ij} is the sample mean of the n_{ij} measurements, $\hat{\mu}_{ij} = \bar{X}_{ij} = \frac{1}{n_{ij}}\sum_{l=1}^{n_{ij}} x_{ijl}$ (because the measurements $x_{ij1},\ldots,x_{ijn_{ij}}$ are assumed to be i.i.d., it follows that $\bar{X}_{ij} \sim N\left(\mu_{ij},\frac{\sigma_j^2}{n_{ij}}\right)$). Since the decision model is linear (Assumption 3), it follows from the invariant property of maximum likelihood estimators (Casella & Berger, 2002) that the maximum likelihood estimators of the alternatives' true decision values are $\bar{\xi}_i = \sum_{j=1}^{k}\lambda_j\bar{X}_{ij}, i=1,\ldots,m$. And further, $\bar{\xi}_i \sim N\left(\xi_i,\sum_{j=1}^{k}\frac{\lambda_j^2\sigma_j^2}{n_{ij}}\right)$. Using maximum likelihood estimation, a decision-maker who selects an alternative using the expected value selection procedure (Section 3.4.1) will select alternative a_s such that $s = \arg\max_i \bar{\xi}_i$

.

The probability of correct selection (*PCS*) is the probability that the alternative identified for selection, a_s, is indeed the most preferred alternative (largest true decision value). The primary goal of this work was to determine, given the provided assumptions, the number of samples (measurements), n_{ij}, required to maximize the *PCS* such that $\sum_{i=1}^{m}\sum_{j=1}^{k} n_{ij}$

does not exceed the total experimental budget, *B*. This sample allocation problem can be expressed by the optimization problem in Equation (4.13).

$$\begin{aligned} &\max_{n_{ij}} PCS = P\left(a_s \text{ is actually the most preferred alternative}\right) \\ &s.t. \quad \sum_{i=1}^{m}\sum_{j=1}^{k} n_{ij} \le B \\ &\quad\; n_{ij} \ge 0 \end{aligned} \tag{4.13}$$

Without loss of generality, we assume that $\xi_1 > \xi_i, \forall i = 2,\ldots,m$. That is, alternative a_1 is truly the most preferred alternative. Given this assumption and that the decision-maker will use the expected value selection procedure using maximum likelihood estimation, we can state the more precise definition of *PCS* as provided in Equation (4.14).

$$PCS = P\left(a_s = a_1\right) = P\left(\bar{\xi}_1 > \bar{\xi}_i, \forall i = 2,\ldots m\right) = P\left\{\bigcap_{i=2}^{m}\left(\bar{\xi}_i - \bar{\xi}_1 < 0\right)\right\} \tag{4.14}$$

We can thus restate the sample allocation problem of Equation (4.13) as Equation (4.15). Note that the final constraint in Equation (4.15) is due to the requirement that $n_{ij} \ge 1, \forall i, j$ to allow for $\bar{X}_{ij}$, and subsequently, $\bar{\xi}_i$, to be computed.

$$\begin{aligned} &\max_{n_{ij}} PCS = P\left\{\bigcap_{i=2}^{m}\left(\bar{\xi}_i - \bar{\xi}_1 < 0\right)\right\} \\ &s.t. \quad \sum_{i=1}^{m}\sum_{j=1}^{k} n_{ij} \le B \\ &\quad\; n_{ij} \ge 1 \end{aligned} \tag{4.15}$$

We first consider the $m = 2$ alternatives case. The objective function in Equation (4.15) thus becomes $\max_{n_{ij}} P\left(\bar{\xi}_2 - \bar{\xi}_1 < 0\right)$ and since $\bar{\xi}_i \sim N\left(\xi_i, \sum_{j=1}^{k} \frac{\lambda_j^2\sigma_j^2}{n_{ij}}\right)$, then $\bar{\xi}_2 - \bar{\xi}_1 \sim N\left(\xi_2 - \xi_1, \sum_{i=1}^{2}\sum_{j=1}^{k} \frac{\lambda_j^2\sigma_j^2}{n_{ij}}\right)$. This yields the definition of *PCS* provided by Equation (4.16).

$$PCS = P\left(\bar{\xi}_2 - \bar{\xi}_1\right) = \Phi\left(\frac{\xi_2 - \xi_1}{\sqrt{\sum_{i=1}^{2}\sum_{j=1}^{k}\frac{\lambda_j^2\sigma_j^2}{n_{ij}}}}\right) \tag{4.16}$$

$\Phi(y)$ in Equation (4.16) denotes the standard normal cumulative distribution function evaluated at *y*. Because Φ is a monotonically increasing function and $\xi_1 - \xi_2 > 0$, maximizing *PCS* requires minimizing $\sum_{i=1}^{2}\sum_{j=1}^{k}\frac{\lambda_j^2\sigma_j^2}{n_{ij}}$. Note that this expression is the sum of the mean squared errors (MSE) of the estimators $\bar{\xi}_1$ and $\bar{\xi}_2$ (see (Casella & Berger, 2002) for discussion of MSE). Therefore, the optimization problem in Equation (4.15), with $m = 2$, is equivalent to that provided in Equation (4.17)

$$\begin{aligned} \min_{n_{ij}} z &= \sum_{i=1}^{2}\sum_{j=1}^{k}\frac{\lambda_j^2\sigma_j^2}{n_{ij}} \\ s.t. \quad & \sum_{i=1}^{2}\sum_{j=1}^{k} n_{ij} \le B \\ & n_{ij} \ge 1 \end{aligned} \tag{4.17}$$

We restated the second constraint of the nonlinear optimization problem of Equation (4.17) as $-n_{ij} \le -1$, and derived the optimal solution displayed in Equation (4.18) using the Kuhn-Tucker conditions (Winston, 2004).

$$n_{ab} = \left(\frac{\lambda_b\sigma_b}{\sum_{j=1}^{k}\lambda_j\sigma_j}\right)\frac{B}{2} \quad a = 1,2,\ b = 1,\ldots,k \tag{4.18}$$

Since the objective function, *z*, in Equation (4.17) is the sum of convex functions, then *z* is too a convex function. As the constraints of this minimization problem are linear, they are also convex. Therefore, the Kuhn-Tucker conditions are necessary and sufficient for the solution displayed in Equation (4.18) to be an optimal solution.

When $m > 2$, there is no closed-form expression for *PCS* as defined in the objective function of Equation (4.15). The solution to the $m = 2$ case suggested an approach to overcome this dilemma, so we derived a sample allocation procedure that minimized the sum of the mean squared errors of the estimators $\bar{\xi}_i, \forall i = 1,\ldots,m$ subject to the constraints provided in Equation (4.15). The solution to this general *m* alternative problem is provided by Equation (4.19).

$$n_{ab} = \left(\frac{\lambda_b \sigma_b}{\sum_{j=1}^{k} \lambda_j \sigma_j} \right) \frac{B}{m} \quad a = 1,\ldots,m, b = 1,\ldots,k \tag{4.19}$$

Note that in Equations (4.18) and (4.19), the sample allocation, n_{ab}, is dependent only on the second index, *b*, which represents the attribute. This means that for this single-stage problem, the sample allocations may vary across attribute, but are equivalent across the alternatives. While equivalent allocations across alternatives was a constraint placed upon the solutions in our empirical studies (Sections 4.3-4.5), it is interesting to note that here, equivalent allocations across alternatives is a property of the derived optimal allocation.

Because the number of measurements made on any attribute must be an integer value greater than or equal to one, and the total number of measurements must not exceed *B*, we implement the following rounding rule, which completes our definition of the sample allocation procedure using the expected value selection procedure and maximum likelihood estimation (*MLE EV allocation procedure*).

1. Calculate the n_{ij} according to Equation (4.19) for alternatives $a_i, i = 1,\ldots,m$ and attributes $j = 1,\ldots k$.
2. Calculate $n'_{ij} = \lceil n_{ij} \rceil$, where $\lceil \bullet \rceil$ is the ceiling function.

3. Calculate $r_{ij} = n'_{ij} - n_{ij}$.

4. Calculate $O = \sum_{i=1}^{m} \sum_{j=1}^{k} n'_{ij} - B$.

5. Order the $n'_{ij} \neq 1$ in decreasing order of r_{ij}. (For any j, $r_{1j} = \cdots = r_{mj}$; thus, for each *j* such that $n'_{ij} \neq 1$, the n'_{ij} are ordered in increasing order of *i*.)

6. Subtract 1 from each of the first *O* ordered n'_{ij}; these adjusted n'_{ij} are the sample allocations for alternative a_i and attribute *j*.

We first evaluated the MLE EV allocation procedure using a simulation study similar to those used in the empirical studies of Sections 4.3-4.5 that allowed for us to compute the *rel fcs*. We calculated the *rel fcs* that resulted over a range of 19 decision models when a sample budget of $B = 50$ was allocated to $m = 5$ alternatives and $k = 2$ attributes using the MLE EV allocation procedure and an alternative was selected using the expected value selection procedure. Recall that the *rel fcs* provides a performance measure of the sample allocation suggested by the allocation procedure versus the optimal allocation. This allowed for us to gauge the level of concessions made by adopting the easily computed allocation provided by Equation (4.19) rather than working with the intractable multiple integrals required for the definition of *PCS* when $m > 2$.

It was our desire to draw an evaluation conclusion that was valid over the population of all possible decision cases with $m = 5$ alternatives described by $k = 2$ attributes whose true attribute values formed a concave efficient frontier with a domain $[100, 200] \times [100, 200]$. To accomplish this, we sought a random and representative sample from this population against which the allocation procedures would be evaluated. Because we recognized that the construction methods used in the empirical studies for the

generation of the testing and evaluation cases did not sample concave frontiers from the population space with uniform probability, we turned to the rejection algorithm described in Appendix D to generate decision cases. The rejection algorithm assured that each decision case in the population had an equal likelihood of being accepted as an evaluation case. And therefore, we considered any set of evaluation cases generated using the rejection algorithm as random and representative of the population.

Using the rejection algorithm, we generated 500 concave efficient frontiers (evaluation cases), each consisting of $m = 5$ alternatives described by $k = 2$ attributes. This required generating, on average, 2,922 sets of points for each accepted concave efficient frontier. As in our previous studies of the Gaussian measurement model, the standard deviation of the measurement error for each attribute of each evaluation case was randomly generated from a $Uniform[1,30]$ probability distribution. We simulated, as described in Section 4.2, the experimental evaluation process for each evaluation case using the 11 sample allocations that provided equivalent allocations across alternatives defined as the (n_{i1}, n_{i2}) pairs: $(0,10),(1,9),\ldots,(10,0)$. Based on the outcome of the simulated evaluation process for each evaluation case under each allocation, we calculated the ML estimator for the true attribute value, $\hat{\mu}_{ij} = \bar{X}_{ij} = \frac{1}{n_{ij}} \sum_{l=1}^{n_{ij}} x_{ijl}$. For each of 19 decision models defined by the 19 unique (λ_1, λ_2) pair of attribute weights: $(0.05,0.95),(0.10,0.90),\ldots,(0.95,0.05)$, we calculated the ML estimator of the true decision values, $\bar{\xi}_i = \sum_{j=1}^{k} \lambda_j \bar{X}_{ij}$. An alternative was selected according to the expected value selection procedure and compared to the true best.

We repeated this experimental evaluation simulation a total of 10,000 times. The final result was 104,500 *fcs* values: one for each of the 11 allocations, 19 decision models, and 500 evaluation cases.

For each evaluation case, we calculated, for each decision model, the *rel fcs* provided by the MLE EV allocation procedure as the ratio of the *fcs* of its allocation to the *fcs* of the optimal allocation (assuming equivalent allocations across alternatives). At each λ_1 value, we calculated the average *rel fcs* across the 500 evaluation cases. With one being the ideal performance measure value, we observed a distribution of average *rel fcs* values that ranged from a low of 0.994 at $\lambda_1 = 0.65$, to a high of 0.998 at $\lambda_1 = 0.95$. These observations illustrated that although the MLE EV allocation procedure was not derived as an optimal solution, it does perform at a near optimal level when evaluated against the performance of sample allocations that provide equivalent allocations across alternatives.

Our second evaluation of the MLE EV allocation procedure was to compare its performance against the performance of a uniform allocation procedure that used maximum likelihood estimation and the expected value selection procedure. For reasons that will become clear in the following sections, we chose to perform this comparison using the absolute *fcs* measure as an estimate of *PCS*, rather than consider the *rel fcs*.

In general, for each evaluation case, the evaluation consisted of a number of sample measurements (as defined by the allocation procedure) being simulated and an alternative selected according to a selection procedure and compared to the true best alternative. We thus had a Bernoulli trial for each evaluation (either succeeded or failed to select true best alternative). Because the evaluation decision cases are unique, so too are the associated Bernoulli success probabilities. As an evaluation metric, we were not interested in these

individual success probabilities, but rather, the success probability (*PCS*) of the allocation procedure over the population of all decision cases. Random effects modeling (Searle et al., 2006) provides a method to account for variability amongst the evaluation cases when estimating the overall population success probability. Fleiss et al. (2003) address this "estimation of the marginal mean proportion" problem and suggest the pooled average proportion (Equation (4.20)) as an unbiased estimate.

$$\bar{p} = \frac{1}{n.} \sum_{i=1}^{m} n_i p_i \tag{4.20}$$

Here, m is the number of evaluation cases, n_i is the number of evaluations made on case i, $n. = \sum_{i=1}^{m} n_i$, and $p_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$, where y_{ij} is the Bernoulli outcome of the j^{th} evaluation of case i. Assuming n evaluations are made on each case, then $\bar{p} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{i=1}^{n} y_{ij}$ and the variance of $\bar{p}$ is provided by Equation (4.21).

$$Var(\bar{p}) = \frac{\sum_{i=1}^{m} (p_i - \bar{p})^2}{m(m-1)} \tag{4.21}$$

We note from Equation (4.21) that the variance of the estimator for the overall population success probability is independent of the number of evaluations, n, made on each case, and decreases as the number of evaluation cases, m, increases. To keep the variance of $\bar{p}$ small, we chose an evaluation experiment design that makes only a single evaluation of each of a large number (50,000) evaluation decision cases.

We also used the variance-reduction technique of common random numbers (Law, 2007) across the comparisons of the allocation procedures. To reduce the variability in the comparison of the results from the allocation procedures, for each evaluation case, a set of $N = 10$ sample measurements were randomly drawn from the measurement distributions

that defined each alternative of each attribute (with mean equal to the true attribute value and variance equal to the variance of the measurement error). These sample measurements were used in the evaluations of all of the allocation procedures as needed. For example, the first 5 sample measurements for each of the alternatives and attributes was used in the evaluation of the uniform allocation procedure.

We estimated the *PCS* for the MLE EV and uniform allocation procedures by creating 50,000 evaluation cases with $m = 5$ alternatives described by $k = 2$ attributes using the rejection algorithm. The standard deviation of the measurement error for each attribute of each evaluation case was randomly generated from a $Uniform[1,30]$ probability distribution. We simulated, as described in Section 4.2, a single experimental evaluation process for each evaluation case under the allocation provided by each allocation procedure. Based on the outcome of the simulated evaluation process for each evaluation case, we calculated $\hat{\mu}_{ij} = \bar{X}_{ij} = \frac{1}{n_{ij}} \sum_{l=1}^{n_{ij}} x_{ijl}$ and subsequently $\bar{\xi}_i = \sum_{j=1}^{k} \lambda_j \bar{X}_{ij}$, for each of 19 decision models defined by the 19 unique (λ_1, λ_2) pair of attribute decision weights: $(0.05, 0.95), (0.10, 0.90), \ldots, (0.95, 0.05)$. An alternative was selected according to the expected value selection procedure and compared to the true best. This resulted in 50,000 true or false (correct selection made) observations for each of the 19 decision models for each allocation procedure. To estimate the *PCS*, we computed the pooled average *fcs* (Equation (4.20)) and its variance (Equation (4.21)) over the 50,000 evaluation cases for each decision model and allocation procedure. Using the normality assumption as justified by the Central Limit Theorem, we computed the 95 % pointwise confidence intervals. These results are provided in Figure 4.14.

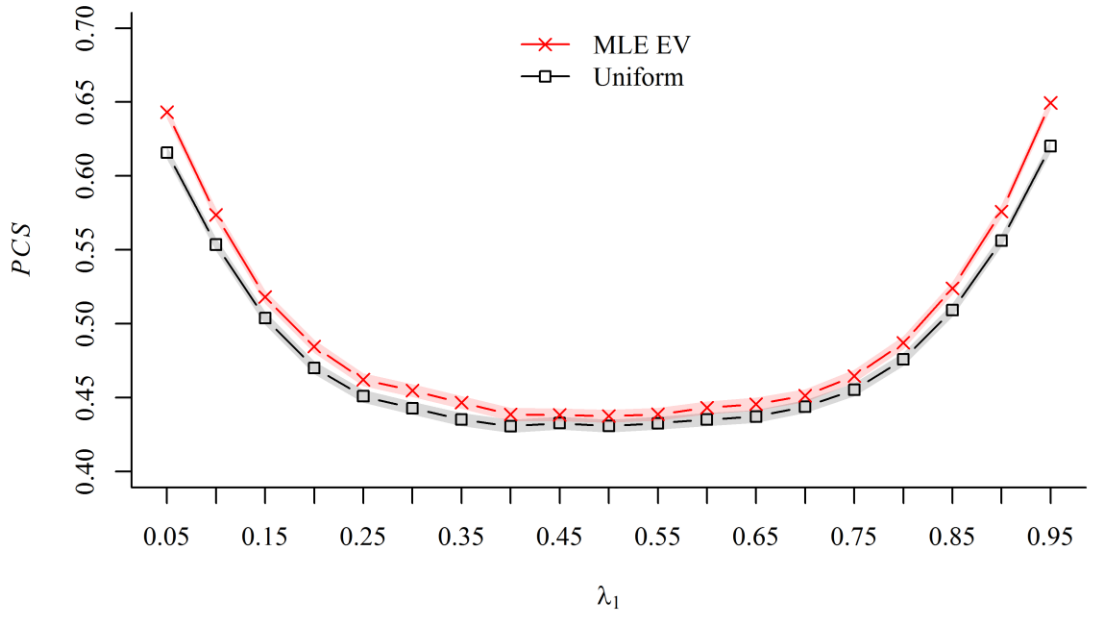

Figure 4.14: Estimated *PCS* values for the MLE EV allocation over the 50,000 evaluation cases; shaded area represents 95 % pointwise confidence bounds.

We observe from Figure 4.14 that the estimated *PCS* values provided by the MLE EV allocation procedure are only slightly larger than those provided by the uniform allocation procedure. These performance values are statistically indistinguishable (overlapping confidence bounds) when $0.4 \leq \lambda_1 \leq 0.7$. We further note that while the *rel fcs* values for the MLE EV allocation procedure was near the optimal value of one, the estimated *PCS* values were rather low. This may be a function of the sample size used in the evaluation or perhaps a result due to the limitations of the allocation and selection procedures, e.g., only allocations providing equivalent allocations across alternatives are produced.

4.6.3 Expected Value Selection using Bayesian Estimation

Under the Bayesian paradigm for estimation, we assume that, before collecting any measurement data, the decision-maker's knowledge of the unknown true attribute value, μ_{ij}, can be described by the conjugate normal prior distribution $N\left(\mu_{0ij}, \tau_{0j}^2\right)$ and *a priori*, the μ_{ij} are independent. Upon observing the normally distributed measurement data, $x_{ij1},\ldots,x_{ijn_{ik}}$, the decision-maker's knowledge of μ_{ij} is updated and presented by the normal posterior distribution (Gelman et al., 2004) in Equation (4.22).

$$\mu_{ij} \mid x_{ij1},\ldots,x_{ijn_{ij}} \sim N\left(\frac{\sigma_j^2\mu_{0ij} + n_{ij}\tau_{0ij}^2\bar{x}_{ij}}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}, \frac{\sigma_j^2\tau_{0ij}^2}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}\right) \tag{4.22}$$

Assumption 3 of Section 4.6.1 provides $\xi_i = \sum_{j=1}^{k} \lambda_j v_j\left(\mu_{ij}\right)$ for each alternative a_i. The posterior distribution of the decision value ξ_i describes the decision-maker's knowledge of the true decision value for alternative a_i, after observing n_{ij} measurements for the estimation of each of the *k* attributes, $\mathbf{x}_i = x_{i11},\ldots x_{i1n_{i1}}, x_{i21},\ldots x_{i2n_{i2}},\ldots, x_{ik1},\ldots x_{ikn_{ik}}$. The posterior distribution of ξ_i is presented in Equation (4.23).

$$\xi_i \mid \mathbf{x}_i \sim N\left(\sum_{j=i}^{k} \lambda_j \frac{\sigma_j^2\mu_{0ij} + n_{ij}\tau_{0ij}^2\bar{x}_{ij}}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}, \sum_{j=1}^{k} \lambda_j^2 \frac{\sigma_j^2\tau_{0ij}^2}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}\right) \tag{4.23}$$

Using Bayesian estimation, a decision-maker who selects an alternative using the expected value selection procedure (Section 3.4.1) will select alternative a_s such that $s = \arg\max_i \left(\sum_{j=i}^{k} \lambda_j \frac{\sigma_j^2\mu_{0ij} + n_{ij}\tau_{0ij}^2\bar{x}_{ij}}{\sigma_j^2 + n_{ij}\tau_{0ij}^2} \right)$.

As we did deriving the allocation procedure using maximum likelihood estimation (Section 4.6.2), we assumed that $\xi_1 > \xi_i, \forall i = 2,\ldots,m$ and defined the *PCS* using the Bayesian estimation of ξ_i (Equation (4.23)) and the rule to select a_s where $s = \arg\max_i \left(\sum_{j=i}^{k} \lambda_j \frac{\sigma_j^2 \mu_{0ij} + n_{ij}\tau_{0ij}^2 \bar{x}_{ij}}{\sigma_j^2 + n_{ij}\tau_{0ij}^2} \right)$. With $m = 2$ alternatives this led to the optimization problem of Equation (4.24).

$$\max_{n_{ij}} P\left(\xi_2 - \xi_1 < 0\right) = \Phi\left(\frac{\sum_{j=1}^{k} \lambda_j \frac{\sigma_j^2 \mu_{01j} + n_{1j}\tau_{01j}^2 \mu_{1j}}{\sigma_j^2 + n_{1j}\tau_{01j}^2} - \lambda_j \frac{\sigma_j^2 \mu_{02j} + n_{2j}\tau_{02j}^2 \mu_{2j}}{\sigma_j^2 + n_{2j}\tau_{02j}^2}}{\sqrt{\sum_{i=1}^{2}\sum_{j=1}^{k} \lambda_j^2 \frac{n_{ij}\tau_{0ij}^4 \sigma_j^2}{\left(\sigma_j^2 + n_{ij}\tau_{0ij}^2\right)^2}}} \right)$$

$$\text{s.t.} \quad \sum_{i=1}^{2}\sum_{j=1}^{k} n_{ij} \le B \qquad (4.24)$$

$$n_{ij} \ge 0$$

Because the unknown attribute values, μ_{ij}, in Equation (4.24) cannot be separated from the decision variables, n_{ij}, we chose to minimize the sum of the mean squared errors of the estimators $p\left(\xi_i \mid \mathbf{x}_i\right)$. We present this revised optimization problem for the general case with *m* alternatives in Equation (4.25).

$$\min_{n_{ij}} z = \sum_{i=1}^{m}\sum_{j=1}^{k} \lambda_j^2 \frac{\sigma_j^2 \tau_{0ij}^2}{\sigma_j^2 + n_{ij}\tau_{0ij}^2}$$

$$\text{s.t.} \quad \sum_{i=1}^{m}\sum_{j=1}^{k} n_{ij} \le B \qquad (4.25)$$

$$n_{ij} \ge 0$$

The optimal solution to this general *m* alternative problem, found using the Kuhn-Tucker conditions, is displayed in Equation (4.26).

$$n_{ab} = \left(\frac{\lambda_b \sigma_b}{\sum_{j=1}^{k} \lambda_j \sigma_j} \right) \left(\frac{B}{m} - \sum_{j=1}^{k} \frac{\lambda_j \sigma_j \sigma_b}{\lambda_b \tau_{0ab}^2} + \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} \frac{\sigma_j^2}{\tau_{0ij}^2} \right) \quad a = 1, ..., m, b = 1, \ldots, k \quad (4.26)$$

Here, the sample allocation is dependent only on both indices, *a* and *b*, which means that the sample allocations may vary across both attributes and alternative. This would happen only when the prior knowledge differs from one alternative to another; more specifically, when the variances of the prior distributions, τ_{0ij}^2, differ from one alternative to another.

When there is little prior knowledge of the true attribute values, μ_{ij}, that is, the prior distribution for the μ_{ij} are diffuse and the variances of the prior distributions, τ_{0ij}^2, are very large, then the limiting posterior distribution for μ_{ij} can be stated as displayed in Equation (4.27).

$$\mu_{ij} \mid x_{ij1}, \ldots, x_{ijn_{ij}} \overset{\rightarrow}{\sim} N\left(\bar{X}_{ij}, \frac{\sigma_j^2}{n_{ij}} \right) \quad as\ \tau_{0ij}^2 \to \infty \quad (4.27)$$

Further, as $\tau_{0ij}^2 \to \infty$, the allocation solution using the expected value selection procedure under Bayesian estimation displayed in Equation (4.26) converges to the solution found using maximum likelihood estimation provided in Equation (4.19).

We finalize our definition of the sample allocation procedure using the expected value selection procedure and Bayesian estimation (*Bayes EV allocation procedure*) by providing a rounding rule that assures that the number of measurements made on any attribute is a non-negative integer and that the total number of measurements does not exceed *B*.

1. Calculate the n_{ij} according to Equation (4.26) for alternatives $a_i, i = 1, \ldots, m$ and attributes $j = 1, \ldots k$.

2. Calculate $n'_{ij} = \lceil n_{ij} \rceil$, where $\lceil \bullet \rceil$ is the ceiling function.

3. Calculate $r_{ij} = n'_{ij} - n_{ij}$.

4. Calculate $O = \sum_{i=1}^{m} \sum_{j=1}^{k} n'_{ij} - B$.

5. Order the n'_{ij} in decreasing order of r_{ij}. (For any j, $r_{1j} = \cdots = r_{mj}$; thus, for each *j*, the n'_{ij} are ordered in increasing order of *i*.)

6. Subtract 1 from each of the first *O* ordered n'_{ij}; these adjusted n'_{ij} are the sample allocations for alternative a_i and attribute j.

It has been observed that, when the variances of the prior distributions, τ^2_{0ij}, are very small, the sample allocations resulting from the rounding rule may be negative. In this case the sample allocation must be set to zero.

As we did in Section 4.6.2, we performed a computational experiment to evaluate the Bayes EV allocation procedure over a range of 19 decision models when a sample budget of $B = 50$ was allocated to $m = 5$ alternatives and $k = 2$ attributes and an alternative was selected using the expected value selection procedure. Because there is no constraint requiring equal allocations across the alternatives in the Bayes EV allocation procedure, the number of possible sample allocations can grow very large. The "*k*-part compositions of *n*" problem from the field of combinatorial mathematics provides us with the binomial coefficient $\binom{n+k-1}{n}$ as the number of distinct allocations of the *n* objects among *k* bins, where the number of objects in any bin can include zero (Feller, 1950). For each of our evaluation cases we have $B = 50$ samples to be allocated among $mk = 5 \times 2 = 10$ alternative

and attribute "bins" which results in $\binom{50+10-1}{50} \approx 12.6 \times 10^9$ possible allocations. Since we cannot possibly evaluate all 12.6 trillion allocations for each evaluation case to determine the optimal allocation, we cannot compute the *rel fcs*. Thus, we simply present the results as absolute *fcs* values; an estimate of the *PCS*.

Using the absolute *fcs* measure as an estimate of the *PCS*, we compared the performance of the Bayes EV allocation procedure to the performance of a uniform allocation procedure that used Bayesian estimation and the expected value selection procedure. We used the same 50,000 evaluation cases and variance-reduction techniques that were used to compare the MLE EV to the uniform allocation procedure that used maximum likelihood estimation and the expected value selection procedure in Section 4.6.2. We assumed that the decision-maker's a priori knowledge of the attribute values, μ_{ij}, were modeled by the $N\left(\mu_{0ij}, \tau_{0j}^2\right)$ prior distribution with $\mu_{0ij} = 150$ and $\tau_{0j}^2 = 35^2$ for all attributes, for all alternatives, for all evaluation cases. We simulated, as described in Section 4.2, a single experimental evaluation process for each evaluation case under the allocation provided by each allocation procedure. Based on the outcome of the simulated evaluation process for each evaluation case, we calculated the parameters of the Bayesian posterior distributions for the decision values (Equation (4.23)), for each of 19 decision models defined by the 19 (λ_1, λ_2) pair of attribute decision weights: $(0.05, 0.95), (0.10, 0.90), \ldots, (0.95, 0.05)$. An alternative was selected according to the expected value selection procedure and compared to the true best. This resulted in 50,000 true or false (correct selection made) observations for each of the 19 decision models for each allocation procedure. To estimate the *PCS*, we computed the pooled average *fcs* (Equation (4.20)) and its variance (Equation (4.21)) over

the 50,000 evaluation cases for each decision model and allocation procedure. Using the normality assumption as justified by the Central Limit Theorem, we computed the 95 % pointwise confidence intervals. These results are provided in Figure 4.15.

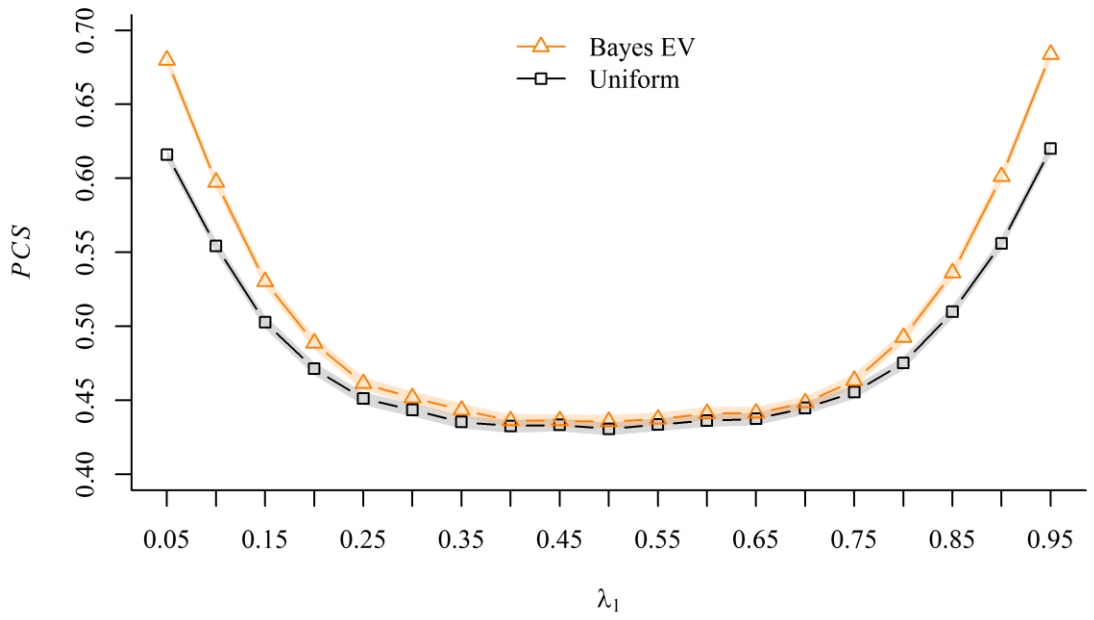

Figure 4.15: Estimated *PCS* values for the Bayes EV allocation over the 50,000 evaluation cases; shaded area represents 95 % pointwise confidence bounds.

We observe from Figure 4.15 that the *PCS* values provided by the Bayes EV allocation procedure are larger than those provided by the uniform allocation procedure when $\lambda_1 < 0.3$ and $\lambda_1 > 0.75$, but otherwise the performance values are statistically indistinguishable (overlapping confidence bounds). As the difference between λ_1 and 0.5 increases, so too does the difference in the performance of the allocation procedures.

We end this section with a comparison of the estimated *PCS* values resulting from the Bayes EV and the uniform allocation procedures that used Bayesian estimation and the

expected value selection procedure to the estimated *PCS* values resulting from the MLE EV and the uniform allocation procedures that used ML estimation and the expected value selection procedure (Section 4.6.2). The estimated *PCS* values (pooled average *fcs*) and the 95 % pointwise confidence intervals for each allocation procedure are presented in Figure 4.16.

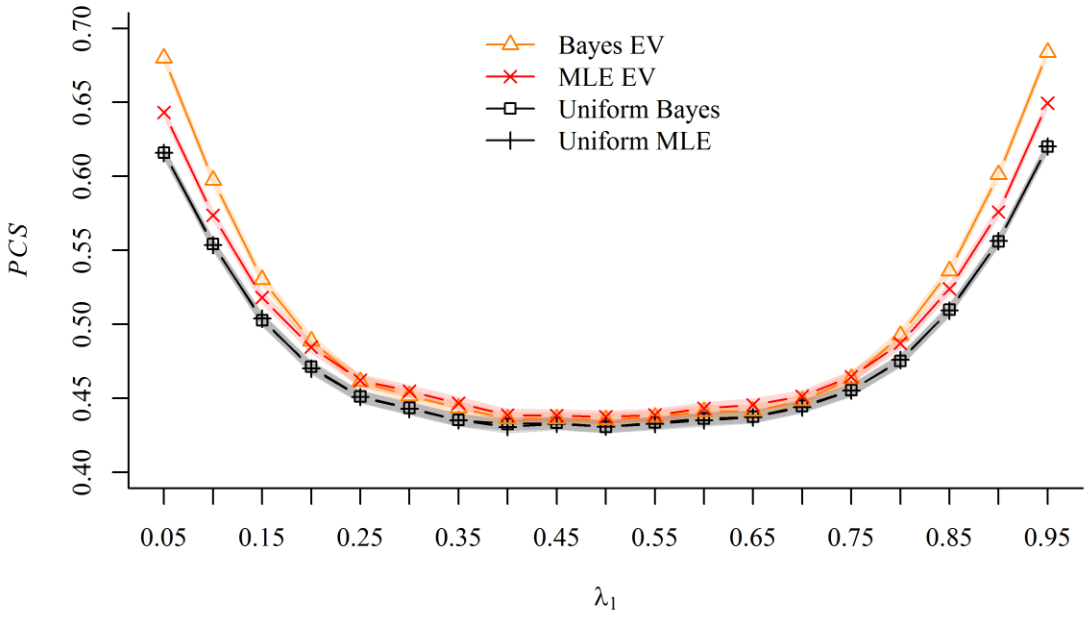

Figure 4.16: Estimated *PCS* values for the Bayes EV and MLE EV over the 50,000 evaluation cases; shaded area represents 95 % pointwise confidence bounds.

We recognize that the Bayesian estimation approach adds information about the attribute values in the form of a prior dissertation. The *PCS* values that resulted when the uniform allocation procedure was used with ML estimation and those that resulted when the uniform allocation procedure was used with Bayesian estimation are nearly identical (as shown by the black overstruck plotting characters of Figure 4.16). This implied that under

the uniform allocation procedure, the information added by the Bayesian priors, given the nonspecific (same for all alternatives and attributes) and diffuse nature of their distributions, had minimal impact on the *PCS* results.

The results of Figure 4.16 do though illustrate the benefit of even the slightest amount of additional information when considering the allocation procedure. When λ_1 is near 0 or 1, we see that the *PCS* results from the Bayes EV allocation procedure are distinguishably larger than those provided by the MLE EV allocation procedure. We ascribe this to the fact that near these extreme values of λ_1, the prior information allows the Bayes EV allocation procedure to allocate all observations to a single attribute while using only the prior information to estimate the other. This maximizes the information gained about the exceedingly important attribute. With the MLE EV allocation procedure at least one observation must be allocated to each attribute. This impact in information gained may be large with only 10 observations made on each alternative.

4.7 Sequential Allocation for Gaussian Measurement Model

The work presented in the previous sections of this chapter considered single-stage sample allocation plans for multiple attribute selection decisions where the complete allocation plan is determined before any samples are collected. Here we consider a sequential allocation approach where the experimental effort is divided into stages, each consisting of a single sample. Within each stage, the decision-maker determines, based on his current

knowledge, which single alternative and attribute pair to sample next. This process is repeated until the experimental budget is exhausted and a selection decision is made (Figure 4.17).

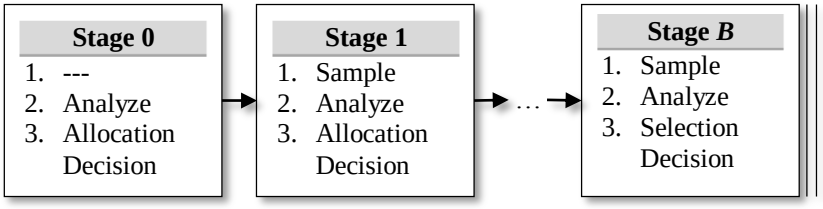

Figure 4.17: Sequential allocation procedure.

In stage $t = 0,\ldots,B-1$, the sequential allocation procedure analyzes the available information and identifies the alternative and attribute to sample in stage $t+1$ (that is, the next sample is allocated to that alternative and attribute). Let $n_{ij}(t)$ be the number of observations and $\mathbf{x}_{ij}(t) = x_{ij1},\ldots,x_{ijn_{ij}(t)}$ be the data collected in stages $1,\ldots,t$ for alternative a_i and attribute j (note that $n_{ij}(0) = 0, \forall i, j$). Let $\mathbf{x}_i(t) = \mathbf{x}_{i1}(t),\ldots,\mathbf{x}_{ik}(t)$ and $\mathbf{X}(t) = \mathbf{x}_1(t),\ldots,\mathbf{x}_m(t)$. In stage B the final sample is collected, that is, $\sum_{i=1}^{m}\sum_{j=1}^{k} n_{ij}(B) = B$, and the selection decision made.

In many multi-staged and sequential allocation procedures, an initialization set of observations are taken to gather preliminary information about each unknown value to be used in the subsequent stages of the allocation procedure (Bechhofer et al., 1995; Kim & Nelson, 2006; Chen & Lee, 2011). For each unknown value, Chen and Lee (2011) suggested an initialization set of between 5 and 20 observations, and Kim and Nelson (2006) suggested 10 or more observations. While the information gained from the initial set of

observations is valuable in making subsequent sample allocations, it may be cost prohibitive in the multiple attribute setting. Initially observing each of the multiple attributes for each of the multiple alternatives 5 or more times can quickly approach the total limited and fixed measurement budget. For this reason, we do not require an initial set of observations for our sequential allocation procedure, but rather allow the allocation procedure to dictate how to allocate all measurement samples.

4.7.1 Allocation Procedure Derivation

To develop this sequential allocation procedure for the Gaussian measurement model, we began with the assumptions provided in Section 4.6.1 (with the modification to Assumption 2 that the measurement data are collected in a sequential experimental effort). Further, we assumed that the decision-maker will describe the attribute values and decision values using Bayesian posterior distributions (Equations (4.22) and (4.23)) and that the decision-maker will use the multinomial selection procedure (Section 3.4.2). Under these assumptions, Equation (4.28) gives the probability, p_i, that alternative a_i has the largest decision value.

$$p_i = P\left(\xi_i > \xi_r, \forall r = 1,\ldots,m, r \neq i\right) \tag{4.28}$$

If the decision-maker selects alternative a_s, then the probability of correct selection is the probability, p_s, that alternative a_s has the largest decision value (Equation (4.29)). The decision-maker, who wants to maximize *PCS*, will select alternative a_s where $s = \arg\max_i p_i$.

$$PCS = p_s = P\left(\xi_s > \xi_r, \forall r = 1,\ldots,m, r \neq s\right) \tag{4.29}$$

When developing the Optimal Computing Budget Allocation (OCBA) method, Chen and Lee (2011) defined *PCS* in a manner similar to Equation (4.29) but suggested that the alternative be selected based upon its expected decision value. We see the multinomial selection procedure to be more directly aligned with the decision-maker's objective of maximizing the *PCS* than a selection based on the expected decision value. We note that the approach presented here could be modified to represent a decision-maker who plans to select the alternative with the best expected value.

Returning to the sequential allocation procedure illustrated in Figure 4.17, we note that in any stage $t = 0,\ldots,B$, the decision-maker's current knowledge of the attributes as given by the posterior distribution $p\left(\mu_{ij} \mid x_{ij1},\ldots,x_{ijn_{ij}(t)}\right)$, and subsequent current knowledge of the decision values, $p\left(\xi_i \mid \mathbf{x}_i(t)\right)$, can be obtained for all alternatives $a_i, i = 1,\ldots,m$ and all attributes $j = 1,\ldots,k$. It then follows that the probability, p_i, that alternative a_i has the largest decision value (Equation (4.28)) can be calculated and the alternative a_s where $s = \arg\max_i p_i$ identified. Thus, the *PCS* at stage *t* is as described in Equation (4.30).

$$\begin{aligned} PCS(t) &= P\left(\xi_s > \xi_r, \forall r = 1,\ldots,m, r \neq s \mid \mathbf{X}(t)\right) \\ &= \int_{\xi_s > \xi_r, \forall r=1,\ldots m, r\neq s} \cdots \int p\left(\xi_1,\ldots,\xi_m \mid \mathbf{X}(t)\right) d\xi_1 \ldots d\xi_m \end{aligned} \tag{4.30}$$

The joint posterior probability distribution of $\xi_1,\ldots,\xi_m$ at stage *t*, $p\left(\xi_1,\ldots,\xi_m \mid \mathbf{X}(t)\right)$, is the product of the individual marginal distributions (probability density of Equation (4.23) given $\mathbf{x}_i(t)$). This follows from the conditional independence of the measurements, X_{ijl}, and the prior independence of the true attribute values, μ_{ij}.

Simply stated, Equation (4.30) provides, given the decision-maker's knowledge at stage *t*, the probability that the selected alternative, a_s, has the largest decision value. While we chose to use the multinomial selection procedure that identified a_s as the alternative that provides the maximum *PCS*, other selection procedures may be used to identify a_s with $PCS(t)$ again calculated according to Equation (4.30).

To make the sample allocation decision at stage *t*, we note that the next sample, $x_{ijn_{ij}(t+1)}$, observed from alternative and attribute pair (a_i, j), will lead to a new *PCS* value (see Figure 4.18). Although the value of the sample and the subsequent new *PCS* cannot be known until the observation is made, the probability distribution of each can be described based upon the decision-maker's current knowledge. The distribution of the new observation is described by its posterior predictive distribution (Press, 1988; Ntzoufras, 2009) with density $p\left(x_{ijn_{ij}(t+1)} \mid \mathbf{x}_i(t)\right) = \int_{-\infty}^{\infty} p\left(x_{ijn_{ij}(t+1)} \mid \mu_{ij}\right) p\left(\mu_{ij} \mid \mathbf{x}_i(t)\right) d\mu_{ij}$. Given our normality assumptions and Bayesian framework, the predictive distribution of $x_{ijn_{ij}(t+1)}$ is provided by the normal distribution (Lee P. M., 1997) in Equation (4.31).

$$x_{ijn_{ij}(t+1)} \mid \mathbf{x}_i(t) \sim N\left(\frac{\sigma_j^2 \mu_{0ij} + n_{ij}(t)\tau_{0ij}^2 \bar{x}_{ij}(t)}{\sigma_j^2 + n_{ij}(t)\tau_{0ij}^2}, \sigma_j^2 + \frac{\sigma_j^2 \tau_{0ij}^2}{\sigma_j^2 + n_{ij}(t)\tau_{0ij}^2}\right) \tag{4.31}$$

For each of the *mk* alternative and attribute pairs of the allocation decision illustrated by the decision tree in Figure 4.18, assuming that the multinomial selection procedure is used, the expected *PCS* in stage *t* + 1 if attribute *j* of alternative a_i is sampled can be calculated according to Equation (4.32).

$$E\left(PCS_{ij}\left(t+1\right)\right)=$$

$$\int_{-\infty}^{\infty}\left(\max_{\forall q=1,\ldots,m}\int_{\substack{\xi_q>\xi_r,\,r\neq q\\ \forall r=1,\ldots,m}}\cdots\int p\left(\xi_1,\ldots,\xi_m\mid\mathbf{X}\left(t\right),x_{ijn_{ij}\left(t+1\right)}\right)d\xi_1\ldots d\xi_m\right)p\left(x_{ijn_{ij}\left(t+1\right)}\mid\mathbf{x}_i\left(t\right)\right)dx_{ijn_{ij}\left(t+1\right)}$$

(4.32)

For a given attribute *j* and alternative a_i, the inner multiple integral of Equation (4.32) provides, given the decision-maker's knowledge at stage *t* and the knowledge from a potential new observation, $x_{ijn_{ij}\left(t+1\right)}$, the probability that alternative a_q has the largest decision value. The maximum such probability over all *m* alternatives (i.e., $q=1,\ldots,m$) provides the *PCS* according to the multinomial selection procedure. Multiplying the *PCS* by the posterior predictive probability density and integrating over the possible values of $x_{ijn_{ij}\left(t+1\right)}$ provides the expected *PCS* in stage *t* + 1.

The sequential allocation approach allocates the sample in stage *t* + 1 to the alternative and attribute pair that yields the maximum $E\left(PCS_{ij}\left(t+1\right)\right)$.

Upon collecting the final observation in stage *B*, the sequential allocation approach calculates the probability p_i that alternative a_i has the largest decision value (Equation (4.28)) and identifies the selected alternative a_s where $s=\arg\max_i p_i$ (the multinomial selection procedure).

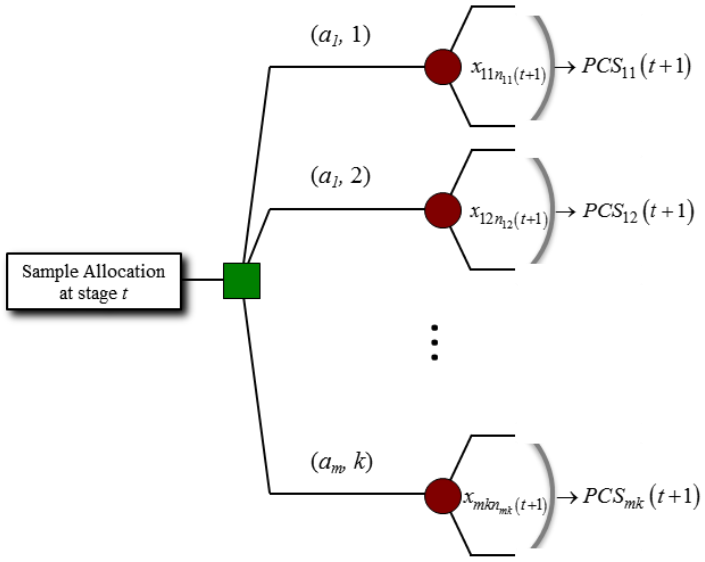

Figure 4.18: Sample allocation decision to occur at each experimental stage.

We summarize the sequential allocation procedure as follows:

1. Assign $t = 0$.
2. Calculate $E\left(PCS_{ij}\left(t+1\right)\right)$ for each attribute and alternative pair, $\left(a_i, j\right)$.
3. Identify the attribute and alternative pair with the largest $E\left(PCS_{ij}\left(t+1\right)\right)$ as the pair to be next sampled.
4. Assign $t = t + 1$.
5. Sample the attribute and alternative pair identified in Step 3.
6. Update the posterior distributions $p\left(\xi_i \mid \mathbf{x}_i\left(t\right)\right)$, based on the sample collected in Step 5.
7. Repeat steps 2 to 6 until the experimental budget is exhausted.
8. Use the multinomial selection procedure to select an alternative.

4.7.2 Numerical Experiments

To investigate the behavior of the sequential allocation procedure, we conducted a pilot study of numerical experiments. We developed 20 concave efficient frontiers (decision cases) where each decision case included $m = 5$ alternatives each characterized by $k = 2$ attributes. The attribute values associated with each alternative were randomly generated from the domain [100, 200] subject to the constraints necessary for non-dominance and concavity using the rejection algorithm described in Section 4.6.2. The attribute values for these 20 decision cases are displayed in Figure 4.19. Each attribute of each decision case was assigned a known standard deviation of the measurement error from a $Uniform[1,30]$ distribution. We assumed that the decision-maker's a priori knowledge of the attribute values, μ_{ij}, were modeled by the $N(\mu_{0ij}, \tau_{0j}^2)$ conjugate prior distribution with $\mu_{0ij} = 150$ and $\tau_{0j}^2 = 35^2$ for all attributes, for all alternatives, for all decision cases. We considered 19 decision models defined by the (λ_1, λ_2) pair $(0.05, 0.95), (0.1, 0.9), \ldots, (0.95, 0.05)$ and the overall experimental budget was limited to $B = 50$ sample measurements.

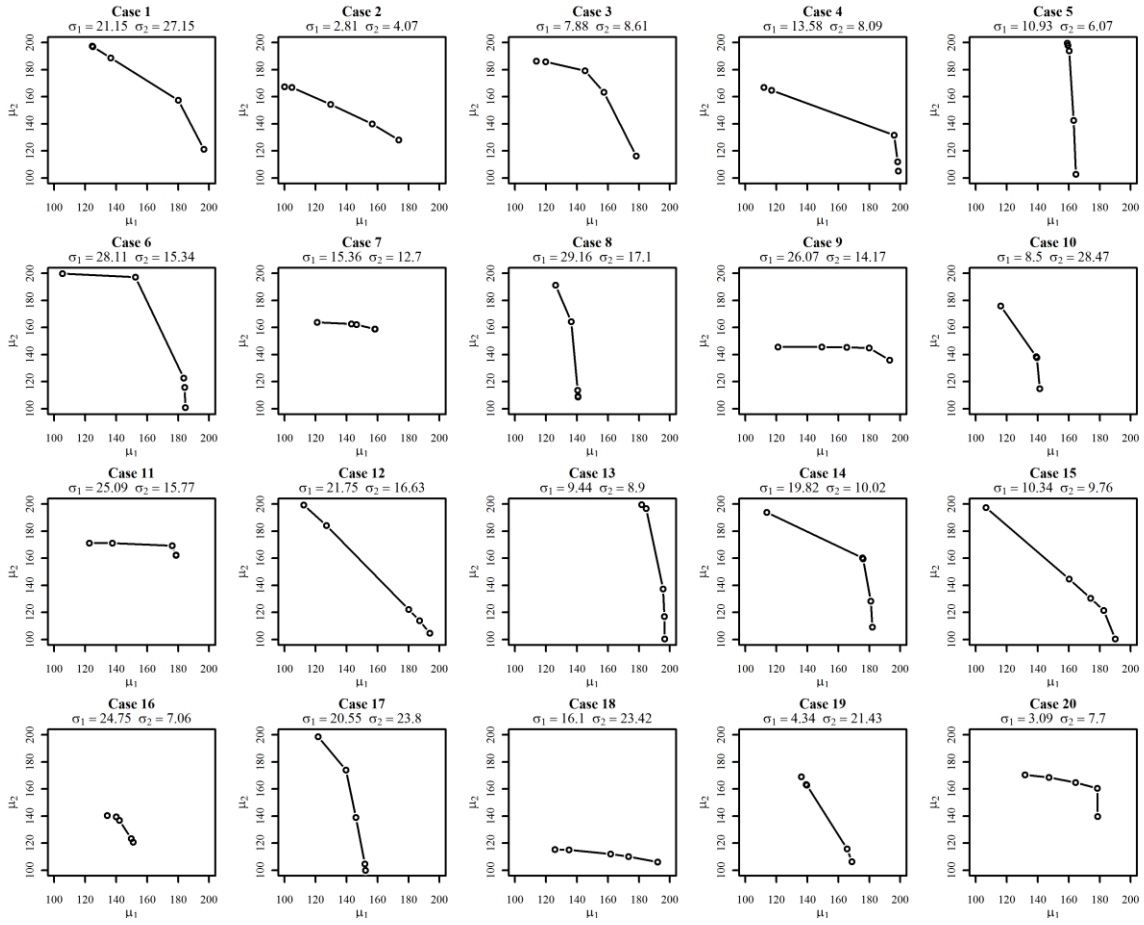

Figure 4.19: Decision cases for sequential allocation pilot study.

In each sampling stage, the multiple integral in Equation (4.32) is computed to provide the expected probability of correct selection if the next sample to be observed were from attribute *j* of alternative a_i. As there is no tractable analytical solution to Equation (4.32) when $m > 2$, we used Monte Carlo simulation to obtain an estimated solution. In implementing the Monte Carlo simulation, a number of random replicates are taken to estimate the outermost integral in Equation (4.32), and another number of random replicates are taken to estimate each of the remaining integrals; we term these numbers of Monte Carlo simulation replicates *MCrep.predict* and *MCrep.compare*, respectively. Recognizing the impact that these two variables may have on the performance of the sequential allocation

approach, we considered four combinations of Monte Carlo replication size in the pilot experiment (Table 4.4).

Table 4.4: Number or Monte Carlo replicates for sequential allocation pilot study.

Combination	*MCrep.predict*	*MCrep.compare*
1	100	1000
2	100	10000
3	1000	1000
4	1000	10000

For each decision case, decision model, and Monte Carlo replication combination, we simulated the selection decision with allocations made according to the sequential allocation procedure. At each stage the simulated sample measurement was a random draw from a normal distribution with a mean equal to the true attribute value and a standard deviation equal to the measurement error for the alternative and attribute pair identified to be sampled. Upon collecting the 50^{th} sample measurement, an alternative was selected according to the multinomial selection procedure. This process was repeated 100 times and the frequency of correct selection was defined to be the proportion of times that the selected alternative was the true best alternative. These results are displayed in Figure 4.20.

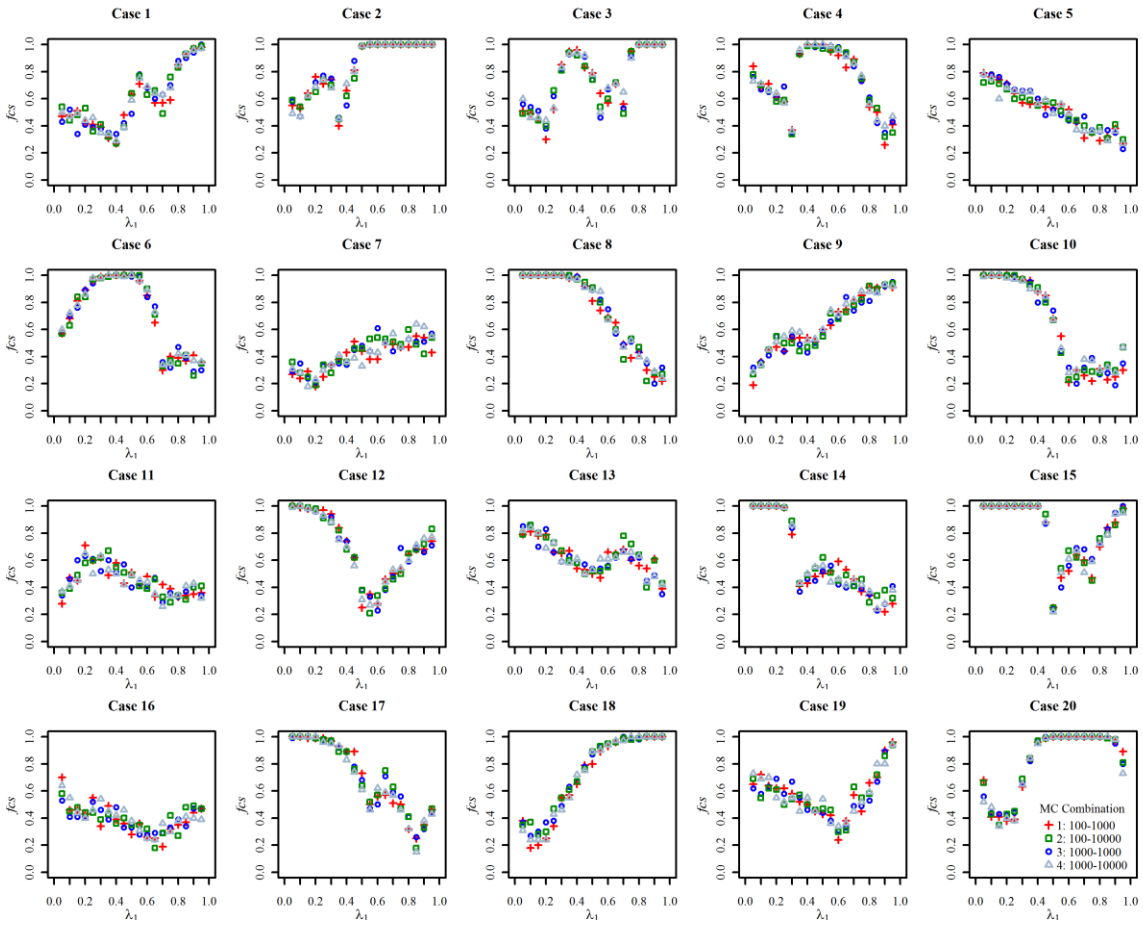

Figure 4.20: *fcs* results for the sequential allocation pilot study. The plotting characters present the four different levels of Monte Carlo simulation replication considered.

The first observation that we made from the results displayed in Figure 4.20 was that the number of Monte Carlo replicates had little impact on the *fcs* values. The results from an ANOVA model that adjusted for the decision case and decision model further supported this observation by providing the insignificant p-value of 0.98 for an effect due to the Monte Carlo replication combination.

We then observed three general trends amongst the 20 pilot study decision cases: (1) an increasing trend in *fcs* as λ_1 increased, e.g., case 9; (2) a decreasing trend in *fcs* as λ_1 increased, e.g., case 5; and (3) one or more inflection points in the *fcs* values over λ_1, e.g.,

case 12. These trends appear to depend on the shapes of the efficient frontiers. The efficient frontier for Case 9 is nearly horizontal, thus it is very difficult to distinguish alternatives based on an uncertain attribute 2 value, resulting in low *fcs* values for decision models that place large importance on attribute 2 (i.e., small λ_1 values). As the importance of attribute 1 (λ_1) increases it becomes easier to discern the alternatives as the alternatives are reasonably spaced in the attribute 1 dimension. The efficient frontier for Case 5 is nearly vertical and thus an opposite argument holds. The efficient frontier for Case 12 has an overall slope of nearly -1. Thus, when the decision model places equal weight on attribute 1 and attribute 2 ($\lambda_1 = \lambda_2 = 0.5$), it is very difficult to discern the alternatives as many alternatives have similar decision values. But as the decision model places greater weight on either of the alternatives, it becomes easier to discern an optimal alternative. These conjectures are reflected in the *fcs* results where *fcs* values are high when λ_1 is small or large, and *fcs* values reach a low point near $\lambda_1 = \lambda_2 = 0.5$. Similar observations and casual relationships were observed to exist between frontier slopes and inflection points in *fcs* results for other pilot study cases.

We expected that the more sequential stages completed (samples observed), the better the quality of the decision (higher *fcs*). By noting, at the end of each sequential stage, the alternative that would have been selected under the multinomial selection procedure, we were able to calculate the *fcs* as a result of each stage. These results are presented in Figure 4.21 for the Monte Carlo replication combination 3 (1000, 1000). The horizontal line on each plot at *fcs* = 0.2 represents the results expected from a random draw (equal probability) from 5 alternatives.

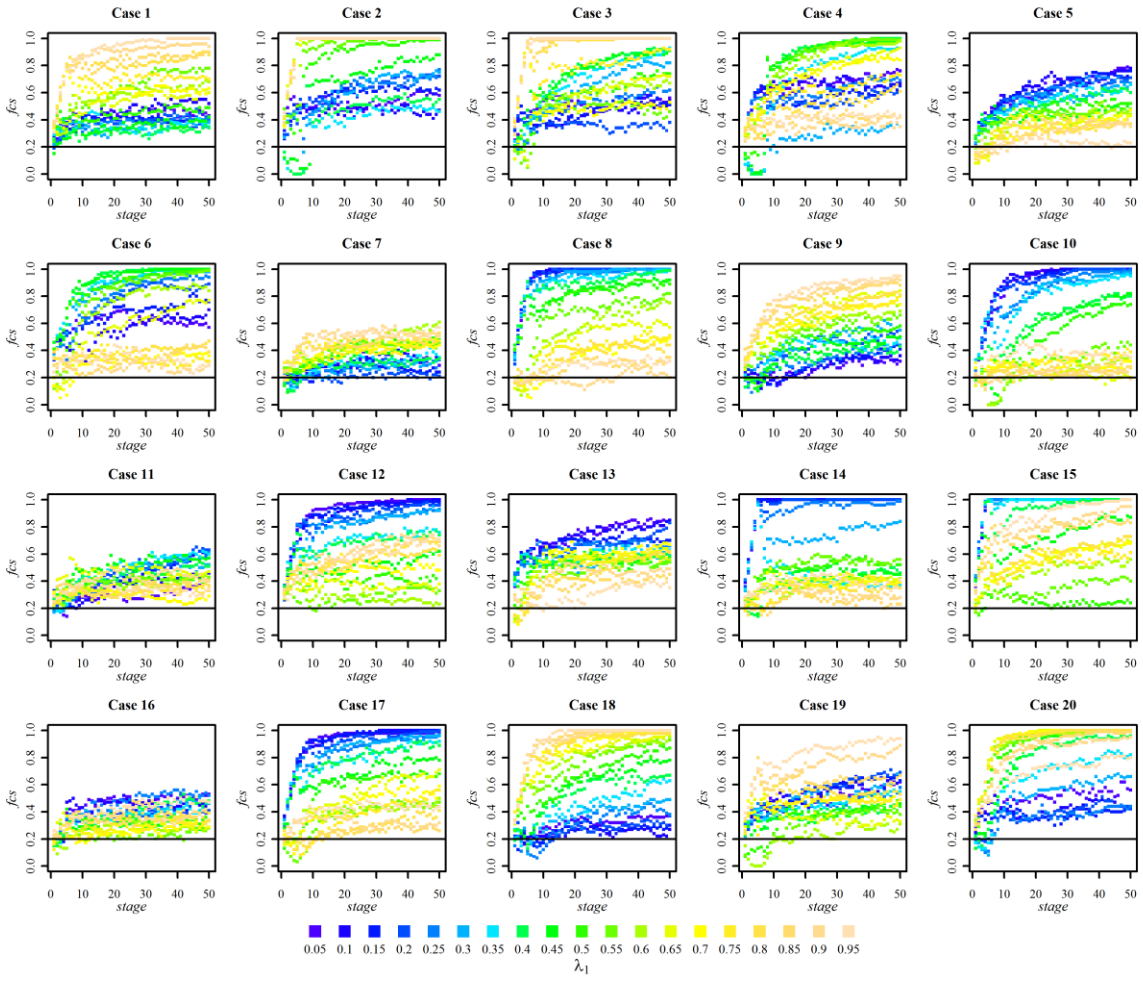

Figure 4.21: *fcs* values at each sequential stage under Monte Carlo replication combination 3.

We observe from Figure 4.21 that in general, the *fcs* does indeed increase as the number of stages (sample measurements) increases. We see however, that the rate of increase is largely dependent on both the decision case and the decision model. Consider, for example, the results from Case 8. Decision models with small λ_1 values achieve *fcs* values near 1 within only a dozen sample measurements. Decision models with large λ_1 values, however, never provide *fcs* values much larger than 0.2.

We also observed from Figure 4.21 that it was not uncommon for the sequential allocation approach to provide *fcs* values worse than a random draw (< 0.2) in early stages

when the sample size was small (< 10). This behavior was predominantly observed for decision cases under decision models that provided multiple alternatives with nearly identical true decision values (one being the true best) and those alternatives were not well separated in attribute value space. In this case, the measurement distributions are very similar across the alternatives and thus there is approximately a 50 % chance that an observation made on any of these alternatives will shift the resulting posterior distribution in the positive direction, increasing the likelihood of the alternative being selected. Because all attributes and alternatives began with the same prior probability distribution, the first sample allocation was influenced only by the attributes' decision weights and measurement errors, but otherwise was assigned at random across the alternatives. This leads to a 20 % chance (1 in 5) that the first sample is allocated to the alternative that is the true best. As a result, the probability that the resulting posterior distribution of the true best alternative is shifted in the positive direction, increasing its likelihood of being selected, is approximately 10 %. With approximately 90 % probability, the resulting posterior distributions are such that the true best alternative appears either inferior to, or indistinguishable from, the other similar alternatives. Furthermore, the alternative with the most favorable posterior distribution in any stage is more likely to be allocated the next sample (the one observed in the next stage). This perpetuates the observed low *fcs* value over the early stages until enough data is collected to begin to distinguish the nearly identical decision values. The sequential approach did however provide results reflective of a self-correcting behavior as the *fcs* most always increased beyond that expected from a random draw as the number of sample measurements increased.

4.7.3 Evaluation Experiment

Similar to the work described elsewhere in this chapter, we evaluated the sequential allocation procedure using a simulation study which calculated, as an estimate of the *PCS*, the frequency of correct selection that resulted when the multinomial selection procedure was applied to the allocations of the sample budget of $B = 50$ made according to the sequential allocation procedure over a range of 19 decision models and across a number of decision (evaluation) cases. These estimated *PCS* results were compared to the estimated *PCS* results obtained using the multinomial selection procedure applied to allocations made according to the uniform allocation and the proportional allocation (defined in Section 4.3.1), and to the estimated *PCS* results obtained using the expected value selection approach applied to allocations made according the Bayes EV allocation (defined in Section 4.6.3). Again, because there is no constraint requiring equal allocations across the alternatives in the sequential allocation procedure and the number of possible sample allocations is extremely large, we simply present the results as absolute *fcs* values.

We used the same 50,000 evaluation cases and variance-reduction techniques that were used in the evaluations of the MLE EV and Bayes EV allocation procedures (Section 4.6). We also applied the variance-reduction technique of common random numbers within the implementation of the sequential allocation procedure. At each sampling stage *t*, Monte Carlo simulations were used to estimate the multiple integrals required in the calculation of the $mk = 10$ $E\left(PCS_{ij}\left(t+1\right)\right)$ values (Equation (4.32)) for each of the 19 decision models. Recall that for each decision model, the 10 $E\left(PCS_{ij}\left(t+1\right)\right)$ values were compared to one another and the alternative and attribute pair with the largest $E\left(PCS_{ij}\left(t+1\right)\right)$ value

was selected to receive the sample allocation in sampling stage $t + 1$. To reduce the variability within this comparison, for sampling stage t, the same set of Monte Carlo simulation samples was used in all 10 calculations of $E\left(PCS_{ij}\left(t+1\right)\right)$ for all 19 decision models.

From the pilot study results presented in Section 4.7.2, we concluded that among the Monte Carlo replication combinations considered, no effect was observed on the *fcs* results in the implementation of the sequential allocation method. We chose to use 1000 Monte Carlo simulation replicates to estimate the outermost integral in the $E\left(PCS_{ij}\left(t+1\right)\right)$ calculation (Equation (4.29)) and 1000 Monte Carlo simulation replicates to estimate each of the remaining integrals. Due to the large number of decision cases and the computational effort needed to estimate $E\left(PCS_{ij}\left(t+1\right)\right)$ at every stage in the sequential sample allocation approach, parallel computing with distributed memory on the University of Maryland's High Performance Computing cluster *Deepthought2* was used to perform this evaluation experiment. Using 200 CPU cores, it required approximately 44 hours to evaluate all 50,000 evaluation cases using all of the considered allocation procedures.

To summarize, we used the 50,000 random evaluation cases generated by the rejection algorithm for the evaluation studies of Section 4.6. Each evaluation case had $m = 5$ alternatives described by $k = 2$ attributes with true attribute values from the domain [100, 200]. The standard deviation of the measurement error for each attribute of each evaluation case was randomly generated from a $Uniform\left[1,30\right]$ probability distribution. Nineteen decision models with $\left(\lambda_1,\lambda_2\right)$ pairs $\left(0.05,0.95\right),\left(0.10,0.90\right),\ldots,\left(0.95,0.05\right)$ were considered. Four sample allocation procedures: uniform, proportional, Bayes EV, and sequential, were used to allocate the overall sample budget of $B = 50$ to the alternatives and attributes of each

evaluation case. One thousand Monte Carlo samples were used in estimating the outermost integral and each of the inner integrals in the $E\left(PCS_{ij}\left(t+1\right)\right)$ calculation for the sequential allocation procedure. Common sample measurements were used across the evaluations of the allocation algorithms, and common Monte Carlo samples were used across the alternative and attribute pair in the implementation of the sequential allocation procedure. The multinomial selection procedure was used with the allocations resulting from the uniform, proportional, and sequential allocation procedure. The expected value selection procedure was used with the allocations resulting from Bayes EV allocation procedure. As a measure of performance of each allocation procedure at each decision model, we used an estimate of the overall probability of correct selection (*PCS*) presented as the pooled average *fcs* value over the 50,000 single-evaluation (Bernoulli trial) results of the evaluation cases.

4.7.4 Comparison of Allocation Procedures

As described in the previous section, we performed a simulation study to compare the performance of the sequential, uniform, proportional, and Bayes EV allocation procedures. For each allocation procedure and each decision model we estimated the *PCS* by computing the pooled average *fcs* (Equation (4.20)) and its variance (Equation (4.21)) over the 50,000 evaluation cases. Using the normality assumption as justified by the Central Limit Theorem, we computed the 95 % pointwise confidence intervals. These results are provided in Figure 4.22.

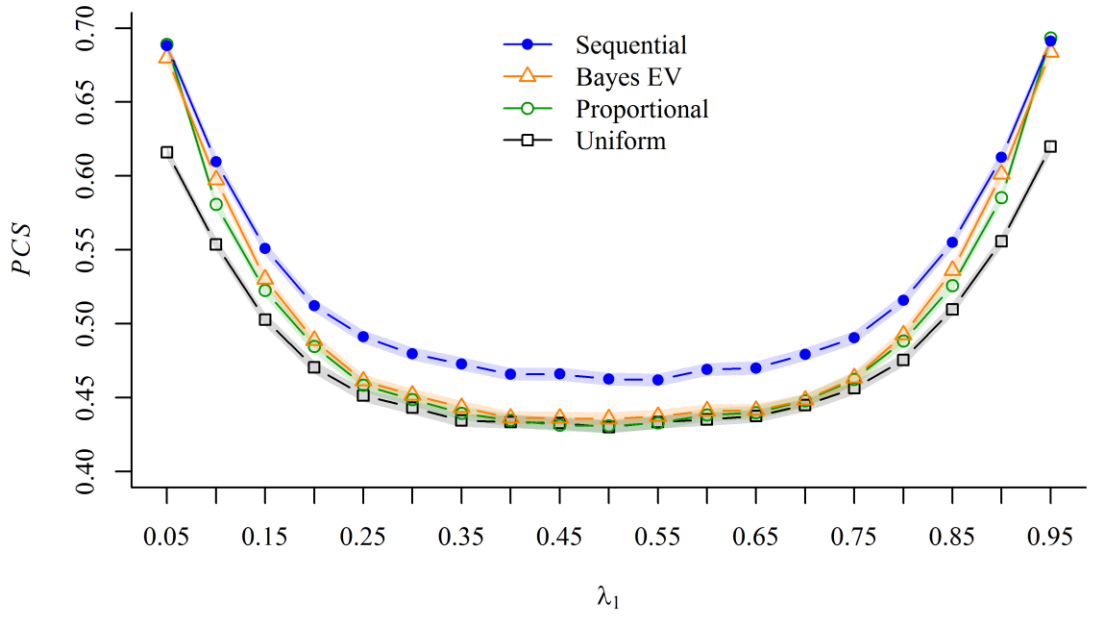

Figure 4.22: Estimated *PCS* values for the sequential allocation over the 50,000 evaluation cases; shaded area represents 95 % pointwise confidence bounds.

We observed from Figure 4.22 that the performance (*PCS*) of all allocation procedures varied in the same way as the attribute decision weights varied: the *PCS* values were lower when the decision weights were nearly equal ($\lambda_1 \approx \lambda_2 \approx 0.5$) and increased when the decision weight for one attribute was much larger than the other (λ_1 or λ_2 near 1). This phenomenon was observed in our previous studies and may occur because, when the weights are nearly equal, there are multiple alternatives with similar decision values.

We also observed in Figure 4.22 that the sequential allocation procedure outperforms the uniform allocation procedure (larger *PCS* values with non-overlapping confidence intervals) across all decision models. This same observation holds when comparing the sequential allocation procedure to the proportional and Bayes EV allocation procedures for

$0.1 < \lambda_1 < 0.9$. When the decision weight for one attribute was much larger than the other (λ_1 or λ_2 near 1), the performances of the sequential, proportional, and Bayes EV allocation procedures were not clearly distinguishable. To emphasize this comparison, we calculated the performance of the sequential allocation procedure relative to each of the other three allocation procedures (Figure 4.23). We propagated the uncertainties of the *PCS* estimates (Bevington & Robinson, 2002) and again displayed them as 95 % pointwise confidence intervals.

Figure 4.23 displays the sequential allocation *PCS* values relative to the *PCS* values for the other allocation procedures evaluated. We observed that only at $\lambda_1 = 0.05$ and $\lambda_1 = 0.95$ does the performance of any of the allocation procedures match or exceed that of the sequential allocation procedure (relative $PCS \leq 1$). At $\lambda_1 = 0.05$ and $\lambda_1 = 0.95$, the performance of the sequential procedure and the proportional allocation procedure are equivalent. Across all other decision models (λ_1 values), the performance of the sequential procedure exceeded that of the proportional allocation procedure by more than 5 % (1.06 average relative *PCS*).

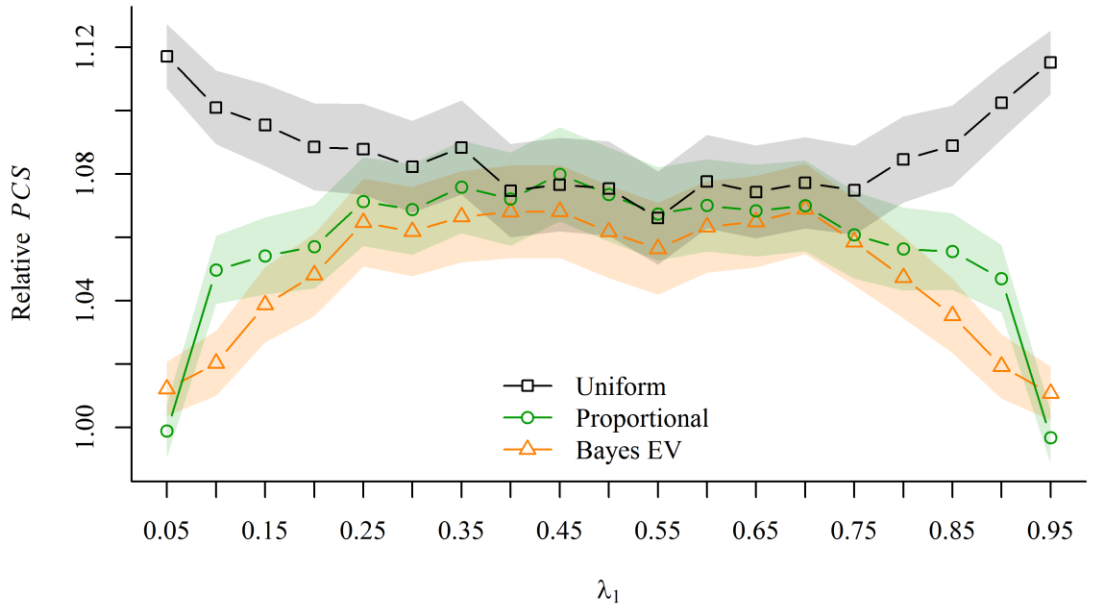

Figure 4.23: Sequential allocation procedure *PCS* results relative to the results of the uniform, proportional, and Bayes EV allocation procedures; shaded area represents 95 % pointwise confidence bounds.

On average, across all decision models, the performance of the Bayes EV allocation procedure was nearest the performance of the sequential allocation (1.05 average relative *PCS*). It did not, however, match or exceed that of the sequential allocation approach for any λ_1 value.

When compared to the commonly used uniform allocation procedure, the sequential allocation procedure produced *PCS* values that are approximately 10 % larger (1.09 average relative *PCS*) across all decision models. Although the lower 95 % confidence bounds on the uniform relative *PCS* values was as low as 1.05, these results show that the sequential allocation procedure produced significantly better results than the uniform allocation procedure.

4.8 Summary

Through a series of simulation studies that investigated various decision situations, we demonstrated that, when collecting data to support a selection decision with multiple uncertain attributes, the allocation of the experimental budget across the multiple attributes and alternatives does impact the decision-maker's ability to select the true best alternative, i.e., the probability of correct selection. In this chapter we developed and evaluated a number of procedures to allocate the samples of a fixed experimental budget to the multiple attributes and alternatives of a multiple attribute selection decision with the goal of maximizing the probability of correct selection.

We first considered the multiple attribute selection decision where the attributes were success probabilities that were evaluated using separate and independent series of Bernoulli trials (e.g., pass-fail testing). We formulated several single-stage sample allocation procedures based on insights gained from an empirical simulation study of decision cases with $m = 5$ alternatives and $k = 2$ attributes. The uniform allocation procedure allocated an equivalent number of samples to each attribute of each alternative. The proportional allocation procedure allocates samples to attributes proportional to the attribute's decision weight. If, in addition to knowledge of the decision model, the decision-maker had knowledge of the shape of the efficient frontier formed by the attribute values, then he could implement the step allocation procedure. Through a second simulation study, we evaluated each of these allocation procedures by considering its frequency of correct selection (*fcs*) performance when applied to 100 new decision cases with $m = 5$ alternatives and $k = 2$ attributes. The results of our evaluation showed that the step allocation procedure

provided *rel fcs* that were nearly 25 % higher, on average (0.93 versus 0.75), than those provided by the common uniform allocation procedure.

We next conducted studies that considered the multiple attribute selection decision where the attributes were continuous measurands that were evaluated using separate and independent measurement processes. We developed single-stage allocation procedures for both the $k = 2$ attributes and $k = 3$ attributes decision cases. For the $k = 2$ attributes study, we developed and evaluated the performance of the uniform, proportional, 1-parameter step, and 2-parameter step allocation procedures. From the *fcs* and *rel fcs* results provided by an evaluation against 500 new decision cases with $m = 5$ alternatives and $k = 2$ attributes, we again saw that the additional information provided by the decision-maker proved beneficial. That is, the 2-parameter step outperformed the 1-parameter step, which outperformed the proportional, which outperformed the uniform allocation procedure in terms of *rel fcs*. The best performing 2-parameter step allocation procedure provided *rel fcs* that were 15 % higher, on average (0.92 versus 0.80), than those provided by the common uniform allocation procedure.

In the $k = 3$ attributes study of the Gaussian measurement model, we developed and evaluated the performance of the uniform, proportional, 3-zone, and 4-zone allocation procedures. The evaluation of these allocation procedures using 500 new decision cases with $m = 5$ alternatives and $k = 3$ attributes provided *fcs* and *rel fcs* results that demonstrated the inferiority of these extreme allocation approaches. The zone allocation procedures were outperformed by the proportional allocation procedure for all decision models investigated, and they were outperformed by the uniform allocation procedure for most decision models investigated.

We conducted an analytical study of the Gaussian measurement model. We began by developing an optimal, single-stage allocation procedure for $m = 2$ alternatives when maximum likelihood estimation was used in conjunction with the expected value selection procedure. We extended the ideas used in developing the optimal $m = 2$ allocation procedure to the derivation of a general, near-optimal, m alternatives procedure using maximum likelihood estimation and the expected value selection procedure. We called this general procedure the MLE EV allocation procedure. In an evaluation using 500 decision cases with $m = 5$ alternatives and $k = 2$ attributes, we found that for all decision models the MLE EV provided, on average, *fcs* results that were within 1 % of *fcs* values provided by the optimal allocation (over all allocations that provide equivalent allocations across alternatives).

We extended the analytical study of the Gaussian measurement model to develop a general single-stage allocation procedure using Bayesian estimation and the expected value selection procedure. Though the Bayes EV allocation procedure was not derived to be optimal, an evaluation using 50,000 decision cases with $m = 5$ alternatives and $k = 2$ attributes showed it to perform well when compare to the uniform allocation using Bayes estimation and expected value selection, the MLE EV, and the uniform allocation using ML estimation and expected value selection. We saw that the additional information provided by the Bayesian prior distribution benefited the allocation procedure for some decision models.

In our final study of the Gaussian measurement model, we developed and evaluated a general sequential allocation procedure using Bayesian estimation and the multinomial selection procedure. Using 50,000 decision cases with $m = 5$ alternatives and $k = 2$ attributes, we evaluated the performance of the sequential allocation procedure and compared it to the

performance of the uniform and proportional allocations procedures using Bayesian estimation and multinomial selection, and to the performance of the Bayes EV allocation procedure using Bayesian estimation and the expected value selection procedure. Although the sequential allocation requires additional overhead in terms of evaluating a multiple integral at each stage, the resulting *fcs* values were, on average, 10 % larger than those provided the common uniform allocation and, on average, 5 % larger than the *fcs* values provided by the proportional and Bayes EV allocation procedures.

In the evaluation of each of these allocation procedures, we observed an increase in the probability of correct selection when additional information was used in the allocation procedure and in the evaluation of the selection decision model. This information was either in the form of knowledge of the shape of the efficient frontiers, or provided by Bayesian prior distributions on the attribute values. While providing nonlinearity and general angle measures for the frontier shape may be difficult in most decision situations, providing some general information or bounding for the attribute values in terms of a Bayesian prior distribution may prove to be a beneficial and easier task.

The allocation procedures developed in this chapter were based on the set of assumptions provided in Section 4.1, with one of the more restrictive assumptions being the identity form of the individual value functions, i.e., $v_j\left(\mu_{ij}\right) = \mu_{ij}$. When all attributes are measured on the same scale, using the same measurement units, and are of comparable size, this assumption is reasonable. However, when attributes are measured on different scales, a change of scale transformation may be necessary. When the individual value function is a linear transformation, e.g., $v_j\left(\mu_{ij}\right) = \alpha_j \mu_{ij} + \beta_j$, where α_j is the scaling factor and β_j is the location constant for attribute *j*, the approaches presented in this dissertation can be

modified appropriately with straightforward changes. The most notable change that must be accounted for is due to the change in the definition of the decision value, ξ_i, and more specifically, the change in the probability distribution of ξ_i. To account for this change in the MLE EV allocation procedure (Equation (4.19)) and the Bayes EV allocation procedure (Equation (4.26)), λ is replaced by $\gamma = \alpha\lambda$ (with appropriate subscripts) . Both the mean and the variance parameters of the posterior probability distributions of the decision values that underlie the sequential allocation procedure (Equation (4.23) must be appropriately updated, but otherwise the sequential allocation procedure remains as defined. While it is believed that nonlinear value functions will have greater impact on the analyses and results of this chapter, further work is required to understand the extent of their impact.

Chapter 5 Conclusions, Contributions, and Future Work

In this dissertation we studied the problem of allocating a fixed experimental budget across multiple attributes and alternatives in a selection decision where the results of the experimental evaluations lead to uncertain estimates of the true attribute values. Subject to the budget constraint, the goal was to identify an allocation approach that maximized the probability of correct selection. Through a series of simulation studies, we developed and evaluated a number of allocation procedures under both the Bernoulli measurement model (success probability attributes) and the Gaussian measurement model (continuous measurand attributes). We considered several approaches to incorporating the uncertainty about the attribute values into the selection decision model including the multinomial selection procedure.

As a baseline for comparisons, we used a uniform allocation procedure that allocated the sample budget equally across the alternatives and attributes. This simple and common approach is consistent with the principle of balance in the traditional design of experiments discipline. We observed the performance of the uniform allocation procedure to be inferior to that of allocation procedures that leveraged additional information, such as specific knowledge of the decision model or general knowledge about the attribute values. Of the allocation procedures developed, the best performing was found to be the sequential allocation procedure, which uses specific knowledge of the decision model and general knowledge of the attribute values provided by Bayesian prior distributions that is updated

with each observed sample. These observations lead us to the important conclusion that the manner in which the information gathering budget is allocated across the multiple attributes and alternatives does impact the quality of the selection made. Further, any available information should be used in developing the allocation plan and evaluating the selection decision model. We saw that even vague information in the form of nonspecific and diffuse Bayesian prior probability distributions describing the attribute values can lead to an increase in selection performance for some decision models. In practical terms, 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 common). Our work demonstrates that such a cooperative approach can improve the overall selection results of the project.

We summarize the contributions of this dissertation as follows:

1. We identified and evaluated methods to incorporate the uncertainty in the attribute value estimates into a normative model for a multiple attribute selection decision. We demonstrated the idea of propagating the attribute value uncertainty and describing the decision values for each alternative as probability distributions and used these distributions in the selection of an alternative. We noted that, by selecting the alternative that has the greatest probability of having the largest decision value, the multinomial selection procedure most closely aligns with the goal of maximizing the probability of correct selection. This is particularly true when a Bayesian approach is used to model the attribute and decision values.
2. We provided a set of allocation procedures that were developed under a variety of assumptions and decision situations. From the results of simulation experiments,

we produced evidence of the relative performance of the allocation procedures. The allocation procedures that included decision-maker knowledge outperformed those that neglected such information.

Thus, we have answered the research questions posed in Chapter 1. We expect these contributions will lead to the further advancement of selection decision and allocation methods that promote the use and recognition of the uncertainty in attribute values that results when the values are estimated based on a limited set of sample data collected from observations such as measurements, experimental evaluations, or simulation runs.

We assumed in this work that, for a given decision, all attributes were of the same type (e.g., success probabilities), were evaluated or measured on the same scale, the values of each attribute were described by simply linear individual value functions, and the cost to collect information about each attribute and alternative was the same. While these assumptions may hold for some decision situations, they will not for many. We will look to relax each of these assumptions in future work and specifically provide allocation and selection methods to handle mixed attribute error types with differing individual value functions.

Another general topic that we would like to contemplate in future work is the consideration of evaluation metrics beyond the probability of correct selection. For example, a measure of regret that indicates how far, in decision value space, a selected alternative is from the true best alternative would be of interest. Such a measure may help us to further understand what impact the proximity of the alternatives in attribute space has on the quality of the selection decision.

Simulation studies that allowed for the true best alternative to be known were the basis for the evaluations performed in this dissertation. We relied upon the several methods described in Appendix A, Appendix C, and Appendix D to generate the true attribute values for the random decision cases used in the evaluations. While using the rejection algorithm of Appendix D assured that each frontier in the population of all concave efficient frontiers with m alternatives described by $k = 2$ attributes had an equal probability of being constructed, it came with a computational cost. As noted, for $m = 5$ alternatives, the generation of an average of 2,922 sets of points was required for each accepted concave efficient frontier. This number of required sets of points increased to 3.5 million for each accepted concave efficient frontier when $m = 7$, and the algorithm was unable to return a result when $m = 8$. Further research on the development of a random set of concave efficient frontiers for the general m alternatives with k attributes case would be beneficial in further progressing this work. An immediate use of such random concave frontiers with $k > 2$ attributes would be to evaluate the general allocation procedures developed in Sections 4.6 and 4.7.

One focus of the work of this dissertation was how to allocate a provided information gathering budget to best support a selection decision. The ideas of the value of information (Clemen & Reilly, 2001) focus on whether the decision-maker should collect additional information to improve the quality of the decision outcome. Though different, there is a similarity between these topics that may be worthy of further exploration. The similarity is most apparent in the sequential decision allocation procedure that considers, at each sampling stage, whether or not to sample from each alternative and attribute pair based on the "value" that such a sampling is expected to provide in terms of probability of correct selection.

The sequential allocation procedure, which provided the most favorable results and is the most complicated allocation procedure that was studied, also raises many questions for future work. The complex relationships illustrated in Figure 4.21 between decision case, measurement error, decision model, and the *fcs* result at each stage provide for a rich topic of research. Also of interest is updating the sequential allocation procedure to allow for a sample size greater than one in each stage. By pushing this idea to the extreme of allocating all B samples in the first stage of the sequential allocation procedure would provide another single-stage allocation procedure.

Finally, we noted that the OCBA method (Chen & Lee, 2011) selected an alternative based upon the expected decision value. We believe this approach may be inconsistent with the stated goal of maximizing the probability of correct selection and view the multinomial selection procedure as a better choice for a selection procedure. We plan to further investigate these views by formulating and evaluating an OCBA-like approach using a multinomial selection procedure and comparing the results to those that resulted from numerical experiments performed by Chen and Lee.

Appendix A Constructing Frontiers with Two Attributes

Here we describe the construction algorithm used to generate a concave efficient frontier with $m = 5$ alternatives, each described by $k = 2$ attributes such as that presented in Figure A.1. For this discussion, let X and Y denote the values of attribute 1 and attribute 2. We use the ordered statistic notation $x_{[i]}$ to denote the i^{th} smallest value of attribute X.

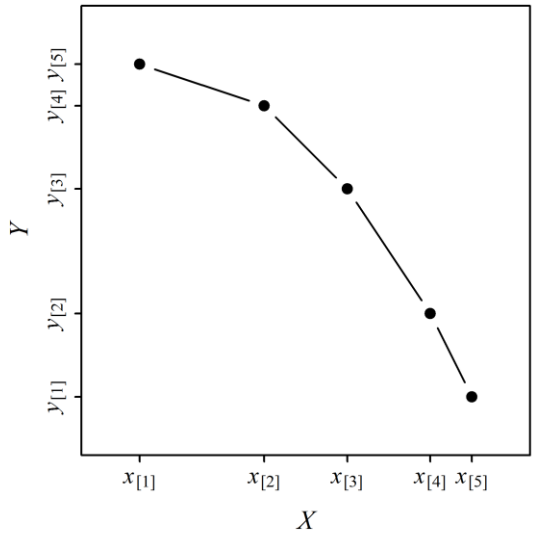

Figure A.1: Concave efficient frontier with 5 alternatives and 2 attributes.

The construction algorithm proceeds as follows:

1. Define the domain of the attribute values as $\left[X_{\min}, X_{\max}\right]$ and $\left[Y_{\min}, Y_{\max}\right]$.
2. Randomly select the distance between the smallest and largest value of attribute X by drawing a single value from a $Uniform\left[0, \left(X_{\max} - X_{\min}\right)\right]$ probability distribution. Repeat for attribute Y. Denote these values d_X and d_Y.

3. Determine the attribute values of the extreme alternatives, $\left(x_{[1]}, y_{[5]}\right)$ and $\left(x_{[5]}, y_{[1]}\right)$ by assigning a single value drawn from a $Uniform\left[0, 1-d_X\right]$ distribution to $x_{[1]}$ and $x_{[5]} = x_{[1]} + d_X$; similarly, assign a single value drawn from a $Uniform\left[0, 1-d_Y\right]$ distribution to $y_{[1]}$ and $y_{[5]} = y_{[1]} + d_Y$.

4. Define the attribute values of the third attribute, $\left(x_{[3]}, y_{[3]}\right)$ such that $L_1 \le x_{[3]} \le x_{[5]}$ and $L_1 \le y_{[3]} \le y_{[5]}$, where L_1 is the line segment with endpoints $\left(x_{[1]}, y_{[5]}\right)$ and $\left(x_{[5]}, y_{[1]}\right)$ as displayed in Figure A.2. We do this by drawing a single value *r* from a $Uniform\left[0,1\right]$ probability distribution.

    a. If $r < 0.5$, then:

        i. $x_{[3]}$ is randomly selected from a $Uniform\left[x_{[1]}, x_{[5]}\right]$ probability distribution.

        ii. Define $L_1(x) = \frac{y_{[1]} - y_{[5]}}{x_{[5]} - x_{[1]}}\left(x - x_{[1]}\right) + y_{[5]}$; $y_{[3]}$ is randomly selected from a $Uniform\left[L_1\left(x_{[3]}\right), y_{[5]}\right]$ probability distribution.

    b. If $r \ge 0.5$, then:

        i. $y_{[3]}$ is randomly selected from a $Uniform\left[y_{[1]}, y_{[5]}\right]$ probability distribution.

        ii. Define $L_1^{-1}(y) = \frac{x_{[5]} - x_{[1]}}{y_{[1]} - y_{[5]}}\left(y - y_{[5]}\right) + x_{[1]}$; $x_{[3]}$ is randomly selected from a $Uniform\left[L_1^{-1}\left(y_{[3]}\right), x_{[5]}\right]$ probability distribution.

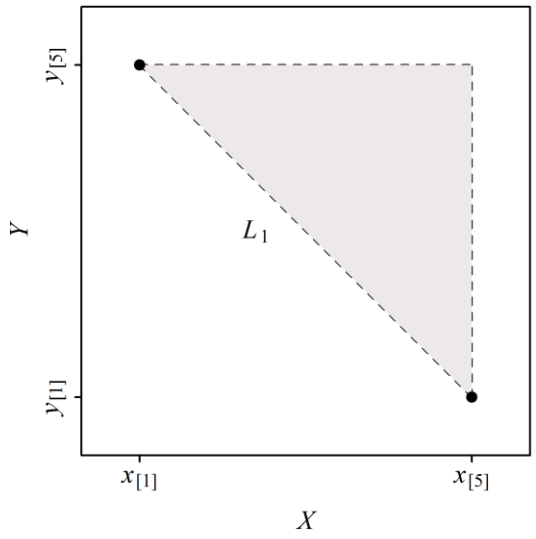

Figure A.2: Constraints for the attribute values of the third alternative.

5. Define the attribute values of the fourth and fifth alternatives, $\left(x_{[2]}, y_{[4]}\right)$ and $\left(x_{[4]}, y_{[2]}\right)$, by randomly determining the order that they are created. If $\left(x_{[2]}, y_{[4]}\right)$ is the fourth alternative created then, $L_2 \leq x_{[2]} \leq L_3$ and, if $x_{[2]} < X' = \frac{x_{[5]} - x_{[3]}}{y_{[1]} - y_{[3]}}\left(y_{[5]} - y_{[3]}\right) + x_{[3]}$, then $L_2 \leq y_{[4]} \leq y_{[5]}$, else $L_2 \leq y_{[4]} \leq L_3$. Where L_2 is the line that extends through the points $\left(x_{[1]}, y_{[5]}\right)$ and $\left(x_{[3]}, y_{[3]}\right)$ and L_3 is the line that extends through the points $\left(x_{[3]}, y_{[3]}\right)$ and $\left(x_{[5]}, y_{[1]}\right)$. These constraints are illustrated by the shaded area in the left panel of Figure A.3. It follows that the values of the fifth alternative $\left(x_{[4]}, y_{[2]}\right)$ are such that $L_3 \leq y_{[2]} \leq L_4$ and if $y_{[2]} < X'' = \frac{y_{[3]} - y_{[4]}}{x_{[3]} - x_{[2]}}\left(x_{[5]} - x_{[2]}\right) + y_{[4]}$, then $L_3 \leq x_{[4]} \leq x_{[5]}$, else $L_3 \leq x_{[4]} \leq L_4$. Where L_4 is the line that extends through the points $\left(x_{[2]}, y_{[4]}\right)$ and

$\left(x_{[3]}, y_{[3]}\right)$. These constraints are illustrated by the shaded area in the right panel of Figure A.3.

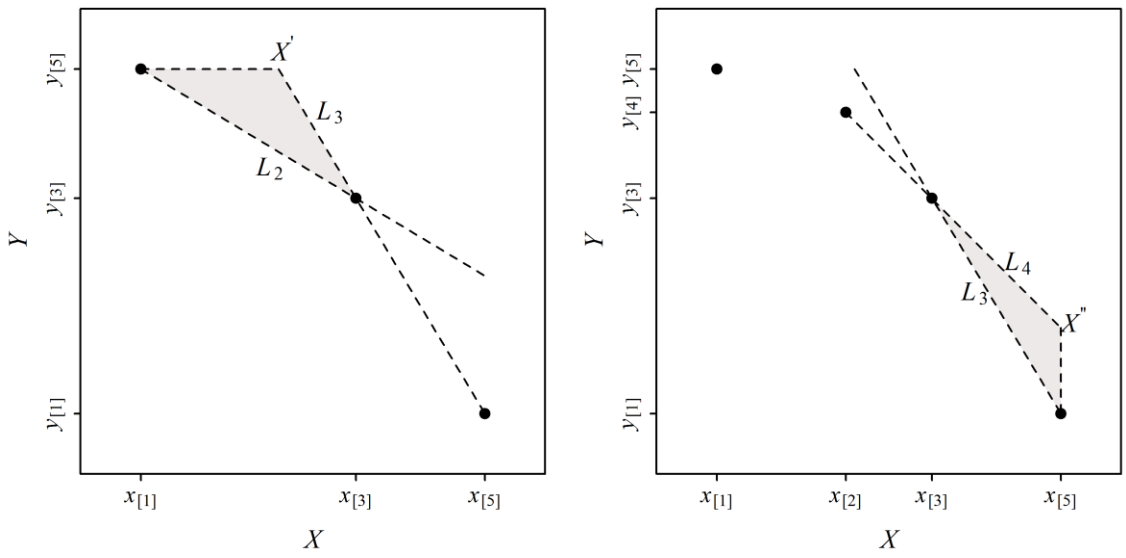

Figure A.3: Constraints for the attribute values of the fourth (left panel) and fifth (right panel) attributes, if attribute $\left(x_{[2]}, y_{[4]}\right)$ is the fourth attribute created.

If, on the other hand, $\left(x_{[4]}, y_{[2]}\right)$ is the fourth alternative created, then, $L_3 \le y_{[2]} \le L_2$ and, if $y_{[2]} < X^* = \frac{y_{[3]} - y_{[5]}}{x_{[3]} - x_{[1]}}\left(x_{[5]} - x_{[1]}\right) + y_{[5]}$, then $L_3 \le x_{[4]} \le x_{[5]}$, else $L_3 \le x_{[4]} \le L_2$. These constraints are illustrated by the shaded area in the left panel of Figure A.4. It follows that the values of the fifth alternative, $\left(x_{[2]}, y_{[4]}\right)$, are such that $L_2 \le x_{[2]} \le L_5$ and if $x_{[2]} < X^{**} = \frac{x_{[4]} - x_{[3]}}{y_{[2]} - y_{[3]}}\left(y_{[5]} - y_{[3]}\right) + x_{[3]}$, then $L_2 \le y_{[4]} \le y_{[5]}$, else $L_2 \le y_{[4]} \le L_5$. Where L_5 is the line that extends through the points $\left(x_{[3]}, y_{[3]}\right)$ and $\left(x_{[4]}, y_{[2]}\right)$. These constraints are illustrated by the shaded area in the right panel of Figure A.4.

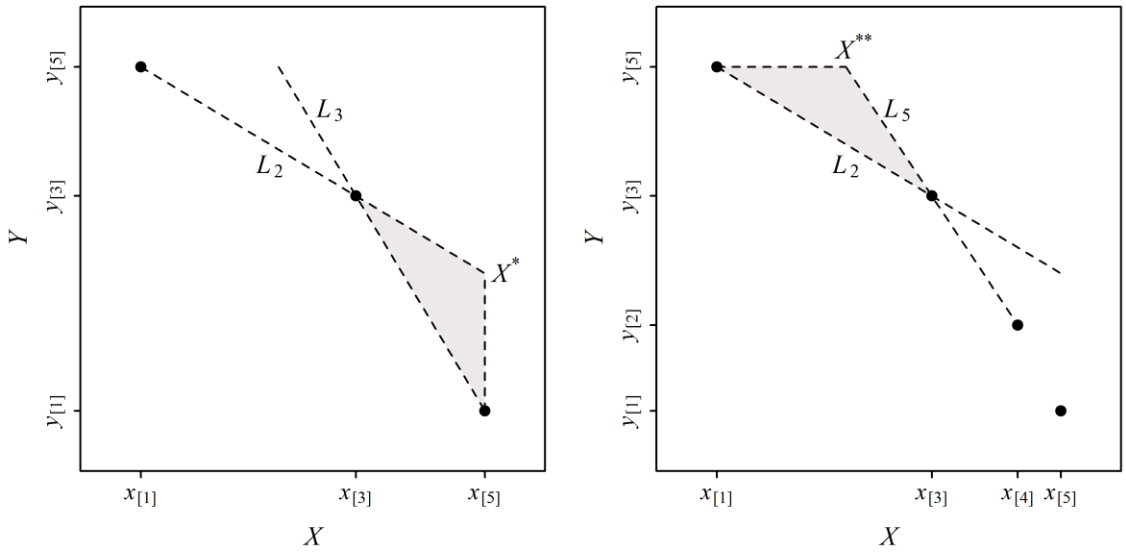

Figure A.4: Constraints for the attribute values of the fourth (left panel) and fifth (right panel) attributes, if attribute $\left(x_{[4]}, y_{[2]}\right)$ is the fourth attribute created.

We accomplished this by defining:

$$L_2(x) = \frac{y_{[3]} - y_{[5]}}{x_{[3]} - x_{[1]}}\left(x - x_{[1]}\right) + y_{[5]} \text{ and } L_2^{-1}(y) = \frac{x_{[3]} - x_{[1]}}{y_{[3]} - y_{[5]}}\left(y - y_{[5]}\right) + x_{[1]}$$

$$L_3(x) = \frac{y_{[1]} - y_{[3]}}{x_{[5]} - x_{[3]}}\left(x - x_{[3]}\right) + y_{[3]} \text{ and } L_3^{-1}(y) = \frac{x_{[5]} - x_{[3]}}{y_{[1]} - y_{[3]}}\left(y - y_{[3]}\right) + x_{[3]}$$

$$L_4(x) = \frac{y_{[3]} - y_{[4]}}{x_{[3]} - x_{[2]}}\left(x - x_{[2]}\right) + y_{[4]} \text{ and } L_4^{-1}(y) = \frac{x_{[3]} - x_{[2]}}{y_{[3]} - y_{[4]}}\left(y - y_{[4]}\right) + x_{[2]}$$

$$L_5(x) = \frac{y_{[2]} - y_{[3]}}{x_{[4]} - x_{[3]}}\left(x - x_{[3]}\right) + y_{[3]} \text{ and } L_5^{-1}(y) = \frac{x_{[4]} - x_{[3]}}{y_{[2]} - y_{[3]}}\left(y - y_{[3]}\right) + x_{[3]}$$

Randomly select a single value *u* from a $Uniform[0,1]$ probability distribution.

a. If $u \le 1/8$, then:

   i. $x_{[2]}$ is randomly selected from a $Uniform\left[x_{[1]}, x_{[3]}\right]$ probability distribution.

ii. If $x_{[2]} < L_3^{-1}\left(y_{[5]}\right)$, then $y_{[4]}$ is randomly selected from a $Uniform\left[L_2\left(x_{[2]}\right), y_{[5]}\right]$ probability distribution; else, $y_{[4]}$ is randomly selected from a $Uniform\left[L_2\left(x_{[2]}\right), L_3\left(x_{[2]}\right)\right]$ probability distribution.

iii. $x_{[4]}$ is randomly selected from a $Uniform\left[x_{[3]}, x_{[5]}\right]$ probability distribution.

iv. $y_{[2]}$ is randomly selected from a $Uniform\left[L_3\left(x_{[4]}\right), L_4\left(x_{[4]}\right)\right]$ probability distribution.

b. If $1/8 < u \leq 1/4$, then:

i. $x_{[2]}$ is randomly selected from a $Uniform\left[x_{[1]}, x_{[3]}\right]$ probability distribution.

ii. If $x_{[2]} < L_3^{-1}\left(y_{[5]}\right)$, then $y_{[4]}$ is randomly selected from a $Uniform\left[L_2\left(x_{[2]}\right), y_{[5]}\right]$ probability distribution; else, $y_{[4]}$ is randomly selected from a $Uniform\left[L_2\left(x_{[2]}\right), L_3\left(x_{[2]}\right)\right]$ probability distribution.

iii. $y_{[2]}$ is randomly selected from a $Uniform\left[y_{[1]}, y_{[3]}\right]$ probability distribution.

iv. If $y_{[2]} < L_4\left(x_{[5]}\right)$, then $x_{[4]}$ is randomly selected from a $Uniform\left[L_3^{-1}\left(y_{[2]}\right), x_{[5]}\right]$ probability distribution; else, $x_{[4]}$ is randomly selected from a $Uniform\left[L_3^{-1}\left(y_{[2]}\right), L_4^{-1}\left(y_{[2]}\right)\right]$ probability distribution.

c. If $1/4 < u \leq 3/8$, then:

i. $y_{[4]}$ is randomly selected from a $Uniform\left[y_{[3]}, y_{[5]} \right]$ probability distribution.

ii. $x_{[2]}$ is randomly selected from a $Uniform\left[L_2^{-1}\left(y_{[4]}\right), L_3^{-1}\left(y_{[4]}\right) \right]$ probability distribution.

iii. $x_{[4]}$ is randomly selected from a $Uniform\left[x_{[3]}, x_{[5]} \right]$ probability distribution.

iv. $y_{[2]}$ is randomly selected from a $Uniform\left[L_3\left(x_{[4]}\right), L_4\left(x_{[4]}\right) \right]$ probability distribution.

d. If $3/8 < u \leq 1/2$, then:

i. $y_{[4]}$ is randomly selected from a $Uniform\left[y_{[3]}, y_{[5]} \right]$ probability distribution.

ii. $x_{[2]}$ is randomly selected from a $Uniform\left[L_2^{-1}\left(y_{[4]}\right), L_3^{-1}\left(y_{[4]}\right) \right]$ probability distribution.

iii. $y_{[2]}$ is randomly selected from a $Uniform\left[y_{[1]}, y_{[3]} \right]$ probability distribution.

iv. If $y_{[2]} < L_4\left(x_{[5]}\right)$, then $x_{[4]}$ is randomly selected from a $Uniform\left[L_3^{-1}\left(y_{[2]}\right), x_{[5]} \right]$ probability distribution; else, $x_{[4]}$ is randomly selected from a $Uniform\left[L_3^{-1}\left(y_{[2]}\right), L_4^{-1}\left(y_{[2]}\right) \right]$ probability distribution.

e. If $1/2 < u \leq 5/8$, then:

i. $x_{[4]}$ is randomly selected from a $Uniform\left[x_{[3]}, x_{[5]}\right]$ probability distribution.

ii. $y_{[2]}$ is randomly selected from a $Uniform\left[L_3\left(x_{[4]}\right), L_2\left(x_{[4]}\right)\right]$ probability distribution.

iii. $x_{[2]}$ is randomly selected from a $Uniform\left[x_{[1]}, x_{[3]}\right]$ probability distribution.

iv. If $x_{[2]} < L_5^{-1}\left(y_{[5]}\right)$, then $y_{[4]}$ is randomly selected from a $Uniform\left[L_2\left(x_{[2]}\right), y_{[5]}\right]$ probability distribution; else, $y_{[4]}$ is randomly selected from a $Uniform\left[L_2\left(x_{[2]}\right), L_5\left(x_{[2]}\right)\right]$ probability distribution.

f. If $5/8 < u \leq 3/4$, then:

i. $x_{[4]}$ is randomly selected from a $Uniform\left[x_{[3]}, x_{[5]}\right]$ probability distribution.

ii. $y_{[2]}$ is randomly selected from a $Uniform\left[L_3\left(x_{[4]}\right), L_2\left(x_{[4]}\right)\right]$ probability distribution.

iii. $y_{[4]}$ is randomly selected from a $Uniform\left[y_{[3]}, y_{[5]}\right]$ probability distribution.

iv. $x_{[2]}$ is randomly selected from a $Uniform\left[L_2^{-1}\left(y_{[4]}\right), L_5^{-1}\left(y_{[4]}\right)\right]$ probability distribution.

g. If $3/4 < u \leq 7/8$, then:

i. $y_{[2]}$ is randomly selected from a $Uniform\left[y_{[1]}, y_{[3]} \right]$ probability distribution.

ii. If $y_{[2]} < L_2\left(x_{[5]}\right)$, then $x_{[4]}$ is randomly selected from a $Uniform\left[L_3^{-1}\left(y_{[2]}\right), x_{[5]} \right]$ probability distribution; else, $x_{[4]}$ is randomly selected from a $Uniform\left[L_3^{-1}\left(y_{[2]}\right), L_2^{-1}\left(y_{[2]}\right) \right]$ probability distribution.

iii. $x_{[2]}$ is randomly selected from a $Uniform\left[x_{[1]}, x_{[3]} \right]$ probability distribution.

iv. If $x_{[2]} < L_5^{-1}\left(y_{[5]}\right)$, then $y_{[4]}$ is randomly selected from a $Uniform\left[L_2\left(x_{[2]}\right), y_{[5]} \right]$ probability distribution; else, $y_{[4]}$ is randomly selected from a $Uniform\left[L_2\left(x_{[2]}\right), L_5\left(x_{[2]}\right) \right]$ probability distribution.

h. If $7/8 < u \leq 1$, then:

i. $y_{[2]}$ is randomly selected from a $Uniform\left[y_{[1]}, y_{[3]} \right]$ probability distribution.

ii. If $y_{[2]} < L_2\left(x_{[5]}\right)$, then $x_{[4]}$ is randomly selected from a $Uniform\left[L_3^{-1}\left(y_{[2]}\right), x_{[5]} \right]$ probability distribution; else, $x_{[4]}$ is randomly selected from a $Uniform\left[L_3^{-1}\left(y_{[2]}\right), L_2^{-1}\left(y_{[2]}\right) \right]$ probability distribution.

iii. $y_{[4]}$ is randomly selected from a $Uniform\left[y_{[3]}, y_{[5]} \right]$ probability distribution.

iv. $x_{[2]}$ is randomly selected from a $Uniform\left[L_2^{-1}\left(y_{[4]} \right), L_5^{-1}\left(y_{[4]} \right) \right]$ probability distribution.

Appendix B Frontier Measures

Here we define the measure of nonlinearity, *NL*, and a measure of general angle, θ, that are used to describe a concave efficient frontier with *m* alternatives, each described by $k = 2$ attributes. For this discussion, let *X* and *Y* denote the values of attribute 1 and attribute 2. For notational purposes, when considering *m* alternatives, alternative 1 will maintain the smallest value for attribute 1 and the largest value for attribute 2 when compared to the remaining alternatives. Alternative 2 will maintain the second smallest value for attribute 1 and the second largest value for attribute 2, and so on. The values for the attribute value pair will be represented as (x_i, y_i) for alternative $i = 1, \ldots, m$. Figure B.1 illustrates this alternative number convention for a concave efficient frontier with $m = 5$ attributes.

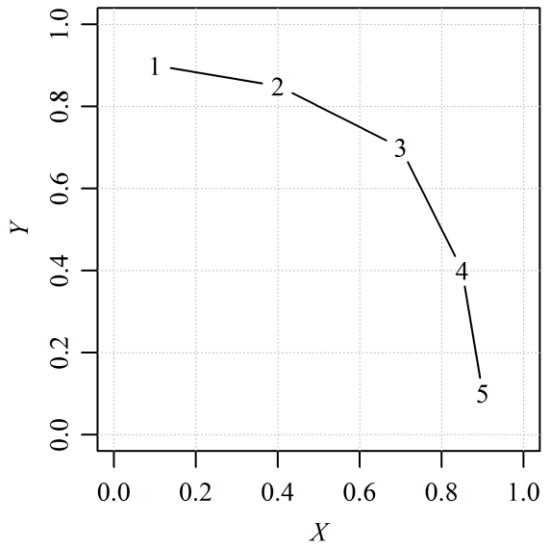

Figure B.1: Concave efficient frontier with 5 alternatives and 2 attributes.

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 value pairs (x_1, y_1) and (x_m, y_m). Specifically, all points on the piecewise linear concave curve are scaled such that the distance between (x_1, y_1) and (x_m, y_m) is 1. The nonlinearity measure is the area between the scaled piecewise linear concave curve and the line segment connecting the scaled points (x'_1, y'_1) and (x'_m, y'_m). This measure can range from 0 (linear) to 0.25 ("extreme concave" curve or "90° knee").

The algorithm to calculate the measure of nonlinearity is as follows:

1. For each alternative *i*, represented by $(x_i, y_i), i = 1, \ldots, m$, calculate:

   a. $J_i = \sqrt{(x_i - x_1)^2 + (y_i - y_1)^2}$, the distance between (x_i, y_i) and (x_1, y_1).

   b. $K_i = \sqrt{(x_i - x_m)^2 + (y_i - y_m)^2}$, the distance between (x_i, y_i) and (x_m, y_m).

   c. $L = \sqrt{(x_1 - x_m)^2 + (y_1 - y_m)^2}$, the distance between (x_1, y_1) and (x_m, y_m).

   d. Scaled points $x'_i = \dfrac{J_i^2 - K_i^2 + L^2}{2L^2}$ and $y'_i = \sqrt{\dfrac{J^2}{L^2} - \left[\dfrac{J^2 - K^2 + L^2}{2L^2}\right]^2}$.

2. For each of the $m - 1$ line segments in the piecewise linear concave curve formed by the scaled points $(x'_i, y'_i), i = 1, \ldots, m$, calculate area between the line segment and the horizontal line $y' = 0$.

3. The measure of nonlinearity, *NL*, is the sum of the areas calculated in Step 2.

The measure of general angle, θ, is defined as the acute angle formed by the line segment connecting the attribute value pairs (x_1, y_1) and (x_m, y_m) and the horizontal line $y = y_1$. This measure can range from 0° to 90°. A measure of 0° indicates that all alternatives fall on a horizontal line. A measure of 90° indicates that all alternatives fall on a vertical line.

The algorithm to calculate the measure of general angle is as follows:

1. Calculate the angle, in radians, formed by the line segment connecting the attribute value pairs (x_1, y_1) and (x_m, y_m) and the horizontal line $y = y_1$ as $\theta_R = \tan^{-1}(y_1 - y_m / x_m - x_1)$.
2. Calculate the measure of general angle as $\theta = \frac{180}{\pi}\theta_R$.

The measures of nonlinearity, *NL*, and a measure of general angle, θ, are illustrated in Figure B.2.

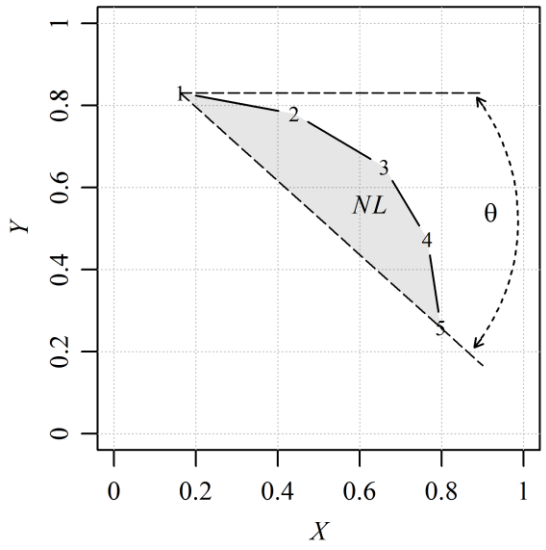

Figure B.2: Illustration of the measures of nonlinearity (scaled shaded area) and general angle, θ.

The measures of nonlinearity and overall angle are related. Piecewise linear concave curves with an overall angle measure near 0° (nearly horizontal) or near 90° (nearly vertical) will have small nonlinearity measures. A piecewise linear concave curve consisting of *m* points provides a maximum nonlinearity measure when all points fall on one of two line segments and these line segments form a 90° angle. The maximum nonlinearity measure, as a function of the overall angle measure, can thus be described by $\max NL(\theta) = \frac{1}{4}\sin\left(\frac{\theta\pi}{90}\right)$.

Appendix C Constructing Frontiers with Three Attributes

Here we describe the construction algorithm used to generate a concave efficient frontier with m = 5 alternatives, each described by k = 3 attributes. The algorithm used to generate a concave efficient frontier in R^3 space is as follows:

1. For each attribute $j = 1, 2, 3$, an attribute space was defined by:
    a. The distance between the minimum attribute value and the maximum attribute value, denoted $dist_j$, was randomly selected from a $Uniform[0,100]$ distribution.
    b. The attribute value for the alternative with the minimum attribute value, $\mu_{[1]j}$, was randomly selected from a $Uniform[100, 200 - dist_i]$ distribution.
    c. The attribute value for the alternative with the maximum attribute value, $x_{[5]j} = x_{[1]j} + dist_j$.
2. A normalized space was defined such that the domain of each random variable, Z_{ij}, $j = 1, 2, 3$ is $[0,1]$.
3. A random concave surface in normalized space was defined by the curve $z_{i1}^s + z_{i2}^s + z_{i3}^s = 1$, 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_{i1}, z_{i2}, z_{i3}) for each of five alternatives were randomly selected from the concave surface. For each alternative $a_i, i = 1,\ldots,5$, the following steps were performed:

    a. A value of z_{i1} was randomly drawn from a $Uniform[0,1]$ distribution.

    b. A value of z_{i2} was randomly drawn from a $Uniform\left[0, \sqrt[s]{1-z_{i1}^s}\right]$ distribution.

    c. $z_{i3} = \sqrt[s]{1-z_{i1}^s+z_{i2}^s}$.

5. The normalized attribute values for each alternative $a_i, i = 1,\ldots,5$, were translated to the attribute space that was defined in step 1 by:

    a. Assigning $\mu_{i1} = z_{ia}, \mu_{i2} = z_{ib}, \mu_{i3} = z_{ic}$ where (a, b, c) is a random permutation of (1, 2, 3), with each permutation having equal probability.

    b. Scaling (by $dist_j$) and shifting (by $\mu_{[1]j}$) each μ_{ij}, $j = 1,2,3$.

Appendix D Rejection Algorithm for Frontier Construction

Here we describe the rejection algorithm used to generate a random concave efficient frontier with m alternatives, each described by $k = 2$ attributes.

1. Generate a set of $m = 5$ points from a bivariate uniform probability distribution over the two-dimensional region [100, 200] × [100, 200].
2. Order the points over the first dimension, we denote these m ordered pair as $\left(x_{[1]}, y_1\right), \ldots, \left(x_{[m]}, y_m\right)$.
3. Calculate the $m - 1$ gradients $g_i = \dfrac{y_{i+1} - y_i}{x_{[i+1]} - x_{[i]}}, i = 1, \ldots, m-1$.
4. Accept the set of points as a concave efficient frontier if $g_i \geq g_{i+1}\ \forall i = 1, \ldots, m-1$.

This rejection algorithm assures that each frontier in the population of all concave efficient frontiers with m alternatives described by $k = 2$ attributes has an equal probability of being constructed.

Bibliography

Albert, J. (2014). LearnBayes: Functions for Learning Bayesian Inference. (R package version 2.15). Retrieved from http://CRAN.R-project.org/package=LearnBayes

Anderson, M., Daponte, B. O., Fienberg, S. E., Kadane, J. B., Spencer, B. D., & Steffey, D. L. (2000). Sampling-Based Adjustment of the 2000 Census A Balanced Perspective. *Jurimetrics, 40*, 341-356.

Arrow, K. J. (1951). *Social Choice and Individual Values* (First ed.). New York: Wiley.

Balinski, M. L., & Laraki, R. (2007, May). A theory of measuring, electing, and ranking. *Proceedings of the National Academy of Sciences of the United States of America, 104*(21), 8720-8725.

Balinski, M. L., & Laraki, R. (2010). *Majority Judgment: Measuring, Ranking, and Electing.* Cambridge, MA: The MIT Press.

Balinski, M. L., & Young, H. P. (1982). *Fair Representation.* New Haven: Yale University Press.

Bechhofer, R. E. (1954). A Single-sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances. *The Annals of Mathematical Statistics, 25*, 16-39.

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

Bevington, P. R., & Robinson, D. K. (2002). *Data Reduction and Error Analysis for the Physical Sciences* (Third ed.). Boston: McGraw-Hill.

Box, G. E., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for Experimenters* (Second ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Branke, J., Chick, S. E., & Schmidt, C. (2007). Selecting a Selection Procedure. *Management Science*, 1916-1932.

Burnett, K. D. (2011). *Congressional Apportionment.* Washington, DC: U.S. Census Bureau.

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

Casella, G., & Berger, R. L. (2002). *Statistical Inference* (Second ed.). Pacific Grove, CA: Duxbury.

Center for the Study of Services. (2011, Spring/Summer). Who's on Top? Ratings of Roofers. *Washington Consumers' Checkbook, 15*(4), pp. 50-65.

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

Chambal, S. P., Weir, J. D., Kahraman, Y. R., & Gutman, A. J. (2011). A Practical Procedure for Customizable One-Way Sensitivity Analysis in Additive Value Models. *Decision Analysis, 8*(4), 303-321.

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

Chen, C.-H., & Lee, L. H. (2011). *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation.* Hackensack, NJ: World Scientific Publishing Co. Pte. Ltd.

Clemen, R. T., & Reilly, T. (2001). *Making Hard Decisions* (Second ed.). Pacific Grove, CA: Duxbury.

Committee on National Statistics. (2009). *Coverage Measurement in the 2010 Census.* (R. M. Bell, & M. L. Cohen, Eds.) Washington, DC: The National Academies Press.

Davis, P. P., & Mulligan, J. (2012). *Net Coverage for the Household Population in the United States.* Washington, DC: U.S. Census Bureau.

DeGroot, M. H. (1989). *Probability and Statistics* (Second ed.). Reading, MA: Addison-Wesley.

Dyer, J. S. (2005). MAUT - Multiattribute Utility Theory. In J. Figueira, S. Greco, & M. Ehrgott (Eds.), *Multiple Criteria Decision Analysis - State of the Art Surveys* (pp. 265-295). New York: Springer.

Dyer, J. S., & Sarin, R. K. (1979). Measurable Multiattribute Value Functions. *Operations Research, 27*(4), 810-822.

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* New York: Chapman & Hall.

Einhorn, H. J., & Hogarth, R. M. (1986). Decision Making Under Ambiguity. *The Journal of Business, 59*(4), S225-S250.

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics, 75*(4), 643-669.

Farquhar, P. H., & Keller, R. L. (1989). Preference Intensity Measurement. *Annals of Operations Research*, 205-217.

Feller, W. (1950). *An Introduction to Probability Theory and Its Applications* (Vol. One). New York: John Wiley & Sons, Inc.

Fishman, G. S. (1996). *Monte Carlo Concepts, Algorithms, and Applications.* New York: Springer.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions* (Third ed.). New York: John Wiley & Sons, Inc.

French, S. (1995). Uncertainty and Imprecision: Modelling and Analysis. *The Journal of the Operational Research Society, 46*, 70-79.

Frisch, D., & Baron, J. (1988). Ambiguity and Rationality. *Journal of Behavioral Decision Making, 1*, 149-157.

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

Goldsman, D. M. (1984a). A Multinomial Ranking and Selection Procedure: Simulation and Applications. In S. Shepard, U. W. Pooch, & C. D. Pegden (Eds.), *Proceedings of the 1984 Winter Simulation Conference* (pp. 259-264). Piscataway, NJJ: IEEE Press.

Goldsman, D. M. (1984b). On Selecting the Best of K Systems: An Expository Survey of Indifference-zone Multinomial Procedures. In S. Shepard, U. W. Pooch, & C. D. Pegden (Eds.), *Proceedings of the 1984 Winter Simulation Conference* (pp. 107-112). Piscataway, NJ: IEEE Press.

Gupta, S. S., & Panchapakesan, S. (1979). *Multiple Decision Procedures.* New York: John Wiley & Sons.

Hadar, J., & Russell, W. R. (1969). Rules for Ordering Uncertain Prospects. *The American Economic Review, 59*, 25-34.

Herrmann, J. W. (2015). *Engineering Decision Making and Risk Managment.* Hoboken, NJ: John Wiley and Sons, Inc.

Howard, R. A. (1966). Decision Analysis: Applied decision theory. *Proceedings of the Forth International Conference on Operational Research* (pp. 55-71). New York: John Wiley.

Hyde, K., Maier, H. R., & Colby, C. (2003). Incorporating Uncertainty in the PROMETHEE MCDA Method. *Journal of Multi-Criteria Decision Analysis, 12*, 245-259.

Joint Committee for Guides in Metrology. (2008). *Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement.* Sevres Cedex, France: Joint Committee for Guides in Metrology.

Kahn, B. E., & Meyer, R. J. (1991, March). Consumer Multiattribute Judgments under Attribute-Weight Uncertainty. *Journal of Consumer Research Inc.*, *17*(4), 508-522.

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

Kim, S.-H., & Nelson, B. L. (2006). Selecting the Best System. In S. G. Henderson, & B. L. Nelson (Eds.), *Handbook in Operations Research and Management Science* (Vol. 13, pp. 501-534). Oxford: Elsevier.

Kirkwood, C. W. (1997). *Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets.* Belmont, CA: Duxbury Press.

Lahdelma, R., & Salminen, P. (2001). SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. *Operations Research, 49*, 454-469.

Lahdelma, R., Hokkanen, J., & Salminen, P. (1998). SMMA - Stochastic multiobjective acceptability analysis. *European Journal of Operations Research, 106*, 137-143.

LaPorte, G. J., Branke, J., & Chen, C.-H. (2012). Optimal Computing Budget Allocation for Small Computing Budgets. *Proceedings of the 2012 Winter Simulation Conference* (pp. 3324-3336). Berlin, Germany: IEEE.

Law, A. M. (2007). *Simulation Modeling and Analysis* (Fourth ed.). New York: McGraw-Hill.

Leber, D. D., & Herrmann, J. W. (2012). Incorporating Attribute Value Uncertainty into Decision Analysis. In G. Lim, & J. W. Herrmann (Eds.), *Proceedings of the 2012 Industrial and Systems Engineering Research Conference.* Orlando, FL.

Leber, D. D., & Herrmann, J. W. (2013a). Allocating Attribute-Specific Information-Gathering Resources to Improve Selection Decisions. In R. Pasupathy, S. -H. Kim, A. Tolk, R. Hill, & M. E. Kuhl (Eds.), *Proceedings of the 2013 Winter Simulation Conference* (pp. 101-112). Washington DC: IEEE.

Leber, D. D., & Herrmann, J. W. (2013b). Decision Analysis Methods for Selecting Consumer Services with Attribute Value Uncertainty. In M.-L. T. Lee, M. Gail, R. Pfeiffer, G. Satten, T. Cai, & A. Gandy (Eds.), *Risk Assessment and Evaluation of Predictions* (pp. 405-424). New York: Springer.

Leber, D. D., & Herrmann, J. W. (2014a). Resource Allocation for Selection Decisions with Measurement Uncertainty. In Y. Guan, & H. Liao (Eds.), *Proceedings of the 2014 Industrial and Systems Engineering Research Conference.* Montreal.

Leber, D. D., & Herrmann, J. W. (2014b). Sample Allocation for Multiple Attribute Selection Problems. In A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, & J. A. Miller (Eds.), *Proceedings of the 2014 Winter Simulation Conference* (pp. 722-732). Savanah, GA: IEEE.

Leber, D. D., & Herrmann, J. W. (2015). Impact of Decision and Allocation Rules on Selection Decisions. In S. Cetinkaya, & J. K. Ryan (Eds.), *Proceedings of the 2015 Industrial and Systems Engineering Research Conference*, (pp. 1787-1796). Nashville, TN.

Lee, L. H., Chew, E. P., Teng, S., & Goldsman, D. (2004). Optimal Computing Budget Allocation for Multi-Objective Simulation Models . *Proceedings of the 2004 Winter Simulation Conference*, (pp. 586-594).

Lee, L. H., Chew, E. P., Teng, S., & Goldsman, D. (2010). Finding the non-dominated Pareto set for multi-objective simulation models. *IIE Transactions, 42*, 656-674.

Lee, P. M. (1997). *Bayesian Statistics - An Introduction* (Second ed.). London: Arnold.

Lindley, D. V. (2006). *Understanding Uncertainty*. Hoboken, NJ: John Wiley & Sons, Inc.

Luce, R. D., & Raiffa, H. (1957). *Games and Decision.* New York: John Wiley & Sons, Inc.

Miller, J. O., Nelson, B. L., & Reilly, C. H. (1998). Efficient Multinomial Selection in Simulation. *Naval Research Logistics, 45*, 459-482.

Montgomery, D. C. (2013). *Design and Analysis of Experiments* (Eighth ed.). New York: John Wiley & Sons.

Mule, T. (2012). *Summary of Estimates of Coverage for Persons in the United States.* Washington, DC: U.S. Census Bureau.

Mustajoki, J., Hämäläinen, R. P., & Salo, A. (2005). Decision Support by Interval SMART/SWING-Incorpotating Imprecision in the SMART and SWING Methods. *Decision Sciences, 36*(2), 317-339.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Pandey, V., Nikolaidis, E., & Mourelatos, Z. (2011, June). Multi-Objective Decision Making under Uncertainty and Incomplete Knowledge of Designer Preferences. *SAE International Journal of Materials and Manufacturing, 4*(1), 1155-1168.

Press, S. J. (1988). *Bayesian Statistics: Principles, Models, and Applications.* New York: John Wiley & Sons, Inc.

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Raiffa, H. (1961). Risk, Ambiguity, and the Savage Axioms: Comment. *The Quarterly Journal of Economics, 75*(4), 690-694.

Savage, L. J. (1954). *The Foundations of Statistics* (First ed.). New York: John Wiley & Sons, Inc.

Searle, S. R., Casella, G., & McCulloch, C. E. (2006). *Variance Components.* Hoboken, NJ: John Wiley & Sons.

Srivastava, R. P. (1997). Decision Making Under Ambiguity: A Belief-Function Perspective. *Archives of Control Sciences, 6*(1), 5-27.

Stewart, T. J. (2005). Dealing with Uncertainties in MCDA. In J. Figueira, S. Greco, & M. Ehrgott (Eds.), *Multiple Criteria Decision Analysis* (pp. 445-469). New York: Springer.

Tervonen, T., & Figueira, J. R. (2008). A Survey on Stochastic Multicriteria Acceptability Analysis Methods. *Journal of Multi-Criteria Decision Analysis, 15*, 1-15.

Tollefson, E., Goldsman, D. M., Kleywegt, A., & Tovey, C. (2014). Optimal Selection of the Most Probable Multinomial Alternative. *Sequential Analysis, 33*, 491-508.

U.S. Census Bureau . (2010, December 6). *U.S. Census Bureau Releases 2010 Demographic Analysis Population Estimates* . Retrieved November 12, 2012, from United States Census Bureau: http://www.census.gov/newsroom/releases/archives/2010_census/cb10-cn87.html

von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press.

von Winterfeldt, D., & Edwards, W. (1986). *Decision Analysis and Behavior Research.* Cambridge: Cambridge University Press.

Winston, W. L. (2004). *Operations Research* (Fourth ed.). Belmont, CA: Thompson.

Zhang, Y., Fan, Z.-P., & Liu, Y. (2010). A method based on stochastic dominance degrees for stochastic multiple criteria decision making. *Computers and Industrial Engineering, 28*, 544-552.