OPTIMAL RANDOMIZED AND NON-RANDOMIZED PROCEDURES FOR MULTINOMIAL SELECTION PROBLEMS

A Thesis
Presented to
The Academic Faculty

by

Eric Sander Tollefson

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
May 2012

OPTIMAL RANDOMIZED AND NON-RANDOMIZED PROCEDURES FOR MULTINOMIAL SELECTION PROBLEMS

Approved by:

David M. Goldsman, Co-Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Craig A. Tovey, Co-Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Christos Alexopoulos
School of Industrial and Systems Engineering
Georgia Institute of Technology

Anton J. Kleywegt
School of Industrial and Systems Engineering
Georgia Institute of Technology

Susan M. Sanchez
Operations Research Department
Naval Postgraduate School

Date Approved: March 16, 2012

ACKNOWLEDGEMENTS

First, I must acknowledge the invaluable contributions of my principal advisor and mentor, Dr. Dave Goldsman. His guidance, advice, and friendship have made this educational experience most rewarding and enjoyable. I hope to emulate his passion for learning, selfless dedication, and tireless efforts in my own professional career. I also owe a great debt of gratitude to my co-advisor, Dr. Craig Tovey. His initial contribution to the research effort was the primary catalyst for this body of research, and his insightful comments and recommendations along the way have been greatly appreciated and essential to my success. I am grateful to my other Committee members as well — Dr. Anton Kleywegt, for challenging my understanding of the problem and providing significant input into the written documents; Dr. Christos Alexopoulos, for his keen insights into the problem; and Dr. Susan Sanchez, for her welcomed input and for agreeing to fly across the country to endure my defense.

To the faculty of the School of Industrial and Systems Engineering (ISyE), I owe my sincere gratitude for their assistance, example, and contagious love for knowledge. And to the most helpful and always friendly ISyE staff: thanks for all that you do behind the scenes. I also thank my fellow students for their companionship and their assistance. In particular, I thank Soheil Shayegh, my officemate, for his great conversations and friendship, and Dimitri Papageorgiou, for our valuable discussions of optimization.

Almost last, but certainly not least, I thank my beautiful wife, Kara, and my three wonderful children, Nathan, Emma, and Claire. Without their unconditional love, support, and understanding, none of this would have been possible.

Finally, I thank God for His innumerable blessings and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS OR ABBREVIATIONS

AS	acceptable selection
AVC	all vectors comparisons
b	budget, or maximum number of observations to be taken
$\mathcal{B}$	the set $\{\boldsymbol{\eta} : \sum_{i=1}^{k} \eta_i = b\}$
$\mathcal{B}'$	the set of $\boldsymbol{\eta}'$ at the end of the directed network
B	batch size
BEM	Bechhofer, Elmaghraby, and Morse
BG	Bechhofer and Goldsman
BK	Bechhofer and Kulkarni
BKS	Bechhofer, Kiefer, and Sobel
BSG	Bechhofer, Santner, and Goldsman
BSP	Bernoulli selection problem
C	random variable representing total observation cost taken through experiment termination
CS	correct selection
CV	coefficient of variation
$c_{\boldsymbol{\eta}}$	marginal observation cost at $\boldsymbol{\eta}$
D_B	break-even batch discount rate
EPC	equal probability configuration
$\boldsymbol{\eta}$	cumulative success vector $(\eta_1, \eta_2, \ldots, \eta_k)$
$\boldsymbol{\eta}'$	left-lexicographic cumulative success vector $(\eta_{[k]}, \eta_{[k-1]}, \ldots, \eta_{[1]})$
$\boldsymbol{\eta}''$	$\boldsymbol{\eta}'$ uncurtailed with respect to budget b
$\boldsymbol{\eta}^{\nu}$	origination node for arc ν ending at component ν of vector $\boldsymbol{\eta}$
$f_{\boldsymbol{\eta}}$	continuous variable representing flow from node $\boldsymbol{\eta}$ to the sink
$f_{\boldsymbol{\eta}}^{i}$	continuous variable representing flow from node $\boldsymbol{\eta}$ due to success by alternative i
$\mathcal{G}$	set of groups $\{\mathbf{g^1}, \mathbf{g^2}, \ldots, \mathbf{g^G}\}$ in a node
G	number of groups in a node, or $\lvert\mathcal{G}\rvert$

g_j^i	jth element of group i
γ	stopping policy
Γ	set of all stopping policies γ
IZ	indifference zone
k	number of alternatives
LFC	least favorable configuration
LP	linear program or programming
MC	marginal cost
MC	Monte Carlo
MCP	multiple comparison procedure
MIP	mixed integer program or programming
MSP	multinomial selection problem
N	random variable representing total number of observations taken through experiment termination
n	truncation stopping parameter
$\mathcal{N}$	the set $\{\boldsymbol{\eta} : \sum_{i=1}^{k} \eta_i \leq b\}$
$\mathcal{N}'$	the set $\{\boldsymbol{\eta}' : \sum_{i=1}^{k} \eta_{[i]} \leq b\}$
$\mathcal{N}''$	the set $\{\boldsymbol{\eta}'' : \sum_{i=1}^{k} \eta_{[i]} \leq b\}$
OCBA	optimal computing budget allocation
Ω	probability configuration sample space
Ω_{IZ}	the subspace $\{\boldsymbol{p} \in \Omega : p_{[k]}/p_{[k-1]} < \theta^\star\}$, also called the IZ
$\Omega_{\text{IZ}m}$	subspace of Ω_{IZ} in which m alternatives are in the IZ
Ω_{PZ}	the subspace $\{\boldsymbol{p} \in \Omega : p_{[k]}/p_{[k-1]} \geq \theta^\star\}$, also called the PZ
$\boldsymbol{p}$	general probability configuration $(p_1, p_2, \ldots, p_k)$
P(AS)	probability of acceptable selection
$\mathsf{P}_{\boldsymbol{p}}$(AS)	probability of acceptable selection for general probability configuration $\boldsymbol{p}$
P(CS)	probability of correct selection
$\mathsf{P}_{\boldsymbol{p}}$(CS)	probability of correct selection for general probability configuration $\boldsymbol{p}$

$\mathsf{P}(\mathrm{CS}\|\boldsymbol{\eta})$	conditional (or posterior) $\mathsf{P}(\mathrm{CS})$ given arrival at $\boldsymbol{\eta}$
$\mathsf{P}_{\mathrm{u}}(A)$	probability of event A when $\boldsymbol{p}$ is distributed uniformly in Ω
$p_{\boldsymbol{\eta}}$	conditional probability of stopping at $\boldsymbol{\eta}$ given arrival at $\boldsymbol{\eta}$
$\pi(\boldsymbol{\eta})$	set of permutations of $\boldsymbol{\eta}$
$\pi^{\star}(\boldsymbol{\eta})$	set of permutations of $\boldsymbol{\eta}$ that result in a correct selection
Procedure $\mathcal{M}_{\mathrm{BEM}}$	MSP procedure due to BEM (1959)
Procedure $\mathcal{M}_{\mathrm{BK}}$	MSP procedure due to BK (1984)
Procedure $\mathcal{M}_{\mathrm{C}}$	MSP procedure due to Chen (1988a)
Procedure $\mathcal{M}_{\mathrm{C}'}$	MSP procedure due to Chen (1988a), but modified to include the strong curtailment stopping rule
Procedure $\mathcal{M}_{\mathrm{NR}}$	optimal non-randomized MSP procedure
Procedure $\mathcal{M}_{\mathrm{R}}$	optimal randomized MSP procedure
Procedure $\mathcal{M}_{\mathrm{RA}}$	MSP procedure due to RA (1979)
Procedure $\mathcal{M}_{\mathrm{RA}'}$	MSP procedure due to Chen (1992)
$P^{\star}$	required probability of correct selection
PZ	preference zone
r	difference stopping parameter
RA	Ramey and Alam
RS	ranking and selection
ρ	fraction of total observation costs due to batching
SC	slippage configuration
$\mathsf{SD}(N)$	standard deviation of random variable N
$\mathcal{S}_{\mathrm{NR}}$	set of $\boldsymbol{\eta}$ that are stopping points for Procedure $\mathcal{M}_{\mathrm{NR}}$
$\mathcal{S}_{\mathrm{R}}$	set of ordered pairs $(\boldsymbol{\eta}, p_{\boldsymbol{\eta}})$ that are stopping points for Procedure $\mathcal{M}_{\mathrm{R}}$
$S_{\boldsymbol{\eta}}$	number of sample paths to node $\boldsymbol{\eta}$
S_{γ}	set of all nodes $\boldsymbol{\eta}$ such that $\boldsymbol{\eta}$ is a stopping node under procedure policy γ
t	inverse sampling stopping parameter
TC	total cost

$t(\boldsymbol{\eta})$	number of elements of $\boldsymbol{\eta}$ tied with $\eta_{[k]}$, including $\eta_{[k]}$
Θ	quantity $\sum_{i=1}^{k} 1/\theta_i$
τ	total observation cost through observation b
θ_i	ratio p_1/p_i
$\theta^\star$	relative ratio indifference zone parameter
V	volume or content of a polytope
$\boldsymbol{v}$	a vector or a vertex
$\mathsf{Var}(N)$	variance of random variable N
$\boldsymbol{v}_i$	vertex i
W_{J}	procedure $\mathcal{M}_{\mathrm{J}}$ inefficiency
$\overline{W}_{\mathrm{J}}$	mean procedure $\mathcal{M}_{\mathrm{J}}$ inefficiency
$\overline{V}^{I}_{\mathrm{J,L}}$	mean relative procedure performance of procedure $\mathcal{M}_{\mathrm{J}}$ versus procedure $\mathcal{M}_{\mathrm{L}}$ over $P^\star$-interval I
$Y_{\boldsymbol{\eta}}$	binary variable representing whether or not $\boldsymbol{\eta}$ is a stopping node
Z_γ	binary variable representing whether or not stopping policy γ is active

SUMMARY

Multinomial selection problem (MSP) procedures aim to select the best (most probable) alternative based upon a sequence of multinomial observations. The classical formulation of the procedure design problem is to find a decision rule for terminating sampling. The decision rule should minimize the expected number of observations taken while achieving a specified indifference zone requirement on the prior probability of making a correct selection when the alternative configurations are in a particular subset of the possible probability space called the preference zone. We study the constrained version of the design problem in which there is a given maximum number of allowed observations.

Numerous procedures have been proposed over the past 50 years, all of them suboptimal. In this thesis, we find the optimal selection procedure for any given probability configuration via linear programming. The optimal procedure turns out to be necessarily randomized in many cases. We also find the optimal non-randomized procedure via mixed integer programming. We demonstrate the performance of the methodology on a number of examples.

We then reformulate the mathematical programs to make them more efficient to implement, thereby significantly expanding the set of real world problems that can be modeled. We prove that there exists an optimal policy which has at most one randomized decision point and we develop a procedure for finding such a policy. Additionally, we show that the formulations can be extended to replicate existing bounded procedures from the literature, simply by altering the feasible region of the mathematical programs.

Our formulations also allow us to examine situations in which marginal observation costs are not constant. While variable marginal observation costs are realistic, they have not been considered in the literature, largely because the tools required did not exist. We leverage our formulations to develop a new methodology that guarantees the optimal randomized

and non-randomized procedures for a broad class of variable observation cost functions. We then analyze procedure performance under a representative set of observation cost functions. We also develop two new tools for examining specific cost-related issues — one for deciding whether or not to purchase experimental supplies in batches and one for determining the effect on total observation cost due to changing probability requirements.

Next, we show that there is very little difference between the relative performances of the optimal randomized and non-randomized procedures, particularly for large budgets. Additionally, we compare existing procedures using the optimal procedure as a benchmark, and produce updated tables for a number of those procedures. For our comparisons, we develop a new set of metrics and employ some traditional statistical measures, such as variance, not typically considered in the literature.

Finally, we examine some fundamental assumptions — normally taken for granted in the literature — regarding the application of MSP procedures. In particular, we show how the choice of the indifference zone parameter affects the size of the preference zone when any alternative configuration is equally possible. We then define the concept of an "acceptable selection" for alternatives in the indifference zone and discuss some Monte Carlo sampling results. Finally, we look at issues regarding the conditional (posterior) probability of correct selection upon procedure termination and its implications on the applicability of MSP procedures in general.

CHAPTER I

INTRODUCTION

We consider the problem of selecting the best alternative out of $k \geq 2$ competing alternatives. Each time the alternatives compete, alternative i has probability $p_i > 0$ of winning, where $\sum_{i=1}^{k} p_i = 1$. Let $p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[k]}$ denote the ordered p_i's. The alternative associated with $p_{[k]}$ is denoted with i^* and is called the *most probable* or *best*. The only information that is known in advance is the number k of alternatives, how to conduct independent random observations in which the k alternatives compete, and how to identify the winning alternative in each observation. That is, the probabilities $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$ are not known, and it is not known which alternative is more or less likely to win than another. We want to conduct observations to identify the best alternative i^* with high probability. These types of problems are called multinomial selection problems (MSPs).

Given a method to choose an alternative, the probability that alternative i^* is chosen is called the probability of correct selection and is denoted with $\mathsf{P}_{\boldsymbol{p}}(\mathrm{CS})$, or just $\mathsf{P}(\mathrm{CS})$. Clearly the probability of correct selection depends on $\boldsymbol{p}$, for example if $p_{[k]}$ is small, and $p_{[k]}$ and $p_{[k-1]}$ are close to each other, then $\mathsf{P}_{\boldsymbol{p}}(\mathrm{CS})$ can be small. Relative ratio indifference zone MSPs specify constants $(\theta^\star, P^\star)$ with $\theta^\star > 1$ and $1/k < P^\star < 1$. Ideally, the objective is to minimize the expected number of observations while requiring that

$$\mathsf{P}_{\boldsymbol{p}}(\mathrm{CS}) \;\geq\; P^\star \text{ for all } \boldsymbol{p} \text{ such that } p_{[k]}/p_{[k-1]} \;\geq\; \theta^\star, \tag{1.1}$$

where the constant $\theta^\star$ can be regarded as the "smallest ratio $p_{[k]}/p_{[k-1]}$ worth detecting."

To guarantee (1.1), we require additional information. Let Ω be the set of all possible probability configurations $\boldsymbol{p}$. The *preference zone* (PZ) is denoted $\Omega_{\mathrm{PZ}} \equiv \{\boldsymbol{p} \in \Omega : p_{[k]}/p_{[k-1]} \geq \theta^\star\}$. Its complement, Ω_{IZ}, is the *indifference zone* (IZ).

Definition 1.1 Given a method to choose an alternative, the *least favorable configuration*

(LFC) is the probability configuration $\boldsymbol{p} \in \Omega_{\mathrm{PZ}}$ that minimizes $\mathsf{P}_{\boldsymbol{p}}(\mathrm{CS})$, i.e., the LFC is

$$\operatorname*{argmin}_{\boldsymbol{p}\in\Omega_{\mathrm{PZ}}} \mathsf{P}_{\boldsymbol{p}}(\mathrm{CS}).$$

While guaranteeing (1.1), we minimize the expected number of observations when $\boldsymbol{p}$ is in the LFC. For some methods, the LFC has not yet been identified, so we instead attempt to minimize the expected number of observations when $\boldsymbol{p}$ is in the *slippage configuration* (SC):

$$\boldsymbol{p} = \left(\frac{\theta^\star}{\theta^\star + k - 1}, \frac{1}{\theta^\star + k - 1}, \ldots, \frac{1}{\theta^\star + k - 1}\right),$$

which is assumed to be the LFC. In that case, our objective is to minimize the expected number of observations while requiring that

$$\mathsf{P}_{\boldsymbol{p}}(\mathrm{CS}) \;\geq\; P^\star \text{ when } \boldsymbol{p} \text{ is the SC}, \tag{1.2}$$

a weaker condition than condition (1.1).

When we attempt to minimize the expected number of observations, we are assuming that the marginal cost of each observation is constant. In Chapter 5, we will change our goal slightly to that of minimizing the expected total cost of the observations when marginal observation costs are not constant.

Let $x_{ij} = 1$ if alternative i is the winner on observation j, and $x_{ij} = 0$ otherwise. Let $\eta_{im} \equiv \sum_{j=1}^{m} x_{ij}$ denote the number of observations won by alternative i among the first m observations. Let $\boldsymbol{\eta}_m \equiv (\eta_{1m}, \eta_{2m}, \ldots, \eta_{km})$, and let $\eta_{[1]m} \leq \cdots \leq \eta_{[k]m}$ denote the ordered η_{im}'s. We restrict attention to methods that, after n observations, choose an alternative i with $\eta_{in} = \eta_{[k]n}$. If there are multiple alternatives with $\eta_{in} = \eta_{[k]n}$, we assume that each of those alternatives is chosen with equal probability. Let $\hat{i}_n$ denote such a chosen alternative i with $\eta_{in} = \eta_{[k]n}$.

Note that the above MSP model can be used as non-parametric model for other types of problems. For example, the experimenter might decide to take an observation, say, cycle time, from each alternative under consideration and use those observations to determine which alternative is 'best' for that observation. The best alternative now is "the alternative with the highest probability of achieving the shortest cycle time." That statement is not the same as that of a parametric problem which seeks to select "the alternative with the

shortest mean cycle time." However, if the rephrased goal is appropriate, an MSP model is a viable, but usually less efficient (in terms of the procedures), alternative to parametric problems and may be useful when normality assumptions about the performance measure of interest are suspect.

Problem Applicability: Before we discuss different versions of the MSP described above, it is important to point out a number of common properties of these problems that determine the types of applications for which these problems are appropriate.

1. The probability of correct selection $\mathsf{P}_{\boldsymbol{p}}(\mathrm{CS})$ is the *prior* probability of correct selection, before any data have been observed. In fact, it is possible that a method that satisfies condition (1.1) or condition (1.2) stops after n observations and chooses an alternative $\hat{i}_n$, but that the *posterior* probability of correct selection, given the observed observation data, is much less than $P^\star$. It is also possible that a method continues with more observations, even if after m observations an alternative $\hat{i}_m$ can be chosen with posterior probability of correct selection, given the observed observation data, significantly exceeding $P^\star$. Since the prior probability of correct selection is the expected value of the posterior probability of correct selection, conditions (1.1) and (1.2) just require that in expectation the sample paths with too many observations compensate for the sample paths with too few observations. This may be a reasonable problem formulation for an application in which many replications of the experiment (each consisting of multiple observations and selection of a winner) will be done, and the major concern is that the correct winner should be chosen in at least fraction $P^\star$ of these experiments. This may not be a good problem formulation for an application in which the experiment will be conducted only once or a small number of times, or in which it is important to ensure that the posterior probability of correct selection is sufficiently high.

2. The probability of correct selection $\mathsf{P}_{\boldsymbol{p}}(\mathrm{CS})$ only gives credit for the event in which the alternative i with the largest value of p_i is chosen. For example, there is no partial credit for choosing the alternative i with the second largest value of p_i, and there is

no difference in credit between the event in which the alternative i with the second largest value of p_i is chosen and the event in which the alternative i with the smallest value of p_i is chosen.

3. Condition (1.1) or condition (1.2) is only required to apply in the PZ, that is, for all $\boldsymbol{p} \in \Omega_{\mathrm{PZ}}$. No distinction is made between a procedure with higher probability of correct selection and a procedure with lower probability of correct selection if $\boldsymbol{p} \in \Omega_{\mathrm{IZ}}$.

4. Each of the k alternatives has to compete in each of the observations, that is, cost cannot be reduced by letting only a subset of the alternatives compete in a observation.

5. In each observation we are only concerned with the winning alternative and not other data such as a score earned by each alternative.

While many intuitively appealing MSP procedures have been developed, none can claim optimal performance for a wide range of MSPs. However, if we view the MSP as a network of probability flows, and assume a constraint on the maximum number of observations we are willing to take, we can use mathematical programming techniques, such as linear programming (LP) and mixed integer programming (MIP), to determine the optimal procedure for any given MSP and probability configuration $\boldsymbol{p}$. Furthermore, we can leverage properties of the MSP, such as symmetry, in order to reduce the size of the network, thereby increasing the scope of the problems that we can solve.

Formulating the MSP as a mathematical program has other advantages as well. We can add, modify, and remove constraints to model unique variations of the problem. For example, we can create mathematical programs that replicate existing procedures or require a minimum posterior conditional probability of correct selection. We can also alter our objective function in order to minimize the expected total cost of the observations instead of the expected number of observations. Such a capability is essential for realistic scenarios in which marginal observation costs vary over time.

Knowledge of the optimal procedure for any given MSP also provides us a benchmark against which we can compare existing procedures. Whereas previous comparisons and evaluations of procedure performance depended upon the relative performance between

procedures, we can now evaluate a procedure's performance with respect to the optimal procedure. Such a benchmark allows us to quantify a procedure's efficiency, which is more desirable than quantifying its relative performance versus another arbitrary procedure.

1.1 Research Objectives

The objectives of this research are to:

- Develop efficient LP and MIP formulations of the MSP in order to identify the optimal stopping policy for any given MSP.
- Reformulate the mathematical programs to identify the optimal stopping policy under variable observation cost functions, and develop insights into the effects of variable observation cost functions on procedure performance.
- Examine the performance of existing procedures with respect to the optimal procedures for a representative set of MSPs.
- Examine key assumptions concerning MSPs in order to develop initial insights and to set the stage for future research.

1.2 Research Outline

This thesis is arranged as follows. Chapter 2 reviews the literature related to ranking and selection (RS), in particular MSPs, and the integration of observation costs into RS procedures. Chapter 3 introduces and describes our mathematical formulations of the MSP to include some preliminary results. In Chapter 4, we improve the algorithmic efficiency of the original formulations, prove some general results about their solutions, and formulate existing procedures as mathematical programs. In Chapter 5, we reformulate the MSP to integrate variable observation costs and we demonstrate results for a representative set of marginal observation cost functions. Chapter 6 includes an analysis of existing MSP procedures, particularly in relation to the optimal procedures. In Chapter 7, we examine some common MSP assumptions concerning the PZ and conditional $\mathsf{P}(\mathrm{CS})$ in order to gain

initial insights and to provide a starting point for future research. Chapter 8 is a summary of our work.

CHAPTER II

LITERATURE REVIEW

In general, our goal is to compare competing alternatives. An alternative can be an individual product or a process, or it can be a system — a set of interrelated products or processes working together to achieve a common purpose. Comparisons between alternatives are made based upon a single measure of performance, e.g., mean queue length, mean cycle time, probability of success, etc. We may be interested in selecting the single best alternative or a subset that includes the best alternative.

Classical hypothesis testing and confidence interval techniques provide ways to determine whether one alternative is better than the other under relevant assumptions about the distribution of the performance measure. Error can be controlled in order to limit the probability of rejecting the null hypothesis when it is true (i.e., choosing an alternative as best when both are the same) and the probability of accepting the null hypothesis when one alternative is better than the other by a specified amount (i.e., concluding that the alternatives are the same, when they are not). Common tests include the Z-test for normal data with known variances, the t-test and paired t-test for normal data with unknown variances, and F-tests for comparing variances. A primary disadvantage of these techniques is that the null hypothesis (equal means) will eventually be rejected if the sample size is large enough.

If there are more than two alternatives under comparison, simultaneous confidence intervals can be used. There exist numerous methods to conduct simultaneous consideration of multiple confidence intervals as part of a hypothesis test, such as the Bonferroni and Tukey multiple comparison statistics. These methods also share the same disadvantage of the single confidence interval approach — eventual rejection of the null hypothesis as the sample size increases. Additional information on these classical techniques can be found in several introductory statistics textbooks, such as Hines et al. (2003). For a detailed coverage

of simultaneous inference, see Miller (1981).

Multiple comparison procedures (MCPs) are part of another body of research that treats multiple alternative comparisons as an inference problem. These procedures are not a departure from the classical approach, but leverage many of the classical techniques to characterize the relationships between alternatives. The most basic of these MCPs uses simultaneous confidence intervals for all possible $k(k-1)/2$ pairwise comparisons between k competing alternatives. Since all pairwise confidence intervals have the same test statistic and pooled variance (under the assumption that all variances are the same), their half-widths are identical. If there is a clear 'winner' among the alternatives, the best can be identified. Unfortunately, there is no guarantee of such an outcome. For multiple comparisons with a control, only $k-1$ intervals are required (each alternative versus the control). These techniques, however, do not provide any additional advantage for choosing the best alternative or a subset that includes the best alternative. Finally, there are MCPs that compare each alternative with the best performing alternative during or after sampling. MCPs, particularly those being applied within a simulation environment, can be used in conjunction with variance reduction techniques, such as common random variates, in which positive correlations are induced between the alternatives. For an example of an MCP procedure for comparisons against the best that leverages common random numbers, see Kolonja et al. (1993). Hsu (1996) provides a good overview of MCPs in general.

Another main body of research for multiple comparisons is the field of RS, which is a departure from, but not unrelated to, classical statistical techniques. In fact, there is a great deal of research that seeks to unify the fields of MCP and RS. Swisher and Jacobson (1999) provide a good review of MCP–RS unifying research with applications to simulation. Matejcik and Nelson (1995) propose a two-stage multiple comparison (with the best alternative) procedure for the expected value of a performance measure, and show that particular indifference zone RS procedures are a special case.

In §2.1, we describe the initial development of RS and provide an overview of the types of procedures that have been developed. In §2.2, we review RS techniques and procedures developed specifically for MSPs. Lastly, in §2.3, we describe research that considers various

types of costs (e.g., sampling and opportunity costs) in RS.

2.1 Ranking and Selection

Ranking and selection has its roots in the late 1940s. Mosteller (1948) proposed a non-parametric test for "the problem of the greatest one" in which the null hypothesis is that of equal means and the alternative hypothesis is that one alternative has "slipped" farther to the right than any other (i.e., has a larger mean). Paulson (1952b) expanded upon Mosteller's work by developing a fixed sample size parametric test for normally distributed data with the goal of deciding whether or not the means are equal, and, if they are not, identifying which is best. Paulson (1949, 1952a) also developed methods to divide a group of alternatives into a "superior" group and an "inferior" group based upon the population means (assumed to be normally distributed), as well as to identify the best alternative when comparing $k-1$ alternatives with a control.

Bechhofer (1954) is the first paper that formalized what we now call RS. In it, he describes an experimental design to determine the ranking of k normal population means with known (but not necessarily equal) variances. The goal of the design is to find the best subset of the populations with or without regard to order (i.e., ranking within the subset). This differs from the goal of previous procedures in which ranking can only be indirectly inferred from the size of the differences between the means. He introduces a parameter, $\delta^{\star}_{\hat{k}_{i+1},\hat{k}_i}$, which represents the smallest difference between the ordered population means that is "worth detecting". There, $\hat{k}_i$ and $\hat{k}_{i+1}$ are indices for the kth and $(k+1)$th ordered population means (or ordered subsets of population means). That type of parameter is what we now call an indifference zone parameter. The purpose of the experimental design is to find the smallest number of samples across the k alternatives that will guarantee a given probability of correctly ranking the alternatives whenever their actual differences exceed the parameters $\delta^{\star}_{\hat{k}_{i+1},\hat{k}_i}$. As such, it is considered a single-stage procedure, since the total number of samples is determined *a priori*. Bechhofer also introduced the concept of the LFC — the configuration of the population means that achieves the lowest probability of a correct ranking under his problem assumptions. He tabulates the experimental design

parameters for varying numbers of alternatives, subsets to be ranked, and probabilities of correct ranking. Bechhofer and Sobel (1954) quickly followed up with an experimental design for ranking normal population *variances* with known means using the same formulation as the previous paper. Bechhofer, Dunnett, and Sobel (1954) proposed a two-stage procedure for ranking normal population means with common, unknown variance, in which the first stage is used to obtain a variance estimate that is subsequently used to determine the remaining number of samples to be taken in the second stage.

Bechhofer's pioneering papers in 1954 led to a large number of RS indifference zone procedures, many of those for ranking normal populations. Paulson (1964) describes a sequential, multi-stage procedure in which inferior populations may be eliminated as sampling continues. Each stage then takes a single observation from each alternative still in contention (i.e., not yet eliminated). Bechhofer, Kiefer, and Sobel (BKS) (1968) present a sequential procedure framework in general for Koopman–Darmois populations (e.g., normal, exponential, Bernoulli, Poisson, and multinomial distributions). Examples of other procedures for ranking normal populations include a two-stage procedure developed by Rinott (1978) and the fully sequential procedure developed by Kim and Nelson (2001).

As mentioned above, procedures have been developed for populations other than normal. For example, there are many procedures in the literature for Bernoulli selection problems (BSPs). Unlike normal population ranking problems that seek to rank population means, BSPs seek to identify the alternative with the highest probability of success (success parameter). Sobel and Huyett (1957) propose a single-stage procedure which uses an indifference zone parameter similar to those discussed already, namely, the smallest difference in probability of success worth detecting. On the other hand, the sequential procedure proposed in BKS (1968) uses an *odds ratio* indifference zone parameter. If we order the success probabilities $p_{[1]} \leq p_{[2]} \leq \ldots p_{[k-1]} \leq p_{[k]}$, then the odds ratio parameter, $\theta^\star$, between the best and second best alternatives is

$$\theta^\star = \frac{p_{[k]}(1 - p_{[k-1]})}{(1 - p_{[k]})p_{[k-1]}}.$$

Our discussion has focused on ranking procedures in which the goal is to choose the best

alternative or to rank a subset of the best alternatives. These fall into the class of RS called indifference zone procedures, since an indifference zone parameter must be specified by the experimenter. There is another class of RS procedures called subset selection procedures. For these, the goal is to select a subset (of a predetermined or random size) which contains the best alternative. Most subset selection procedures, such as those reviewed for MSPs in the next section, do not require the specification of an indifference zone parameter. However, there have been procedures developed, such as that of Sullivan and Wilson (1989), that do allow the experimenter to define an indifference zone.

2.2 Multinomial Selection Problems

A significant body of RS research exists for MSPs. In the introduction in Chapter 1, we described the general setup for the selection-of-the-best MSPs, which are our primary focus. We begin our discussion here in §2.2.1 with those types of MSPs. In §2.2.2, we will briefly review other types of MSPs.

2.2.1 Procedures for Selecting the Best

This section will describe the relevant MSP procedures found in the literature for selecting the best of multiple alternatives. We present, in some detail, those procedures that we will readdress in future chapters. The format of such procedure descriptions and their examples, as well as the notation, mirror that of Bechhofer, Santner, and Goldsman (BSG) (1995). Other MSP procedures will be briefly described.

There are different versions of the MSP described in the introduction (Chapter 1). In the single-stage MSP, the problem is to choose in advance the smallest number n of observations such that if $\hat{i}_n$ is chosen as described above, then condition (1.1) or condition (1.2) holds. For example, the classic single-stage procedure due to Bechhofer, Elmaghraby, and Morse (BEM) (1959) proceeds as follows.

Procedure $\mathcal{M}_{\text{BEM}}$

- For the given k, $\theta^\star$, and $P^\star$, choose the fixed sample size n from an appropriate table, e.g., BEM (1959) or BSG (1995).

- Take n multinomial observations in a single stage.
- Select $\hat{i}_n$ as the best alternative, randomizing to break ties.

Remark 2.1 Kesten and Morse (1959) prove that the SC is the LFC for Procedure $\mathcal{M}_{\text{BEM}}$. Thus, the parameter n in BEM (1959) and BSG (1995), chosen based upon the procedure performance in the SC, satisfies condition (1.1) for the given k, $\theta^\star$, and $P^\star$.

Note that the version of the problem above does not consider or allow the possibility of stopping before n observations have been completed, even if the observations up to observation $m < n$ seem to clearly indicate the winner. Another version of the problem considers *sequential* procedures which dynamically choose the number n of observations such that if $\hat{i}_n$ is chosen as described above, then condition (1.1) or condition (1.2) holds, and the (prior) expected number of observations is minimized in the LFC or SC. Sequential procedures can be one of three types:

- *Unbounded* sequential procedures for which there is no bound on the number of observations taken during an experiment.
- *Bounded* sequential procedures for which the chosen procedure parameters provide an upper bound on the number of observations taken during an experiment.
- *Constrained* sequential procedures, the version of the problem considered in this thesis, for which there is a given maximum number of observations allowed, called the budget b.

Existing sequential procedures do not necessarily minimize the expected number of observations, but are heuristics for the MSP described above.

We note that Procedure $\mathcal{M}_{\text{BEM}}$ sometimes undertakes what can be regarded as needless sampling. For example, if $k = 2$ and $n = 100$, and we obtain $\eta_{1,100} = 90$ and $\eta_{2,100} = 10$, then we clearly could have stopped sampling quite a bit earlier, yet still reached the conclusion that alternative 1 is the most probable for the given $P^\star$-requirement on (prior) $\mathsf{P}(\text{CS})$. The curtailed sequential procedure of Bechhofer and Kulkarni (BK) (1984) achieves efficiencies over Procedure $\mathcal{M}_{\text{BEM}}$ by capitalizing on such favorable sample paths, that is, sample

paths that allow the procedure to stop before taking all n multinomial observations called for by the single-stage procedure. When otherwise ambiguous, we place a subscript on n to indicate the appropriate procedure, for example n_{BEM} for Procedure $\mathcal{M}_{\text{BEM}}$.

Procedure $\mathcal{M}_{\text{BK}}$

- For the given k, $\theta^\star$, and $P^\star$, specify n prior to the start of sampling. (Typically, n will be the same value as n_{BEM}.)
- At the mth stage of experimentation, $m \geq 1$, take a random multinomial observation.
- Calculate the ordered cumulative successes $\eta_{[i]m}$, $i = 1, 2, \ldots, k$. Stop sampling at the first stage when

$$\eta_{[k]m} - \eta_{[k-1]m} \;\geq\; n - m. \tag{2.1}$$

- Select $\hat{i}_m$ as the best alternative, randomizing to break ties.

In other words, Procedure $\mathcal{M}_{\text{BK}}$ employs a curtailment strategy that stops sampling at the first stage m for which the alternative currently in first place can do no worse than tie if the remaining $n - m$ observations were to be taken. Let N be a random variable denoting the value of m at the termination of multinomial observations. It can be shown that the curtailed Procedure $\mathcal{M}_{\text{BK}}$ yields the same $\mathsf{P}(\text{CS})$ as the single-stage Procedure $\mathcal{M}_{\text{BEM}}$, yet with a smaller expected number of observations, i.e., for all $\boldsymbol{p}$,

$$\mathsf{P}_{\boldsymbol{p}}(\text{CS using Procedure } \mathcal{M}_{\text{BK}}) \;=\; \mathsf{P}_{\boldsymbol{p}}(\text{CS using Procedure } \mathcal{M}_{\text{BEM}})$$

and

$$\mathsf{E}_{\boldsymbol{p}}[N_{\text{BK}}] \;\leq\; n_{\text{BEM}},$$

where $\mathsf{E}_{\boldsymbol{p}}[N_{\text{BK}}]$ is the expectation of N_{BK} under configuration $\boldsymbol{p}$.

Remark 2.2 Since $\mathsf{P}_{\boldsymbol{p}}(\text{CS})$ for both procedures is identical when $n_{\text{BK}} = n_{\text{BEM}}$, the SC for Procedure $\mathcal{M}_{\text{BK}}$ must be the LFC as Kesten and Morse (1959) proved for Procedure $\mathcal{M}_{\text{BEM}}$. Thus, the parameter n_{BK}, chosen based upon the procedure performance in the SC, satisfies condition (1.1) for the given k, $\theta^\star$, and $P^\star$.

In the course of our research, we developed and implemented algorithms to calculate the performance parameters of the primary procedures we discuss here. Appendix A.1 includes tables for Procedure $\mathcal{M}_{\mathrm{BK}}$ with $\mathsf{E}_{\mathrm{SC}}[N]$ and $\mathsf{E}_{\mathrm{EPC}}[N]$ for a range of k, $\theta^\star$, and n values, where EPC is the *equal probability configuration*, or $\boldsymbol{p} = (1/k, 1/k, \ldots, 1/k)$. Estimates for the expected number of observations in the EPC are considered worst-case estimates.

Bechhofer and Kulkarni (1984) describe the curtailment in their procedure as *strong* curtailment. An earlier bounded sequential procedure developed by Gibbons et al. (1977) is similar, but uses a strict inequality in the curtailment stopping condition. Such curtailment is known as *weak* curtailment. It also results in the same $\mathsf{P}_{\boldsymbol{p}}(\mathrm{CS})$ but with a slightly larger expected number of observations.

Example 2.1 For $k = 3$ and $n = 2$, stop sampling if

m	x_{1m}	x_{2m}	x_{3m}	η_{1m}	η_{2m}	η_{3m}
1	0	0	1	0	0	1

and select alternative 3, because $\eta_{[k]m} - \eta_{[k-1]m} = 1 \geq n - m = 2 - 1 = 1$. □

Example 2.2 For $k = 3$ and $n = 3$ or 4, stop sampling if

m	x_{1m}	x_{2m}	x_{3m}	η_{1m}	η_{2m}	η_{3m}
1	0	0	1	0	0	1
2	0	0	1	0	0	2

and select alternative 3, because $\eta_{[k]m} - \eta_{[k-1]m} = 2 \geq n - m = n - 2$ for $n = 3$ or $n = 4$. □

Example 2.3 For $k = 3$ and $n = 3$, suppose that

m	x_{1m}	x_{2m}	x_{3m}	η_{1m}	η_{2m}	η_{3m}
1	0	1	0	0	1	0
2	1	0	0	1	1	0
3	0	0	1	1	1	1

Because $n = m$, we stop sampling and randomize among the three alternatives. □

The above examples involving Procedure $\mathcal{M}_{\mathrm{BK}}$ suggest that sequential procedures might prove efficacious in terms of sampling efficiency; and, in fact, a number of intuitively appealing sequential procedures have been studied in the literature. Cacoullos and Sobel (1966) propose a bounded sequential procedure in which sampling continues until the alternative with the greatest number of successes reaches a predetermined inverse sampling parameter t (i.e., stops sampling when $\eta_{[k]m} = t$). The inverse sampling parameter is chosen for a given k and $\theta^\star$ to guarantee a particular $P^\star$. The authors show that the SC is the LFC for their procedure.

Alam (1971) recommends an unbounded sequential procedure in which sampling continues until the difference between the two best alternatives is equal to the parameter r (i.e., $\eta_{[k]m} - \eta_{[k-1]m} = r$). He shows that the SC is the LFC for his procedure when $k = 2$. He conjectures that the same is true for $k > 2$ and substantiates his claim via Monte Carlo methods. Alam and Thompson (1972) develop a similar procedure for identifying the *least* probable multinomial alternative. Bhandari and Ali (1994) generalize Alam's (1971) procedure to selection of the s best cells ($s \geq 1$) and prove Alam's conjecture about the LFC for $k > 2$.

Ramey and Alam (RA) (1979) propose a procedure that combines the inverse sampling of the Cacoullos and Sobel procedure with the difference stopping rule of the Alam procedure.

Procedure $\mathcal{M}_{\mathrm{RA}}$

- For the given k, $P^\star$, and $\theta^\star$, find (r, t) from an appropriate table (e.g., Bechhofer and Goldsman (BG) 1985a or Appendix A.3 of this thesis).
- At the mth stage of experimentation, $m \geq 1$, take a random multinomial observation.
- Calculate the ordered cumulative successes $\eta_{[i]m}$, $i = 1, 2, \ldots, k$. Stop sampling at the first stage when

$$\eta_{[k]m} \;=\; t \quad \text{or} \quad \eta_{[k]m} - \eta_{[k-1]m} \;=\; r. \tag{2.2}$$

- Select $\hat{i}_m$ as the best alternative; ties are not possible.

Remark 2.3 RA show that the SC is the LFC for their procedure when $k = 2$, and conjecture that it is so for $k > 2$, based upon empirical evidence. Therefore, the (r, t)-values in BG (1985a) and in Appendix A.3 of this thesis have been chosen to minimize the expected number of observations taken by Procedure $\mathcal{M}_{\mathrm{RA}}$ when $\boldsymbol{p}$ is the SC, satisfying condition (1.2), but not necessarily condition (1.1).

Remark 2.4 Empirical evidence also led RA to conclude that when there are several pairs (r, t) that satisfy condition (1.2), the pair with the lowest value of r corresponds to the lowest expected number of observations. While this is true in most cases, it is not true in general. Consider $k = 3$, $\theta^\star = 2$, and $P^\star = 0.75$. For parameter pair $(4, 5)$, $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS}) = 0.756$ and $\mathsf{E}_{\mathrm{SC}}[N] = 8.809$, but for parameter pair $(3, 6)$, $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS}) = 0.773$ and $\mathsf{E}_{\mathrm{SC}}[N] = 8.825$. Thus, the pair with the higher r has the lower expected number of observations in this case. This is bad news when searching for an optimal pair across the feasible domains for r and t because it makes the search somewhat more burdensome.

Remark 2.5 For $k = 2$, Procedure $\mathcal{M}_{\mathrm{RA}}$ is a gambler's ruin problem subject to a constraint on the total number of wins or losses dictated by t. We discuss this further in Chapter 6.

Example 2.4 For $k = 3$, $\theta^\star = 3$, and $P^\star = 0.75$, the Procedure $\mathcal{M}_{\mathrm{RA}}$ table reveals that the required (r, t)-pair is $(2, 3)$. Suppose that

m	x_{1m}	x_{2m}	x_{3m}	η_{1m}	η_{2m}	η_{3m}
1	1	0	0	1	0	0
2	0	0	1	1	0	1
3	1	0	0	2	0	1
4	0	1	0	2	1	1
5	1	0	0	3	1	1

Because $\eta_{[3]5} = 3 = t$, we stop sampling due to the inverse sampling rule. We would also stop due to the difference rule, since $\eta_{[3]5} - \eta_{[2]5} = 2 = r$. $\square$

The next procedure is from BG (1985b, 1986), and is a truncated version of an unbounded sequential procedure due to BKS (1968), which is itself based on a sequential

probability ratio test. This procedure stops sampling as soon as one of several stopping criteria are met, including curtailment.

Procedure $\mathcal{M}_{\mathrm{BG}}$

- For the given k, $P^\star$, and $\theta^\star$, find the truncation number n from the table given in BG (1986) or BSG (1995).
- At the mth stage of experimentation, $m \geq 1$, take a random multinomial observation.
- Calculate the ordered cumulative successes $\eta_{[i]m}$, $i = 1, 2, \ldots, k$, and the quantity

$$z_m \;=\; \sum_{i=1}^{k-1} \Big(\frac{1}{\theta^\star}\Big)^{\eta_{[k]m} - \eta_{[i]m}}.$$

  Stop sampling at the first stage when either

$$z_m \;\leq\; (1 - P^\star)/P^\star \quad \text{or} \quad \eta_{[k]m} - \eta_{[k-1]m} \;\geq\; n - m. \tag{2.3}$$

- Select $\hat{i}_m$ as the best alternative, randomizing to break ties.

Remark 2.6 For the unbounded procedure upon which Procedure $\mathcal{M}_{\mathrm{BG}}$ is based, BKS prove that the LFC is the SC; see also Levin (1984). BG acknowledge that both the BKS procedure and Procedure $\mathcal{M}_{\mathrm{BEM}}$ share the same LFC, but they do not prove that combining the stopping rules of these two procedures by adding a truncation point to the BKS procedure actually preserves the LFC in the new procedure. We have seen no evidence to doubt that conjecture; however, we do not have a proof that it is true. Therefore, the tabulated n-values in BG (1986) and BSG (1995) minimize the expected number of observations taken by Procedure $\mathcal{M}_{\mathrm{BG}}$ when $\boldsymbol{p}$ is the SC, satisfying condition (1.2), but not necessarily condition (1.1).

Remark 2.7 Many of the procedures that have a strong curtailment stopping rule also include the fixed sample size stopping rule. In this procedure, for example, both $\eta_{[k]m} - \eta_{[k-1]m} \geq n - m$ and $m = n$ are often explicitly listed as stopping criteria. We have chosen to remove this redundancy here and for other similar procedures, since the first of the two rules is always satisfied whenever the second is.

Example 2.5 Suppose $k = 3$, $\theta^\star = 3$, and $P^\star = 0.75$. The table tells us to truncate sampling at $n = 5$ observations. For the data

m	x_{1m}	x_{2m}	x_{3m}	η_{1m}	η_{2m}	η_{3m}
1	1	0	0	1	0	0
2	1	0	0	2	0	0

we stop sampling by the first criterion in (2.3) because $z_2 = (1/3)^2 + (1/3)^2 = 2/9 \leq (1 - P^\star)/P^\star = 1/3$, and we select alternative 1. □

Example 2.6 Again suppose $k = 3$, $\theta^\star = 3$, and $P^\star = 0.75$ (so that $n = 5$). For the data

m	x_{1m}	x_{2m}	x_{3m}	η_{1m}	η_{2m}	η_{3m}
1	0	1	0	0	1	0
2	0	0	1	0	1	1
3	0	0	1	0	1	2
4	0	1	0	0	2	2
5	0	0	1	0	2	3

we stop sampling by the second criterion in (2.3) because $\eta_{[k]m} - \eta_{[k-1]m} = 1 \geq m - n = 5 - 5 = 0$, and we select alternative 3. □

Example 2.7 Yet again suppose $k = 3$, $\theta^\star = 3$, and $P^\star = 0.75$ (so that $n = 5$). For the data

m	x_{1m}	x_{2m}	x_{3m}	η_{1m}	η_{2m}	η_{3m}
1	0	0	1	0	0	1
2	1	0	0	1	0	1
3	1	0	0	2	0	1
4	0	0	1	2	0	2
5	0	1	0	2	1	2

we stop according to the second criterion in (2.3) since $\eta_{[k]m} - \eta_{[k-1]m} = 0 \geq m - n = 5 - 5 = 0$, but we now have a tie between $\eta_{1,5}$ and $\eta_{3,5}$, so we randomly select between alternatives 1 and 3. □

Example 2.8 Suppose $k = 4$, $\theta^\star = 1.6$, $P^\star = 0.75$. The single-stage Procedure $\mathcal{M}_{\rm BEM}$ requires 46 observations to guarantee (1.1), whereas Procedure $\mathcal{M}_{\rm BG}$ (with $n_{\rm BG} = 57$) has $\mathsf{E}_{\rm SC}[N] = 31.1$ and $\mathsf{E}_{\rm EPC}[N] = 37.65$. $\square$

Chen (1988a) proposes a bounded sequential procedure that combines the curtailment of Procedure $\mathcal{M}_{\rm BEM}$ with Cacoullos and Sobel's inverse sampling procedure.

Procedure $\mathcal{M}_{\rm C}$

- For the given k, $\theta^\star$, and $P^\star$, find (n, t) from an appropriate table (e.g., Chen 1988a).
- At the mth stage of experimentation, $m \geq 1$, take a random multinomial observation.
- Calculate the ordered cumulative successes $\eta_{[i]m}$, $i = 1, 2, \ldots, k$. Stop sampling at the first stage when

$$\eta_{[k]m} \;=\; t \quad \text{or} \quad m = n. \tag{2.4}$$

- Select $\hat{i}_m$ as the best alternative, randomizing to break ties.

Remark 2.8 Chen proves that the SC is the LFC. Thus, his tabulated (n, t)-pairs, based upon procedure performance in the SC, satisfy condition (1.1) for a given k, $\theta^\star$, and $P^\star$.

Remark 2.9 Chen states that the strong curtailment stopping rule (2.1) of Procedure $\mathcal{M}_{\rm BK}$ could be used to reduce the expected number of observations for his procedure without affecting $\mathsf{P}(\text{CS})$, but he does not implement the change. In Chapter 6, we modify his procedure by incorporating curtailment, renaming it Procedure $\mathcal{M}_{\rm C'}$, and tabulate the results for common choices of k, $\theta^\star$, and $P^\star$, in Appendix A.2.

Example 2.9 For $k = 3$, $\theta^\star = 3$, and $P^\star = 0.75$, Chen's table reveals that the required (n, t)-pair is $(5, 3)$. Suppose that

m	x_{1m}	x_{2m}	x_{3m}	η_{1m}	η_{2m}	η_{3m}
1	1	0	0	1	0	0
2	0	0	1	1	0	1
3	1	0	0	2	0	1
4	0	1	0	2	1	1
5	0	0	1	2	1	2

Because $m = 5 = n$, we stop sampling and select the best alternative by randomizing between alternatives 1 and 3. □

Example 2.10 Again let $k = 3$, $\theta^\star = 3$, and $P^\star = 0.75$, so the required (n, t)-pair is $(5, 3)$. Suppose that

m	x_{1m}	x_{2m}	x_{3m}	η_{1m}	η_{2m}	η_{3m}
1	1	0	0	1	0	0
2	0	0	1	1	0	1
3	1	0	0	2	0	1
4	1	0	0	3	0	1

Because $\eta_{[k]m} = 3 = t = 3$, we stop sampling, and choose alternative 1 as the best. □

Chen (1992) proposes a modified Procedure $\mathcal{M}_{\mathrm{RA}'}$ that adds truncation at point n (with curtailment) to Procedure $\mathcal{M}_{\mathrm{RA}}$.

Procedure $\mathcal{M}_{\mathrm{RA}'}$

- For the given k, $\theta^\star$, and $P^\star$, find the (n, r, t)-triplet from an appropriate table (e.g., Chen 1992).
- At the mth stage of experimentation, $m \geq 1$, take a random multinomial observation.
- Calculate the ordered cumulative successes $\eta_{[i]m}$, $i = 1, 2, \ldots, k$. Stop sampling at the first stage when

$$\eta_{[k]m} = t \quad \text{or} \quad \eta_{[k]m} - \eta_{[k-1]m} = r \quad \text{or} \quad \eta_{[k]m} - \eta_{[k-1]m} \geq n - m. \tag{2.5}$$

- Select $\hat{i}_m$ as the best alternative, randomizing to break ties.

Chen points out that $\mathsf{P}(\mathrm{CS})$ is unchanged for all t such that $t \geq n/2$ with fixed r and n. Similarly, he shows that $\mathsf{P}(\mathrm{CS})$ is also unchanged for all r such that $r \geq t$ for fixed t and n. We can then restrict our search for the optimal procedure parameter triplet (n, r, t) to $r \leq t \leq n/2$.

Remark 2.10 Chen conjectures that the LFC of Procedure $\mathcal{M}_{\mathrm{RA}'}$ is the SC. He then proves a partial result to the conjecture, namely, that the LFC is of the form

$$\boldsymbol{p} = \{0, 0, \ldots, 0, s, p, p, \ldots, p, \theta^{\star} p\},$$

where $0 \leq s \leq p$. Since his procedure is a generalization of Procedure $\mathcal{M}_{\mathrm{RA}}$, this is also a partial proof of Ramey and Alam's (1979) conjecture concerning the LFC of their procedure. The (n, r, t)-values in Chen (1992) have been chosen to minimize the expected number of observations taken by Procedure $\mathcal{M}_{\mathrm{RA}'}$ when $\boldsymbol{p}$ is the SC, satisfying condition (1.2), but not necessarily condition (1.1).

For the sake of brevity, we will not include examples here, since the previous examples for other procedures that use one or more of the same stopping rules have already been discussed.

2.2.2 Other MSP Procedures

We briefly review other types of MSP procedures that have a primary goal other than selection of the single best alternative or that use the MSP as a non-parametric approach to RS.

A great deal of research has been carried out for subset selection problems in which the experimenter wishes to find a subset of alternatives which includes the best (or worst). Gupta and Nagel (1967) propose a single-stage procedure for selecting a subset containing the best or worst alternative. Bechhofer and Chen (1991) improve Gupta and Nagel's procedure by integrating the strong curtailment stopping rule. Panchapakesan (1971) developed a subset selection procedure based upon an inverse sampling rule. Chen and Hsu

(1991a) create a procedure that combines four stopping rules based upon fixed sample size, curtailment, inverse sampling, and the difference in number of successes between the best two alternatives. Chen and Hsu (1991b) also develop a similar procedure for selecting the subset with the least probable alternative. Chen and Sobel (1987) and Chen (1988b) present a procedure that integrates the indifference zone and subset selection procedures by guaranteeing a $\mathsf{P}(\mathrm{CS})$ of the best alternative if the probability configuration is in the PZ and a probability of selecting a subset containing the best alternative if the configuration is in the IZ. Vieira et al. (2012) generalize Procedure $\mathcal{M}_{\mathrm{BG}}$ for application to subset selection problems with a large number of alternatives under a restriction on the maximum subset size.

Ramey and Alam (1980) propose a sequential procedure for selecting the most probable alternative that is based upon a Bayesian approach when the parameters of a multinomial distribution are known to follow a Dirichlet distribution. Bechhofer et al. (1989) extend the single-stage Procedure $\mathcal{M}_{\mathrm{BEM}}$, designed for single factor experiments, to a single-stage procedure designed for a multiple factor experiment.

Aoshima and Chen (1999) and Aoshima et al. (2003) propose procedures for selecting the best multinomial alternative in the presence of a nuisance alternative. In such a problem, the experimenter takes observations from k alternatives, but wants only to know the best of a particular subset of size $k-1$. The remaining alternative is considered the nuisance alternative.

Chen and Hwang (1984), with reference to Marshall and Olkin (1979), and Bhandari and Bose (1987), address the LFC for selecting the best multinomial alternative when the indifference zone is defined by a location parameter (i.e., $p_{[k]} \geq p_{[k-1]} + a$, where a is the specified indifference zone parameter) instead of a relative risk indifference zone parameter. Bhandari and Bose (1989) address least favorable considerations for fixed sample size subset selection procedures.

Miller et al. (1998) propose a procedure for experiments that compare alternatives via a numerical performance measure, but formulate the problem as an MSP rather than a parametric selection problem. Their procedure is called All Vector Comparisons (AVC) and

involves creating "pseudoreplications" out of all possible permutations of the k independent vectors of observations for each alternative. They then tabulate the total number of replications required to achieve a specified $P^\star$ given k and $\theta^\star$, and compare the performance to Procedure $\mathcal{M}_{\text{BEM}}$. Procedure efficiency under varying performance measure probability distributions is also examined. Vieira et al. (2010) report a more thorough comparison of AVC and Procedure $\mathcal{M}_{\text{BEM}}$. One of their primary conclusions is that AVC does not work in populations for which the best alternative has the greatest probability of achieving the highest value of the performance measure, but does not have the largest expected value of the measure.

2.3 Observation Cost

For all of the procedures discussed thus far, the goal of interest for procedure efficiency is the minimization of the expected number of observations. The underlying assumption is that each observation has constant cost (i.e., the first observation is no more expensive than the last). None of the literature for MSP procedures addresses or considers variable observation costs. Other papers have, however, considered particular aspects of cost in RS problems other than MSPs. All of the examples we highlight below assume normally distributed alternative outputs. Note that we use the term 'cost' as a proxy for any expenditure (e.g., effort, resources, penalties, opportunity cost, etc.), not just monetary.

Hong and Nelson (2005) consider the setup costs of switching between alternatives during sampling. Their procedures minimize switching while still providing the same statistical guarantees of traditional RS procedures. Chick and Inoue (2001) develop Bayesian procedures (for both constrained and unconstrained observation budgets) that build upon earlier Bayesian formulations of Gupta and Miescke (1996). Their procedures trade off the cost of continued observations with expected opportunity costs, i.e., the expected loss due to potentially choosing an inferior alternative. Their procedures also allow for a different observation cost for each alternative, but each remains constant throughout the experiment. Chick and Gans (2009) solve a selection problem in which the goal is to maximize the net present value of the alternative that is eventually selected, accounting for discounted cash

flows and observation costs. They assume that all observation costs are constant (but discounted based upon sampling duration). Their procedure extends that of Chick and Inoue by using a Bayesian formulation that treats the decision take an observation as a real option.

Chen (1995) and subsequent papers by his colleagues use an alternate Bayesian formulation called optimal computing budget allocation (OCBA) procedures. The goal of their procedures is to allocate simulation lengths efficiently to the alternatives under consideration in order to minimize computation time (observation costs) while still guaranteeing an approximate confidence probability. He et al. (2007) extend Chen's work by using OCBA to minimize the expected opportunity cost (instead of improving confidence probabilities). For a comparison of indifference zone, expected opportunity cost, and OCBA procedures, see Branke et al. (2007).

We were unable to find any RS literature that considers variable observation costs for within-alternative sampling, with the exception of the very narrow application to switching costs in Hong and Nelson (2005). Furthermore, none of the literature considering costs specifically addresses alternative output other than normally distributed measures.

CHAPTER III

OPTIMAL PROCEDURES

We next formulate the selection problem in ways that allow the performance characteristics of the procedures to be evaluated via linear programming (LP) and mixed integer programming (MIP) techniques. In particular, we propose an LP formulation that yields an optimal randomized sequential procedure, as well as an MIP formulation that yields a non-randomized sequential procedure possessing certain optimality properties.

3.1 Optimal Multinomial Selection

In this section, we describe the development of our new procedures, which are based upon mathematical programming techniques. §3.1.1 provides the motivation behind our research. §3.1.2 describes our generalized procedure, followed by the mathematical formulation of the problem in §3.1.3.

3.1.1 Motivation

In the constrained version of the MSP, considered in this thesis, there is a given maximum number of observations allowed, called the budget b, and the problem is to choose dynamically the number $n \leq b$ of observations such that if $\hat{i}_n$ is chosen as described in Chapter 1, then condition (1.2) holds, and the (prior) expected number of observations is minimized.

In the most general sense, any vector of cumulative successes $\boldsymbol{\eta}$ is a viable stopping vector or point. All of the reviewed procedures use stopping rules that are based on some simple relationships between the components of the cumulative success vector and the specified procedure parameters. Therefore, those stopping rules are special cases of a general procedure that specifies the particular cumulative success vectors that are stopping points for the experiment. We remark that the specified set of stopping points may be infinite. Stopping rules based upon simple relationships between components of $\boldsymbol{\eta}$, such as those reviewed in Chapter 2, are much easier to apply in practice and to describe than rules that

specify an entire set of enumerated stopping points; however, given the ubiquity of computing resources today, there is no reason to believe that the specification of even thousands of stopping points is impractical to implement.

In any case, we do not need to enumerate every possible stopping point. Since our assignment of alternative i to component η_i of the cumulative success vector is arbitrary for all $i = 1, 2, \ldots, k$, then our procedure must be invariant to those assignments. As an example with $k = 3$, if the vector $(3, 2, 1)$ is a stop, then so is any member of the set of its permutations $\{(3, 2, 1), (3, 1, 2), \ldots, (1, 2, 3)\}$. That symmetry means that we could enumerate our set of stopping points in terms of only one of each set of permutations, say the *left-lexicographic* permutation denoted $\boldsymbol{\eta}' \equiv (\eta_{[k]}, \eta_{[k-1]}, \ldots, \eta_{[1]})$, with the understanding that if $\boldsymbol{\eta}'$ is a stop, then so are all of its permutations.

The discretized nature of multinomial trials ensures that the set of potential stopping points is countable. If our problem is bounded, then our set is also finite. As a result, we can enumerate all of the possible subsets of stopping points, although the number of such subsets may be prohibitively large. Not all possible subsets are feasible, however. For instance, if $(2, 0, 0)$ is a stopping point for $k = 3$, then $(3, 0, 0)$ is infeasible, since the latter vector can never be reached. But even the set of feasible subsets can be quite large. Fortunately, we can use mathematical programming techniques as an alternative to complete enumeration.

Mathematical programming techniques give us an additional benefit beyond an ability to optimize over the feasible set of stopping points. All existing MSP procedures assume that the conditional probability of stopping at a particular cumulative success vector $\boldsymbol{\eta}$ given that we have arrived there, is either zero or one. Mathematical programming, however, allows us to consider a generalization of the problem in which the conditional probability of stopping at $\boldsymbol{\eta}$ may be any value in the interval $[0, 1]$. We call this generalized formulation of the problem a *randomized* formulation. If we require that all of the conditional stopping probabilities equal zero or one, we call that formulation a *non-randomized* formulation. Only non-randomized procedures can be considered when using enumeration, since allowing randomized stops makes the number of possible procedures infinite.

The following trivial example shows why we must consider randomization for true optimality. Consider a $k = 2$ problem, with $\theta^\star = 4$ and $P^\star = 0.75$. In the SC, $\boldsymbol{p} = (0.8, 0.2)$. Taking one observation and choosing the alternative that wins is the optimal non-randomized procedure, with $\mathsf{E}_{\mathrm{SC}}[N] = 1$ and $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS}) = 0.8$. However, suppose we stop without taking any observations with probability 1/6 and choose an alternative at random. If we do not stop, we will take one observation and choose the alternative that wins. Then,

$$\mathsf{E}_{\mathrm{SC}}[N] = \left(\frac{1}{6}\right)(0) + \left(\frac{5}{6}\right)(1) = \frac{5}{6},$$

and

$$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS}) = \left(\frac{1}{6}\right)\left(\frac{1}{2}\right) + \left(\frac{5}{6}\right)\left(\frac{4}{5}\right) = \frac{3}{4} = P^\star.$$

Thus, randomizing stops results in a lower expected number of observations in the SC than the non-randomized procedure, while still achieving $P^\star$.

3.1.2 Procedure

To define our new procedures, we need some additional notation. Let the set of all possible cumulative success vectors be denoted $\mathcal{N} \equiv \left\{\boldsymbol{\eta} : \sum_{i=1}^{k} \eta_i \leq b\right\}$. Let $p_{\boldsymbol{\eta}} \equiv \mathsf{P}(\text{stop at } \boldsymbol{\eta} \,|\, \text{arrive at } \boldsymbol{\eta})$ be the conditional probability of stopping at $\boldsymbol{\eta}$. We denote the set of ordered pairs representing the potential stopping points for a randomized procedure as $\mathcal{S}_{\mathrm{R}} \equiv \{(\boldsymbol{\eta}, p_{\boldsymbol{\eta}}) : \boldsymbol{\eta} \in \mathcal{N} \text{ and } p_{\boldsymbol{\eta}} > 0\}$. For non-randomized procedures, $\mathcal{S}_{\mathrm{NR}} \equiv \{\boldsymbol{\eta} : \boldsymbol{\eta} \in \mathcal{N} \text{ and } p_{\boldsymbol{\eta}} = 1\}$ is the set of stopping points. The ordered pair notation is unnecessary for $\mathcal{S}_{\mathrm{NR}}$, since all potential stopping points have $p_{\boldsymbol{\eta}} = 1$.

For example, consider a $k = 2$ problem and a randomized procedure with eight potential stopping points and associated conditional stopping probabilities:

$$\mathcal{S}_{\mathrm{R}} = \left\{ \begin{array}{cccc} \big((2,0), 0.75\big), & \big((3,0), 1\big), & \big((3,1), 1\big), & \big((3,2), 1\big), \\ \big((0,2), 0.75\big), & \big((0,3), 1\big), & \big((1,3), 1\big), & \big((2,3), 1\big) \end{array} \right\}.$$

According to this set, if we arrive at either $(2, 0)$ or $(0, 2)$, we will stop with probability 0.75. If we arrive at any of the other success vectors in $\mathcal{S}_{\mathrm{R}}$, we will stop with probability 1. Although all permutations of a particular stopping point are included in $\mathcal{S}_{\mathrm{R}}$, we could have represented the set more simply by including only the left-lexicographic vectors, as we noted

in §3.1.1, provided it is understood that all permutations of those vectors are also stops with the same conditional stopping probability.

The following generic Procedures $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{NR}}$ are the randomized and non-randomized procedures, respectively.

Procedure $\mathcal{M}_{\mathrm{R}}$

- For the given k, $P^\star$, $\theta^\star$, and b, derive $\mathcal{S}_{\mathrm{R}}$.
- At the mth stage of experimentation, $m \geq 1$, take a random multinomial observation.
- Calculate the cumulative success vector $\boldsymbol{\eta}$. If $\boldsymbol{\eta} \in \mathcal{S}_{\mathrm{R}}$ then stop sampling with probability $p_{\boldsymbol{\eta}}$ (determined by randomization).
- Select $\hat{i}_m$ as the best alternative, randomizing to break ties.

Procedure $\mathcal{M}_{\mathrm{NR}}$

- For the given k, $P^\star$, $\theta^\star$, and b, derive $\mathcal{S}_{\mathrm{NR}}$.
- At the mth stage of experimentation, $m \geq 1$, take a random multinomial observation.
- Calculate the cumulative success vector $\boldsymbol{\eta}$. If $\boldsymbol{\eta} \in \mathcal{S}_{\mathrm{NR}}$ then stop sampling.
- Select $\hat{i}_m$ as the best alternative, randomizing to break ties.

3.1.3 Formulation

Here we describe the randomized and non-randomized implementations of our formulation — a linear program and a mixed integer program, respectively. For conciseness, we will display the MIP and then point out which variables and constraints are not in the LP.

We will discuss our implementation for the case $k = 3$ because it has complexities that are not obvious with $k = 2$ (due to permutations), but is still reasonably straightforward in representation. Consider an experiment with a given $P^\star$, $\theta^\star$, and budget b. We assume that

η_1, η_2, and $\eta_3 \geq 0$. Let $\mathcal{B} \equiv \left\{\boldsymbol{\eta} : \sum_{i=1}^{k} \eta_i = b\right\}$ be the set of possible cumulative success vectors after the bth observation. Without loss of generality, we assume a probability configuration, $\boldsymbol{p} = (p_1, p_2, p_3)$, in which $p_1 \geq p_2 \geq p_3$ and $\sum_{i=1}^{3} p_i = 1$. We use the parameter θ_i to represent the ratio of the best probability of success (alternative 1) to that of alternative i, so $\theta_i \equiv p_1/p_i$, for $i = 1, 2, 3$.

We show our MIP (non-randomized) formulation in full and then explain each component. Our LP (randomized) formulation can be obtained from the MIP by removing the binary variables and the constraints that involve those variables. We will point out the differences between the MIP and the LP as they arise. Let $\mathbf{0}$ be the vector $(0, 0, 0)$. The full formulation, consisting of equations (3.1)–(3.10), is shown in Figure 3.1.

3.1.3.1 Variables

Our key insight is to model the MSP as a network. The flow through the network represents the flow of probability. Each node in the network is a vector $\boldsymbol{\eta}$ through which the multinomial sample paths may go. Figure 3.2 is a graphical depiction of the flow variables discussed in this section. The following is a brief description of each type of variable.

- $f_{\boldsymbol{\eta}}$ are continuous, non-negative variables that represent the probability that flows from node $\boldsymbol{\eta}$ out of the network through an arc to the sink. All nodes have arcs to the sink. A nonzero value for this variable indicates that $\boldsymbol{\eta}$ is a potential stopping node. Specifically,

$$\begin{aligned} f_{\boldsymbol{\eta}} &= \mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and stop at } \boldsymbol{\eta}) \\ &= \mathsf{P}(\text{stop at } \boldsymbol{\eta} \,|\, \text{arrive at } \boldsymbol{\eta})\, \mathsf{P}(\text{arrive at } \boldsymbol{\eta}) \\ &= p_{\boldsymbol{\eta}}\, \mathsf{P}(\text{arrive at } \boldsymbol{\eta}). \end{aligned} \tag{3.11}$$

  Deriving the values of all $p_{\boldsymbol{\eta}}$ is the ultimate goal of the optimization, as they define the procedure.

- $f^i_{\boldsymbol{\eta}}$, $i = 1, 2, 3$, are the probabilities that flow through an outbound arc from $\boldsymbol{\eta}$ with $\sum_{i=1}^{k} \eta_i = m - 1$ due to a success by alternative i on the next observation, m; see

$$\min_{f_{\boldsymbol{\eta}}, f^1_{\boldsymbol{\eta}}, f^2_{\boldsymbol{\eta}}, f^3_{\boldsymbol{\eta}}, Y_{\boldsymbol{\eta}}} \quad \sum_{\boldsymbol{\eta} \in \mathcal{N}} (f^1_{\boldsymbol{\eta}} + f^2_{\boldsymbol{\eta}} + f^3_{\boldsymbol{\eta}}) \tag{3.1}$$

Subject to:

$$\text{(Initialization)} \qquad f_{\mathbf{0}} + f^1_{\mathbf{0}} + f^2_{\mathbf{0}} + f^3_{\mathbf{0}} = 1, \tag{3.2}$$

$$\text{(Flow Conservation)} \qquad \begin{aligned} f_{\boldsymbol{\eta}} + f^1_{\boldsymbol{\eta}} + f^2_{\boldsymbol{\eta}} + f^3_{\boldsymbol{\eta}} - f^1_{(\eta_1-1,\eta_2,\eta_3)} \\ - f^2_{(\eta_1,\eta_2-1,\eta_3)} - f^3_{(\eta_1,\eta_2,\eta_3-1)} = 0, \end{aligned} \quad \forall \boldsymbol{\eta} \in \mathcal{N} \setminus \mathbf{0} \tag{3.3}$$

$$\text{(Balance Outflow)} \qquad \begin{aligned} f^1_{\boldsymbol{\eta}} - \theta_2 f^2_{\boldsymbol{\eta}} = 0, \\ f^1_{\boldsymbol{\eta}} - \theta_3 f^3_{\boldsymbol{\eta}} = 0, \end{aligned} \quad \forall \boldsymbol{\eta} \in \mathcal{N} \setminus \mathcal{B} \tag{3.4}$$

$$\text{(Symmetry)} \qquad \begin{aligned} f_{\boldsymbol{\eta}} - {\theta_2}^{\eta_1-\eta_2} f_{(\eta_2,\eta_1,\eta_3)} = 0, \\ f_{\boldsymbol{\eta}} - {\theta_3}^{\eta_1-\eta_3} f_{(\eta_3,\eta_2,\eta_1)} = 0, \\ f_{\boldsymbol{\eta}} - \left(\tfrac{\theta_3}{\theta_2}\right)^{\eta_2-\eta_3} f_{(\eta_1,\eta_3,\eta_2)} = 0, \\ f_{\boldsymbol{\eta}} - \left(\tfrac{1}{\theta_2}\right)^{\eta_2-\eta_3} {\theta_3}^{\eta_1-\eta_3} f_{(\eta_2,\eta_3,\eta_1)} = 0, \\ f_{\boldsymbol{\eta}} - {\theta_2}^{\eta_1-\eta_2} {\theta_3}^{\eta_2-\eta_3} f_{(\eta_3,\eta_1,\eta_2)} = 0, \end{aligned} \quad \begin{aligned} \forall \boldsymbol{\eta} \in \mathcal{N} \text{ s.t.} \\ \eta_1 \geq \eta_2 \geq \eta_3 \end{aligned} \tag{3.5}$$

$$\text{(Minimum P(CS))} \qquad \sum_{\substack{\boldsymbol{\eta} \in \mathcal{N} \\ \eta_1 > \max\{\eta_2, \eta_3\}}} f_{\boldsymbol{\eta}} + \frac{1}{2} \sum_{\substack{\boldsymbol{\eta} \in \mathcal{N} \\ \eta_1 = \eta_2 > \eta_3 \text{ or} \\ \eta_1 = \eta_3 > \eta_2}} f_{\boldsymbol{\eta}} + \frac{1}{3} \sum_{\substack{\boldsymbol{\eta} \in \mathcal{N} \\ \eta_1 = \eta_2 = \eta_3}} f_{\boldsymbol{\eta}} \geq P^\star, \tag{3.6}$$

$$\text{(Stop)} \qquad f_{\boldsymbol{\eta}} - Y_{\boldsymbol{\eta}} \leq 0, \quad \forall \boldsymbol{\eta} \in \mathcal{N} \tag{3.7}$$

$$\text{(Full Stop)} \qquad f^1_{\boldsymbol{\eta}} + f^2_{\boldsymbol{\eta}} + f^3_{\boldsymbol{\eta}} + Y_{\boldsymbol{\eta}} \leq 1, \quad \forall \boldsymbol{\eta} \in \mathcal{N} \setminus \mathcal{B} \tag{3.8}$$

$$\text{(Non-negativity)} \qquad f_{\boldsymbol{\eta}}, f^1_{\boldsymbol{\eta}}, f^2_{\boldsymbol{\eta}}, f^3_{\boldsymbol{\eta}} \geq 0, \quad \forall \boldsymbol{\eta} \in \mathcal{N} \tag{3.9}$$

$$\text{(Binary)} \qquad Y_{\boldsymbol{\eta}} \quad \text{binary} \quad \forall \boldsymbol{\eta} \in \mathcal{N} \tag{3.10}$$

Figure 3.1: MIP Formulation of the Constrained MSP

Figure 3.2. Then

$$\begin{aligned} f^i_{\boldsymbol{\eta}} &= \mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and do not stop at } \boldsymbol{\eta} \text{ and alternative } i \text{ wins}) \\ &= \mathsf{P}(\text{alternative } i \text{ wins} \mid \text{arrive at } \boldsymbol{\eta} \text{ and do not stop}) \cdot \\ &\qquad \mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and do not stop}) \\ &= \mathsf{P}(\text{alternative } i \text{ wins} \mid \text{arrive at } \boldsymbol{\eta} \text{ and do not stop}) \cdot \\ &\qquad \mathsf{P}(\text{do not stop at } \boldsymbol{\eta} \mid \text{arrive at } \boldsymbol{\eta})\ \mathsf{P}(\text{arrive at } \boldsymbol{\eta}) \\ &= p_i\ (1 - p_{\boldsymbol{\eta}})\ \mathsf{P}(\text{arrive at } \boldsymbol{\eta}). \end{aligned} \tag{3.12}$$

Since $\sum_{i=1}^{k} p_i = 1$,

$$\begin{aligned} f^1_{\boldsymbol{\eta}} + f^2_{\boldsymbol{\eta}} + f^3_{\boldsymbol{\eta}} &= (1 - p_{\boldsymbol{\eta}})\mathsf{P}(\text{arrive at } \boldsymbol{\eta}), \\ &= \mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and do not stop}). \end{aligned}$$

- $Y_{\boldsymbol{\eta}}$'s are binary variables that denote whether or not node $\boldsymbol{\eta}$ is a stopping node and are defined as

$$Y_{\boldsymbol{\eta}} \equiv I_{\left[\boldsymbol{\eta} \text{ is a stopping node}\right]},$$

where $I_{[A]}$ is the indicator function for the occurrence of event A. In the LP (randomized) formulation, these variables are omitted.

3.1.3.2 Objective Function

Our goal is to choose a set of stopping points (and stopping probabilities) that minimizes the expected number of observations subject to the P(CS) constraint (3.6). In the formulation, that goal is reflected by the objective function (3.1), which minimizes the sum of the flows across all of the arcs connecting one node to another (i.e., all of the arcs except those to the sink). To see that (3.1) achieves our goal, consider the random variable, N, which represents the total number of observations taken during a single experiment:

$$N = \sum_{\boldsymbol{\eta} \in \mathcal{N}} I_{\left[\text{observation taken at node } \boldsymbol{\eta}\right]}.$$

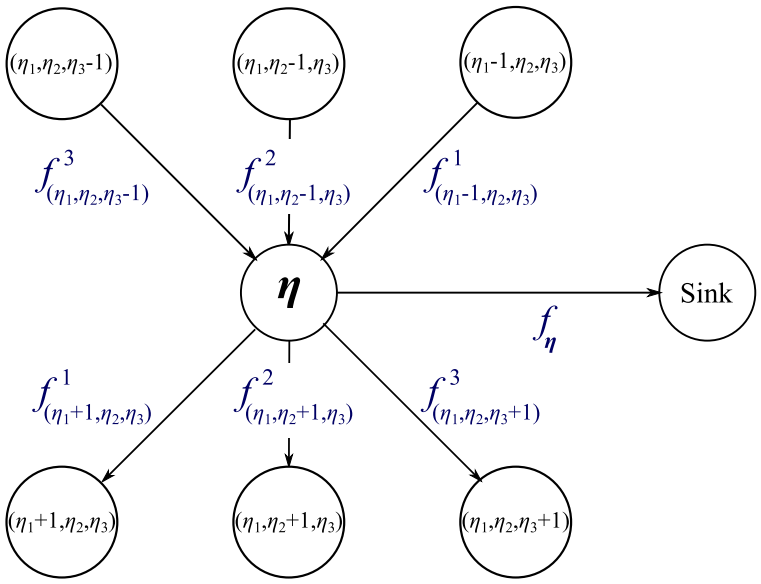

Figure 3.2: Flow Variables

Thus,

$$\begin{aligned}
\mathsf{E}[N] &= \sum_{\boldsymbol{\eta}\in\mathcal{N}} \mathsf{P}(\text{observation taken at node } \boldsymbol{\eta}) \\
&= \sum_{\boldsymbol{\eta}\in\mathcal{N}} \mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and do not stop}) \\
&= \sum_{\boldsymbol{\eta}\in\mathcal{N}} (1-p_{\boldsymbol{\eta}})\, \mathsf{P}(\text{arrive at } \boldsymbol{\eta}) \\
&= \sum_{\boldsymbol{\eta}\in\mathcal{N}} (f^1_{\boldsymbol{\eta}} + f^2_{\boldsymbol{\eta}} + f^3_{\boldsymbol{\eta}}).
\end{aligned}$$

3.1.3.3 Constraints

With the constraints for the case $k = 3$ in mind, we briefly describe the form of the constraints for general k.

- *Initialization*: Constraint (3.2) initializes the total probability in the network by making $\mathbf{0}$ a source node with probability flow 1. Note that the *net* probability flow introduced into the network may be less than 1 if $f_{\mathbf{0}} > 0$, that is, if node $\mathbf{0}$ is also a stopping node.

- *Flow Conservation*: Constraints (3.3), depicted in Figure 3.2, require that the flow of probability into any node equals the flow from that node.

- *Balanced Outflow*: Constraints (3.4) maintain the probability configuration $\boldsymbol{p}$ by ensuring that the success probability for each of the k alternatives is consistent throughout the network. There are $k-1$ such constraints per node, leaving the optimization one degree of freedom to set the probability outflows on the k arcs. We write the constraints in terms of θ_2 and θ_3 for the general configuration $\boldsymbol{p} = (q, \frac{q}{\theta_2}, \frac{q}{\theta_3})$, where $q = (1 + \frac{1}{\theta_2} + \frac{1}{\theta_3})^{-1}$ so that $\sum_{i=1}^{3} p_i = 1$. For the SC, $\boldsymbol{p} = (q, \frac{q}{\theta^\star}, \frac{q}{\theta^\star})$.

- *Symmetry*: Constraints (3.5) require that $p_{\boldsymbol{\eta}}$ — the probability of stopping at any permutation of a particular node $\boldsymbol{\eta}$ given an arrival at that node — is the same for all permutations of the node. The necessity for this set of constraints is discussed in §3.1.1. We note first that if $\eta_1 \geq \eta_2 \geq \eta_3$, then $\boldsymbol{\eta}$ is equal to $\boldsymbol{\eta'}$, its left-lexicographic permutation. Therefore, we require that

$$p_{\boldsymbol{\eta}} = p_{\boldsymbol{\eta}'} \text{ for all } \boldsymbol{\eta} \in \pi(\boldsymbol{\eta}'), \tag{3.13}$$

where $\pi(\boldsymbol{\eta}')$ is the set of all permutations of $\boldsymbol{\eta}' = (\eta_{[3]}, \eta_{[2]}, \eta_{[1]})$.

Note that

$$\mathsf{P}(\text{arrive at } \boldsymbol{\eta}) \;=\; S_{\boldsymbol{\eta}} \prod_{i=1}^{3} {p_i}^{\eta_i} \;=\; S_{\boldsymbol{\eta}}\, q^{\sum_{i=1}^{k} \eta_i} \prod_{i=2}^{3} \frac{1}{{\theta_i}^{\eta_i}}, \tag{3.14}$$

where we leave out $i = 1$ in the product of the last term since $\theta_1 = 1$. By (3.11) and (3.14), we can write (3.13) as

$$\frac{f_{\boldsymbol{\eta}}}{S_{\boldsymbol{\eta}}\, q^{\sum_{i=1}^{k} \eta_i} \prod_{i=2}^{3} \frac{1}{\theta_i^{\eta_i}}} = \frac{f_{\boldsymbol{\eta}'}}{S_{\boldsymbol{\eta}'}\, q^{\sum_{i=1}^{k} \eta_{[i]}} \prod_{i=2}^{3} \frac{1}{\theta_i^{\eta_{[k-1+1]}}}}. \tag{3.15}$$

By symmetry, the number of paths to each of the permutations is the same (i.e., $S_{\boldsymbol{\eta}} = S_{\boldsymbol{\eta}'}$), and $\sum_{i=1}^{3} \eta_i = \sum_{i=1}^{3} \eta_{[i]}$, so (3.15) simplifies to

$$f_{\boldsymbol{\eta}} \;= \prod_{i=2}^{3} {\theta_i}^{\eta_{[k-i+1]} - \eta_i}\; f_{\boldsymbol{\eta}'}. \tag{3.16}$$

Substitution for the appropriate permutation indices, further simplification, and dropping of the order notation leads to the form of the constraints in (3.5). In the SC, θ_2 and θ_3 are replaced by $\theta^\star$.

Remark 3.1 For the $k = 3$ case considered here, there are five constraints governing the six possible permutations of each left-lexicographic node. In general, for each node, we need one less constraint than the number of permutations. That leaves the optimization one degree of freedom for determining the probability of stopping at any permutation of the node. For nodes in which two of the three elements are equal, we have three possible permutations, and thus the five constraints for that node will simplify to two unique constraints. If all three elements are equal, the five symmetry constraints for that node become the trivial constraint $f_{\boldsymbol{\eta}} - f_{\boldsymbol{\eta}} = 0$, which is equivalent to having no symmetry constraints for that node, as expected.

Remark 3.2 For the MIP, we can use the binary variables to force symmetry. In that case, we could use five constraints to represent

$$Y_{\boldsymbol{\eta}} = Y_{\boldsymbol{\eta}'} \text{ for all } \boldsymbol{\eta} \in \pi(\boldsymbol{\eta}'). \tag{3.17}$$

The choice of implementation for the symmetry constraints in the MIP should be based upon user requirements and MIP solver capabilities.

- *Minimum* P(CS): Constraint (3.6) ensures that our achieved P(CS) is at least $P^\star$. It is the only inequality constraint (besides the lower bound constraints) in the LP (randomized) formulation. Nonzero P(CS) is only achieved at stopping nodes for which $\eta_1 \geq \max\{\eta_2, \eta_3\}$. If $\eta_1 = \max\{\eta_2, \eta_3\}$, then the selection of the best alternative is determined by randomization (not to be confused with a randomized stop) to account for any ties. Let $t(\boldsymbol{\eta})$ be the number of alternatives with the same number of successes as the best (including the best). The general equation then is

$$\mathsf{P}(\text{CS}) \;=\; \sum_{\substack{\boldsymbol{\eta}\in\mathcal{N} \\ \eta_1 \geq \max\{\eta_2,\eta_3\}}} \frac{1}{t(\boldsymbol{\eta})} \mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and stop}) = \sum_{\substack{\boldsymbol{\eta}\in\mathcal{N} \\ \eta_1 \geq \max\{\eta_2,\eta_3\}}} \frac{f_{\boldsymbol{\eta}}}{t(\boldsymbol{\eta})}.$$

- *Stop*: Constraints (3.7) only apply to the MIP (non-randomized) formulation. They ensure that there is no stopping probability at a node unless it is a complete stop (i.e., $f_{\boldsymbol{\eta}} = 0$ when $Y_{\boldsymbol{\eta}} = 0$).

- *Full Stop*: Constraints (3.8) only apply to the MIP formulation. They ensure that if a node has been selected as a stopping node (i.e., $Y_{\boldsymbol{\eta}} = 1$), the stop is a full stop; i.e., there is no flow on the node's outbound arcs. This set of constraints, combined with the previous set, ensures that $p_{\boldsymbol{\eta}} = 0$ or 1.

- *Non-negativity*: Constraints (3.9) ensure that all continuous variables are non-negative.

- *Binary*: Constraints (3.10) declare the binary variables for the MIP formulation.

3.1.3.4 Solutions

We now consider the solution (assuming that it exists) to the optimization. A solution will not exist if our budget is insufficient to achieve $P^\star$ for the given problem parameters. In particular, if the budget b is less than the truncation parameter n_{BEM} for Procedure $\mathcal{M}_{\text{BEM}}$, the problem is infeasible. Our goal is to find $\mathcal{S}_{\text{R}}$ or $\mathcal{S}_{\text{NR}}$ that minimizes the expected number of observations. Pursuant to the discussion of our procedure in §3.1.2, we must extract from our solution the $p_{\boldsymbol{\eta}}$ for all $\boldsymbol{\eta}$ as follows:

$$\begin{aligned} p_{\boldsymbol{\eta}} &= \mathsf{P}(\text{stop at } \boldsymbol{\eta} \,|\, \text{arrive at } \boldsymbol{\eta}) \\ &= \frac{\mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and stop})}{\mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and stop}) + \mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and do not stop})} \\ &= \frac{f_{\boldsymbol{\eta}}}{f_{\boldsymbol{\eta}} + f^1_{\boldsymbol{\eta}} + f^2_{\boldsymbol{\eta}} + f^3_{\boldsymbol{\eta}}}. \end{aligned} \tag{3.18}$$

In the case that the numerator and denominator are both zero (e.g., for nodes through which there is no inflow of probability), we arbitrarily set $p_{\boldsymbol{\eta}} = 0$. With the set $\{p_{\boldsymbol{\eta}}\}$ in hand, we can construct the sets required to define our procedures, which leads us to the following theorems.

Theorem 3.1 *A randomized procedure created from the solution to the LP formulation described in §3.1.3 is an optimal solution to the constrained, indifference zone MSP for a specified probability configuration* $\boldsymbol{p}$.

Theorem 3.2 *A non-randomized procedure created from the solution to the MIP formulation described in §3.1.3 is an optimal* non-randomized *solution to the constrained, indifference zone MSP for a specified probability configuration* $\boldsymbol{p}$.

The theorems acknowledge that our procedures satisfy condition (1.2) when $\boldsymbol{p}$ is the SC, but not necessarily condition (1.1). We address that further in §3.2.1.

3.2 Results

In this section, we give preliminary results concerning the performance of our new procedures. §3.2.1 briefly discusses the LFC for our procedure. In §3.2.2, we present examples that compare our new procedures to existing procedures.

3.2.1 Least Favorable Configuration

For all of the MSP procedures that we reviewed in Chapter 2, it is either proved or conjectured that the SC is the LFC. Every known proof is based on a knowledge of the structure of the procedure's stopping points. Unfortunately, beyond symmetry, there is no structure guaranteed by either Procedure $\mathcal{M}_{\mathrm{R}}$ or $\mathcal{M}_{\mathrm{NR}}$. In fact, our anecdotal evidence has demonstrated optimal sets of stopping points with no discernible pattern and which change completely in structure given small changes to $P^{\star}$. As a result, we have been unable to apply similar tools to prove that the SC, or any other configuration, is the LFC for our procedures. Furthermore, by Definition 1.1, the LFC is specific to the method of choosing an alternative. Since the optimal stopping structures for Procedures $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{NR}}$ vary so significantly, we must consider the possibility that the LFC for our procedures is unique to a set of optimal stopping points, not Procedures $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{NR}}$ overall, as it is for, say, Procedure $\mathcal{M}_{\mathrm{BEM}}$. Nonetheless, we make the following conjecture.

Conjecture 3.1 *The SC is the LFC for Procedures* $\mathcal{M}_{\mathrm{R}}$ *and* $\mathcal{M}_{\mathrm{NR}}$.

To substantiate our conjecture, we conducted Monte Carlo (MC) sampling in which we randomly and uniformly drew 100,000 probability vectors $\boldsymbol{p}$ from the preference zone, Ω_{PZ}. For each of the sampled probability vectors, we calculated $\mathsf{P}_{\boldsymbol{p}}(\mathrm{CS})$ for the stopping points and compared it to $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$. We conducted these MC sampling experiments for over 60 different sets of optimal stopping points for both Procedures $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{NR}}$. In every case, $\mathsf{P}_{\boldsymbol{p}}(\mathrm{CS}) \geq \mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$. Note that the randomness of our experiments was in the drawing of

Table 3.1: Comparative Results for $k = 2$, $\theta^\star = 1.6$, $P^\star = 0.9$, $b = 41$

Procedure	Parameters	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	% Increase
$\mathcal{M}_{\mathrm{R}}$	Stops = 22	16.865	0.9000	
$\mathcal{M}_{\mathrm{NR}}$	Stops = 21	16.873	0.9000	0.05
$\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 41$	17.001	0.9006	0.81
$\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 41, r = 5, t = 21$	17.001	0.9006	0.81
$\mathcal{M}_{\mathrm{RA}}$	$r = 5, t = 21$	17.001	0.9006	0.81
$\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 31$	25.505	0.9054	51.23
$\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 31$	31.000	0.9054	83.81

the $\boldsymbol{p}$, not in the calculation of $\mathsf{P}_{\boldsymbol{p}}(\mathrm{CS})$ and $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$, which are exact numerical results for the given probability configurations.

3.2.2 Procedure Performance

We provide examples to compare our new procedures to the previously existing procedures introduced in Chapter 2. For our examples, we set a budget constraint equal to the optimal truncation procedure parameter n_{BG} for Procedure $\mathcal{M}_{\mathrm{BG}}$ — an established, robust procedure, which typically performs at least as well as the other previously existing procedures. (We justify the statement about the superiority of Procedure $\mathcal{M}_{\mathrm{BG}}$ in Chapter 6.) By setting $b = n_{\mathrm{BG}}$, we force our new procedures to perform under conditions very favorable to the best of the existing procedures. All existing procedures are required to have a maximum possible number of observations less than or equal to b, which may affect the feasible parameter space for a particular procedure.

For a first, illustrative example, suppose $k = 2$, $\theta^\star = 1.6$, $P^\star = 0.9$, and $b = 41$. Table 3.1 displays the expected number of observations and the achieved $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ for each procedure. For each existing procedure, the second column includes the procedure parameter settings that minimize $\mathsf{E}_{\mathrm{SC}}[N]$ for the procedure while still achieving $P^\star$. For our new procedures, the second column identifies the total number of left-lexicographic stopping points. The last column shows the percent increase in $\mathsf{E}_{\mathrm{SC}}[N]$ over the optimal, randomized procedure (Procedure $\mathcal{M}_{\mathrm{R}}$).

For our second example, we take $k = 3$, $\theta^\star = 2$, $P^\star = 0.9$, and $b = 34$. Table 3.2

Table 3.2: Comparative Results for $k = 3$, $\theta^\star = 2$, $P^\star = 0.9$, $b = 34$

Procedure	Parameters	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	% Increase
$\mathcal{M}_{\mathrm{R}}$	Stops = 114	16.857	0.9000	
$\mathcal{M}_{\mathrm{NR}}$	Stops = 112	16.859	0.9000	0.01
$\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 34$	17.165	0.9016	1.83
$\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 30, r = 5, t = 12$	18.749	0.9001	11.22
$\mathcal{M}_{\mathrm{RA}}$	$r = 5, t = 12$	18.940	0.9057	12.35
$\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 29$	24.242	0.9044	43.80
$\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 29$	29.000	0.9044	72.03

shows the resulting performance of the procedures. Here, the new procedures significantly outperform all but Procedure $\mathcal{M}_{\mathrm{BG}}$ in terms of $\mathsf{E}_{\mathrm{SC}}[N]$. Note also the large number of left-lexicographic stopping points for Procedures $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{NR}}$, even for this relatively modest-sized problem, suggesting that the new procedures need to be automated in implementation.

Neither of the examples shows a significant improvement over Procedure $\mathcal{M}_{\mathrm{BG}}$; this is a result of our choice of b. Figure 3.3 is a plot of $\mathsf{E}_{\mathrm{SC}}[N]$ as a function of b. The values for k, $\theta^\star$, and $P^\star$ remain the same as in our examples; the chart on the left [right] corresponds to the first [second] example. We first note that for Procedures $\mathcal{M}_{\mathrm{BEM}}$, $\mathcal{M}_{\mathrm{BK}}$, and $\mathcal{M}_{\mathrm{BG}}$, the procedure parameter choices which minimize $\mathsf{E}_{\mathrm{SC}}[N]$ while still achieving $P^\star$ do not change as the budget is increased. For the remaining procedures, an increased budget provides additional flexibility which may result in more efficient procedure parameter settings. For the $k = 2$ example, when $b = 60$, use of Procedure $\mathcal{M}_{\mathrm{BG}}$ results in a 3.2% increase in the expected number of observations over our new procedures. For the $k = 3$ example, when $b = 50$, Procedure $\mathcal{M}_{\mathrm{BG}}$ suffers a 7.0% increase. Thus, the performance of our new procedures versus Procedure $\mathcal{M}_{\mathrm{BG}}$ improves as b increases beyond n_{BG}.

The relative performance between procedures is also dependent upon the particular $P^\star$ of interest. In both of our examples, $P^\star = 0.9$. Figure 3.4 plots $\mathsf{E}_{\mathrm{SC}}[N]$ as a function of $P^\star$ for the original fixed b. Procedure performance is determined for a vector of $P^\star$-values from $1/k$ to 0.99 at 0.01 increments, resulting in the step functions depicted in the figure. In reality, the function for Procedure $\mathcal{M}_{\mathrm{R}}$ is a piecewise linear convex function, but is a step function in the figure because of the increments at which we evaluated the LP. Indeed,

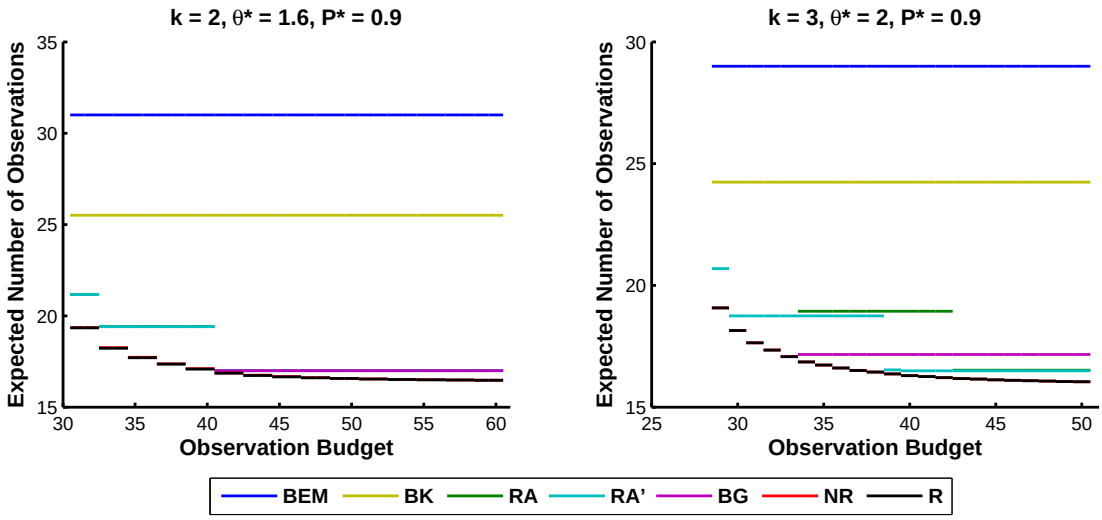

Figure 3.3: Procedure Performance as a Function of the Observation Budget b

in both plots, the improvement in the expected number of observations obtained by our new procedures varies by $P^{\star}$. We also note that Procedure $\mathcal{M}_{\mathrm{R}}$ yields only a very small improvement over Procedure $\mathcal{M}_{\mathrm{NR}}$ in our examples. In fact, in both plots, the separation between the two new procedures' expected number of observations is not even visible over a large portion of their domains. Our empirical results indicate that for larger b, the performance of Procedure $\mathcal{M}_{\mathrm{NR}}$ (in terms of the expected number of observations) is very close to that of Procedure $\mathcal{M}_{\mathrm{R}}$. We explore this further in Chapter 6.

3.3 Summary

We have developed optimal procedures for selection-of-the-best indifference zone MSPs under an observation budget constraint. Our research leverages the field of mathematical programming by modeling the characteristics of the problems as an LP and as an MIP. By construction, our procedures always perform at least as well as existing MSP procedures. Additionally, the optimality properties of our resulting procedures provide a standard against which other MSP procedures can be evaluated in terms of the expected number of observations.

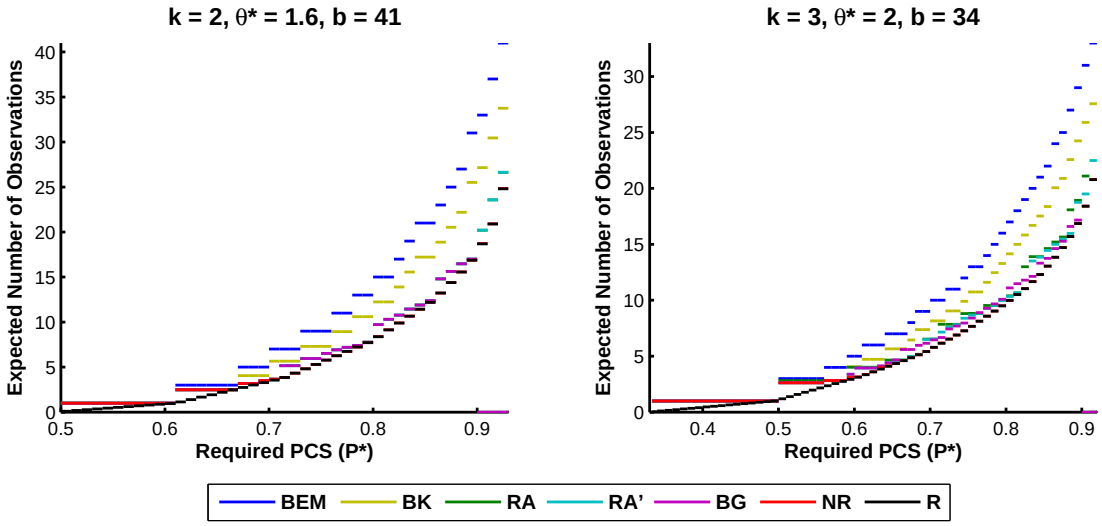

Figure 3.4: Procedure Performance as a Function of $P^\star$

We have also introduced a new type of procedure — the randomized procedure. Randomized procedures provide the same guarantees as non-randomized procedures, but can be more efficient when optimized for the problem of interest. When the budget is large, the performances of the randomized and non-randomized procedures, in terms of expected number of observations, are nearly identical. However, the LP formulations are much easier to solve in implementation than MIP formulations. Furthermore, the concept of randomized stopping points has application beyond MSPs to other types of ranking and selection procedures.

CHAPTER IV

EFFICIENT IMPLEMENTATION

The primary contributions of the current chapter are threefold. In §4.1, we develop efficient reformulations of the LP and MIP upon which the optimal procedures are based. These new formulations expand the range of real world problems that we can address using the new procedures. We then examine some key characteristics of the optimal randomized and non-randomized procedures in §4.2. In §4.3, we extend the LP formulation in order to replicate the classical procedures. We give conclusions in §4.4.

4.1 Efficient Reformulation of the Mathematical Programs

In this section, we reformulate our mathematical programs in order to deal with some daunting computational issues related to problem size and numerical instability concerns. The reformulated problem provides significant computational savings and mitigates many of our numerical instability issues. In §4.1.1, we describe the main implementation problems, and then provide an overview of our approach to mitigate those in §4.1.2. We then derive the reformulation in §4.1.3. Finally, in §4.1.4, we discuss the effects of the reformulation.

4.1.1 Motivation

Our original formulation has a number of significant implementation issues.

Problem Growth The number of nodes in the network increases 'exponentially' in the number of alternatives, k, and the observation budget, b. For example, the randomized formulation of the MSP with $k = 4$ and $b = 50$ has 1.6 million variables (four variables for each of the 316,250 nodes) and 1.5 million constraints (not including non-negativity constraints). The non-randomized formulation for the MIP adds an additional 316,250 binary variables and nearly twice as many additional constraints. It is so large, in fact, that we were unable to solve problems of this size. Thus, as configured, the mathematical

program can handle only a tiny fraction of the problems of interest.

Numerical Instability The large number of equality constraints leads to stability issues, particularly for large non-randomized problems. When running those MIPs, our software had difficulty finding feasible solutions for problems for which we knew such solutions existed. Even when we provided the software a feasible starting point, we would often get that point back immediately as the only feasible (and thus optimal) solution. While we cannot be certain of the cause of these issues, we suspect that minor numerical precision errors can lead to one or more of the equality constraints being violated, resulting in otherwise feasible solutions being considered infeasible (or vice versa). This is especially problematic in larger problems, since the probability flowing through the nodes later in the network, after many observations have been taken, can be orders of magnitude lower than the flows through the nodes at the beginning of the network.

In the next section, we discuss our approach to mitigating these significant implementation problems.

4.1.2 Approach

Our primary goal is to represent the original network using a much smaller subset of the nodes and arcs, without losing any required information. Fortunately, two sets of equality constraints and the concept of curtailment provide us the means to do so. A smaller network will directly address the issue of problem growth, and the elimination of a large portion of the equality constraints will remove much of the numerical instability.

Remark 4.1 A cumulative success vector $\boldsymbol{\eta}_m = (\eta_1, \eta_2, \ldots, \eta_k)$ represents a point on the particular sample path $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_m, \boldsymbol{\eta}_{m+1}, \ldots\}$ — a sequence of random observations — along which an experiment proceeds. There are usually many potential sample paths that include $\boldsymbol{\eta}_m$ as a point. If $\boldsymbol{\eta}_m$ is a stopping point, then the sample path (and the experiment) is terminated. Viewing the cumulative success vectors in this way is helpful for understanding the underlying probabilities in subsequent discussions.

4.1.2.1 Arc Reduction

The balance outflow constraints, (3.4), require that the flows along arcs out of each node (i.e., the respective probabilities of alternative success) maintain the same proportion (θ_i's) throughout the network. Setting the flow on one outbound arc completely determines the flows on the remaining outbound arcs for that node. As a direct result of (3.4), we can represent the flows along each arc from a particular node in terms of any one of its outbound arcs, say $f^1_{\boldsymbol{\eta}}$, in the following manner:

$$f^i_{\boldsymbol{\eta}} = \frac{1}{\theta_i} f^1_{\boldsymbol{\eta}}, \tag{4.1}$$

where $\theta_1 = 1$. All of the flow out of node $\boldsymbol{\eta}$ is then:

$$\begin{aligned} \sum_{i=1}^{k} f^i_{\boldsymbol{\eta}} &= \sum_{i=1}^{k} \frac{1}{\theta_i} f^1_{\boldsymbol{\eta}} \\ &= \Theta f^1_{\boldsymbol{\eta}}, \end{aligned} \tag{4.2}$$

where $\Theta \equiv \sum_{i=1}^{k} \frac{1}{\theta_i}$.

4.1.2.2 Node Reduction via Symmetry

As discussed in Chapter 3 for $k = 3$, if $(\eta_1, \eta_2, \ldots, \eta_k)$ is a stopping point, then so is any permutation of the set $\{\eta_1, \eta_2, \ldots, \eta_k\}$. That symmetry requires that any decision to stop must be the same for all permutations of a potential stopping vector as observed by the experimenter. The symmetry constraints, (3.5), accomplish this by requiring that the conditional probability of stopping at a particular node, given an arrival at that node, is the same for all of its permutations. But that implies that the flows to the sink from all of the permutations of a particular node should be representable in terms of the flow from any one of its permutations; we choose the left-lexicographic permutation $\boldsymbol{\eta}' = (\eta_{[k]}, \eta_{[k-1]}, \ldots, \eta_{[1]})$. Let $\mathcal{N}' \equiv \{\boldsymbol{\eta}' : \sum_{i=1}^{k} \eta_{[i]} \leq b\}$.

In developing the symmetry constraints in Chapter 3, we derive (3.16), which we generalize here for arbitrary k:

$$f_{\boldsymbol{\eta}} = \prod_{i=2}^{k} \theta_i^{\eta_{[k-i+1]} - \eta_i} \; f_{\boldsymbol{\eta}'}. \tag{4.3}$$

We now have a relationship between the flows to the sink from a node and its left-lexicographic permutation. We claim that we can apply this same formula to the flows along the arcs between nodes as well, to arrive at the following relationship:

$$f_{\boldsymbol{\eta}}^i = \prod_{i=2}^{k} \theta_i^{\eta_{[k-i+1]}-\eta_i} \; f_{\boldsymbol{\eta}'}^i. \tag{4.4}$$

The derivation follows. Let $S_{\boldsymbol{\eta}}$ be the number of sample paths into node $\boldsymbol{\eta}$, and let the general probability configuration be $\boldsymbol{p} = (q, \frac{q}{\theta_2}, \ldots, \frac{q}{\theta_k})$. Then, using (3.12),

$$\begin{aligned}
f_{\boldsymbol{\eta}}^i &= p_i \; (1 - p_{\boldsymbol{\eta}}) \; \mathsf{P}(\text{arrive at } \boldsymbol{\eta}) \\
&= p_i \; (1 - p_{\boldsymbol{\eta}}) \; S_{\boldsymbol{\eta}} \; \prod_{i=1}^{k} p_i^{\eta_i} \\
&= p_i \; (1 - p_{\boldsymbol{\eta}}) \; S_{\boldsymbol{\eta}} \; q^{\sum_{i=1}^k \eta_i} \prod_{i=1}^{k} \left(\frac{1}{\theta_i}\right)^{\eta_i} \\
&= p_i \; (1 - p_{\boldsymbol{\eta}'}) \; S_{\boldsymbol{\eta}'} \; q^{\sum_{i=1}^k \eta_{[i]}} \prod_{i=2}^{k} \left(\frac{1}{\theta_i}\right)^{\eta_i},
\end{aligned}$$

since $p_{\boldsymbol{\eta}} = p_{\boldsymbol{\eta}'}$ by design, $S_{\boldsymbol{\eta}} = S_{\boldsymbol{\eta}'}$ by symmetry, and $\sum_{i=1}^k \eta_i = \sum_{i=1}^k \eta_{[i]}$ since the terms are permutations of the same vector. Then

$$\begin{aligned}
f_{\boldsymbol{\eta}}^i &= p_i \; (1 - p_{\boldsymbol{\eta}'}) \; S_{\boldsymbol{\eta}'} \; q^{\sum_{i=1}^k \eta_{[i]}} \prod_{i=2}^{k} \left(\frac{1}{\theta_i}\right)^{\eta_i} \left(\frac{1}{\theta_i}\right)^{\eta_{[k-i+1]}-\eta_{[k-i+1]}} \\
&= \prod_{i=2}^{k} \left(\frac{1}{\theta_i}\right)^{\eta_i-\eta_{[k-i+1]}} \; p_i \; (1 - p_{\boldsymbol{\eta}'}) \; S_{\boldsymbol{\eta}'} \; q^{\sum_{i=1}^k \eta_{[k-i+1]}} \prod_{i=2}^{k} \left(\frac{1}{\theta_i}\right)^{\eta_{[k-i+1]}} \\
&= \prod_{i=2}^{k} \theta_i^{\eta_{[k-i+1]}-\eta_i} \; p_i \; (1 - p_{\boldsymbol{\eta}'}) \; \mathsf{P}(\text{arrive at } \boldsymbol{\eta}) \\
&= \prod_{i=2}^{k} \theta_i^{\eta_{[k-i+1]}-\eta_i} f_{\boldsymbol{\eta}'}^i. \quad \square
\end{aligned}$$

The relationships in (4.3) and (4.4) allow us to reduce our network to one consisting only of left-lexicographic nodes and their arcs.

Remark 4.2 In our current approach, we are still using the network model of Chapter 3, but restricting our attention to the subnetwork of left-lexicographic nodes, from which all of the information we need can be inferred. Alternatively, we could have modeled the problem by starting with the set of *information* states (i.e., the left-lexicographic nodes) and deriving

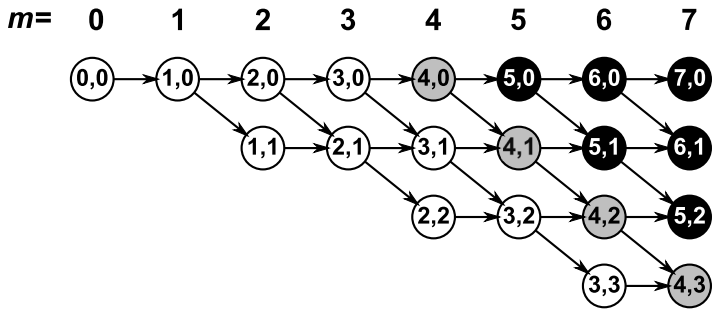

Figure 4.1: Curtailment Example with $k = 3$ and $b = 7$

the objective function and constraints directly from those. We include an example of how to use the information state model to derive a new set of flow conservation constraints in Appendix C.2.

4.1.2.3 Node Reduction via Curtailment

The concept of strong curtailment, introduced by BK (1984), provides us another means of reducing the size of our network. As already discussed in Chapter 2, strong curtailment allows us to stop when the best any alternative can do with respect to the alternative with the most successes is tie, without affecting the P(CS) of the procedure.

For strong curtailment, we stop at stage m if

$$\eta_{[k]m} - \eta_{[k-1]m} \geq b - m.$$

Stopping at a node due to curtailment may terminate some or all of the sample paths into nodes further into the network. Figure 4.1 shows an example with $k = 2$ and $b = 7$. Each circle represents a node in the network and each arrow represents an arc connecting two nodes. Note that the network includes only left-lexicographic nodes. For clarity, we have removed the arcs to the sink as well as the parentheses in the node vector notation. In this example, the gray nodes are the nodes that are stops under strong curtailment and the black nodes are nodes that cannot be subsequently reached. The findings of BK (1984) allow us to eliminate curtailed nodes (i.e., the black nodes in Figure 4.1) without affecting P(CS).

Algorithmically, we eliminate a left-lexicographic node $(\eta_{[k]m}, \eta_{[k-1]m}, \ldots, \eta_{[1]m})$ from

the network if

$$(\eta_{[k]m} - 1) - \eta_{[k-1]m} \geq b - (m-1), \quad \text{or}$$
$$\eta_{[k]m} - \eta_{[k-1]m} - 1 \geq b - m + 1. \tag{4.5}$$

In other words, a node is culled if all of its possible *origination nodes* — a term we use for nodes from which arcs come into the node under discussion — would have met the conditions of a stopping node under strong curtailment. For example, in Figure 4.1, $(5, 1)$ is eliminated since both of its origination nodes, $(5, 0)$ and $(4, 1)$, meet the strong curtailment stopping conditions. On the other hand, $(4, 1)$ is not eliminated, since one of its origination nodes, $(3, 1)$, does not meet the curtailment stopping conditions. We only need to check the origination node least likely to have met the strong curtailment conditions at $m - 1$ observations, which is the node with the first component equal to $\eta_{[k]m} - 1$, since it has the lowest possible value for $\eta_{[k]m} - \eta_{[k-1]m}$.

We have to be careful in one case — when $\eta_{[k]m}$ and $\eta_{[k-1]m}$ are tied. Then, (4.5) has a –1 on the left hand side, which can never result in an elimination by our algorithm. However, the previous difference between the best two must have been either 0 or 1. If $m = b$ (and $\eta_{[k-2]b} = 0$ when $k > 2$), then that difference must have been 1. In that case, sampling would have stopped at $b - 1$, since there would have been only one observation left to be taken. As a result, we also eliminate the one node for which

$$\eta_{[k]b} = \eta_{[k-1]b} \text{ and } \eta_{[k-2]b} = 0 \text{ when } k > 2, \text{ or}$$
$$\eta_{[k]b} = \eta_{[k-1]b} = \frac{b}{2} \text{ when } k = 2. \tag{4.6}$$

This can only occur when b is even. For example, if $k = 3$ and $b = 8$, we eliminate node $(4, 4, 0)$. Note that if $k = 2$ and b is even, we will never take an additional observation at $m = b - 1$. Thus, under curtailment, a procedure constrained by an even budget b is equivalent to the same procedure constrained by a budget of $b - 1$.

We will use $\mathcal{N}'' \subset \mathcal{N}'$ to represent the set of nodes, $\boldsymbol{\eta}''$, that remain after curtailed nodes are eliminated (the white and gray nodes in Figure 4.1), and we define $\mathcal{B}'$ to be those nodes at the end of the directed network, but not necessarily at the budget, b (the gray nodes in Figure 4.1).

4.1.3 Reformulating the Mathematical Program

In §4.1.2, we derived the relationships and tools necessary to reduce the original network to a much smaller one composed of only left-lexicographic, non-curtailed nodes. They also allow us to use a single variable per node to represent the outbound flow to other nodes.

Remark 4.3 There is an important distinction to make here. The symmetry constraints allow us to consider only left-lexicographic nodes, and the arcs from those nodes to other left-lexicographic nodes and to the sink. On the other hand, the balance outflow constraints provide us the means to represent all flows from a node via a single variable, but we still must consider and account for all of the arcs. Without this understanding, the network is no longer intuitive, since flows from most nodes must reach more than one left-lexicographic node.

In the following, we describe how we apply the relationships and tools described above to the existing mathematical programs to develop the final, efficient formulation.

Objective Function Transformation of the objective function involves two steps. First, we represent the sum of the arc flows in terms of the flows associated with successes by alternative 1. Starting with the original objective function, but adapted for general k, we have

$$\min_{f_{\boldsymbol{\eta}}, f^1_{\boldsymbol{\eta}}, \ldots, f^k_{\boldsymbol{\eta}}, Y_{\boldsymbol{\eta}}} \sum_{\boldsymbol{\eta} \in \mathcal{N}} \sum_{i=1}^{k} f^i_{\boldsymbol{\eta}}. \tag{4.7}$$

We use (4.2) to write (4.7) in terms of the alternative 1 success arcs:

$$\min_{f_{\boldsymbol{\eta}}, f^1_{\boldsymbol{\eta}}, Y_{\boldsymbol{\eta}}} \Theta \sum_{\boldsymbol{\eta} \in \mathcal{N}} f^1_{\boldsymbol{\eta}}.$$

Using (4.4), we then eliminate all of the nodes that are not left-lexicographic and consider only those nodes not eliminated by curtailment as follows:

$$\min_{f_{\boldsymbol{\eta}''}, f^1_{\boldsymbol{\eta}''}, Y_{\boldsymbol{\eta}''}} \Theta \sum_{\boldsymbol{\eta}'' \in \mathcal{N}''} \sum_{\boldsymbol{\eta} \in \pi(\boldsymbol{\eta}'')} \prod_{i=2}^{k} {\theta_i}^{\eta_{[k-i+1]} - \eta_i} f^1_{\boldsymbol{\eta}''}, \tag{4.8}$$

where $\pi(\boldsymbol{\eta}'')$ is the set of all permutations of $\boldsymbol{\eta}''$ within which $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)$ is contained. Our new objective function is in terms of $f^1_{\boldsymbol{\eta}''}$ alone for all $\boldsymbol{\eta}'' \in \mathcal{N}''$.

Initialization Constraint Reformulating the initialization constraint is straightforward. We simply use (4.2) to write the constraint in terms of the alternative 1 success arc. The new constraint is

$$f_{\mathbf{0}} + \Theta f^1_{\mathbf{0}} = 1. \tag{4.9}$$

Flow Conservation Constraints We require one flow conservation constraint per node (except the first) represented in the reformulated network, i.e., for all $\boldsymbol{\eta}'' \in \mathcal{N}'' \setminus \mathbf{0}$. Flow conservation for the first node, $\mathbf{0}$, is guaranteed by the initialization constraint. Rewriting the 'flow out' portion of the constraint in terms of $f_{\boldsymbol{\eta}''}$ and $f^1_{\boldsymbol{\eta}''}$ is similar to our reformulation of the initialization constraint.

The 'flow in' portion depends upon the origination nodal arcs, some of which may not even come from left-lexicographic nodes. For example with $k = 3$, the node $(4, 2, 2)$ has origination arcs from $(3, 2, 2)$, $(4, 1, 2)$, and $(4, 2, 1)$; the second origination node is not left-lexicographic. Let $\boldsymbol{\eta}^\nu = (\eta^\nu_1, \eta^\nu_2, \ldots, \eta^\nu_k)$ be the origination node for arc ν ending at component ν of vector $\boldsymbol{\eta}''$. Following our example with $\boldsymbol{\eta} = (4, 2, 2)$, $\boldsymbol{\eta}^1 = (3, 2, 2)$, $\boldsymbol{\eta}^2 = (4, 1, 2)$, and $\boldsymbol{\eta}^3 = (4, 2, 1)$. Note that if one of the alternatives has zero successes, then $\boldsymbol{\eta}^\nu$ does not exist for that component; by convention, we set $f^1_{\boldsymbol{\eta}^\nu} \equiv 0$ in that case. Let $L(\boldsymbol{\eta}^\nu)$ be the left-lexicographic permutation of $\boldsymbol{\eta}^\nu$, or $L(\boldsymbol{\eta}^\nu) = (\eta^\nu_{[k]}, \eta^\nu_{[k-1]}, \ldots, \eta^\nu_{[1]})$.

For each of the origination nodal arcs, we must determine what the origination node is, convert its flow, $f^i_{\boldsymbol{\eta}^\nu}$, to the flow along its left-lexicographic permutation, $f^\nu_{L(\boldsymbol{\eta}^\nu)}$, (if necessary) using (4.4), and then convert that to $f^1_{L(\boldsymbol{\eta}^\nu)}$ (if necessary) using (4.1). In general, the reformulated flow conservation constraints can be written

$$\sum_{\nu=1}^{k} \frac{1}{\theta_\nu} \prod_{i=2}^{k} \theta_i^{\eta^\nu_{[k-i+1]} - \eta^\nu_1} f^1_{L(\boldsymbol{\eta}^\nu)} = f_{\boldsymbol{\eta}''} + \Theta f^1_{\boldsymbol{\eta}''}, \quad \forall \boldsymbol{\eta}'' \in \mathcal{N}'' \setminus \mathbf{0}. \tag{4.10}$$

From an implementation perspective, it would desirable to have an algorithm for efficiently identifying the origination nodes and calculating the coefficients of their arc variables in (4.10). We develop such an algorithm in Appendix C.1.

P(CS) Constraint Let $t(\boldsymbol{\eta}'')$ represent the number of alternatives in node $\boldsymbol{\eta}''$ that have the same number of successes as the first alternative, including the first. Then $1 \leq t(\boldsymbol{\eta}'') \leq k$

with $t(\boldsymbol{\eta}'') = 1$ when none are tied with the first. Additionally, let $\pi^\star(\boldsymbol{\eta}'')$ be the subset of the permutations of $\boldsymbol{\eta}''$ for which the first alternative is greater than or equal to the remaining alternatives. In other words, $\pi^\star(\boldsymbol{\eta}'') \subseteq \pi(\boldsymbol{\eta}'')$ is the subset of permutations that result in correct selections. We can rewrite the $\mathsf{P}(\mathrm{CS})$ constraint, (3.6), in terms of our new network:

$$\sum_{\boldsymbol{\eta}'' \in \mathcal{N}''} \frac{1}{t(\boldsymbol{\eta}'')} \sum_{\boldsymbol{\eta} \in \pi^\star(\boldsymbol{\eta}'')} \prod_{i=2}^{k} \theta_i^{\eta_{[k-i+1]} - \eta_i} f_{\boldsymbol{\eta}''} \;\geq\; P^\star, \tag{4.11}$$

The final representation of the $\mathsf{P}(\mathrm{CS})$ constraint is then in terms of $f_{\boldsymbol{\eta}''}$ alone for all $\boldsymbol{\eta}'' \in \mathcal{N}''$.

4.1.4 Final Reformulation

We combine the transformed objective function and constraints described in §4.1.3 to arrive at our reformulated mathematical program. Note that the integration of the information contained in the original equality constraint sets (3.4) and (3.5) allows us to eliminate those from the reformulation. For brevity, we did not explicitly describe the reformulation of the binary and non-negativity constraints, since their derivation is straightforward. The final reformulated mathematical program is shown in Figure 4.2.

Table 4.1 shows some key problem characteristics for selected values of k and b, where "K" represents thousands and "M" millions. The columns labeled "Old" are the characteristics for the original LP; those marked "New" are the corresponding characteristics for the reformulated LP. Note that the count for the number of constraints does not include the lower bound constraints on the variables. Returning to our original example with $k = 4$ and $b = 50$, we have reduced the number of variables from 1.6 million to 18 thousand — a 98.9% reduction; the number of constraints from 1.5 million to nine thousand — a 99.4% reduction; and the number of nodes from 316 thousand to nine thousand — a 97.1% reduction. This LP, which previously would not even run on our computing resources, now executes in a few seconds. Not only does our reformulation greatly expand the set of solvable problems, but the elimination of many of the equality constraints removes a significant portion of the original source of numerical instability.

$$\min_{f_{\boldsymbol{\eta}''}, f^1_{\boldsymbol{\eta}''}, Y_{\boldsymbol{\eta}''}} \quad \Theta \sum_{\boldsymbol{\eta}'' \in \mathcal{N}''} \sum_{\boldsymbol{\eta} \in \pi(\boldsymbol{\eta}'')} \prod_{i=2}^{k} \theta_i^{\eta_{[k-i+1]} - \eta_i} f^1_{\boldsymbol{\eta}''}$$

Subject to:

$$\text{(Initialization)} \qquad f_{\mathbf{0}} + \Theta f^1_{\mathbf{0}} = 1,$$

$$\text{(Flow Conservation)} \qquad \sum_{\nu=1}^{k} \frac{1}{\theta_\nu} \prod_{i=2}^{k} \theta_i^{\eta^\nu_{[k-i+1]} - \eta^\nu_1} f^1_{L(\boldsymbol{\eta}^\nu)} - f_{\boldsymbol{\eta}''} - \Theta f^1_{\boldsymbol{\eta}''} = 0, \qquad \forall \boldsymbol{\eta}'' \in \mathcal{N}'' \setminus \mathbf{0}$$

$$\text{(Minimum P(CS))} \qquad \sum_{\boldsymbol{\eta}'' \in \mathcal{N}''} \frac{1}{t(\boldsymbol{\eta}'')} \sum_{\boldsymbol{\eta} \in \pi^\star(\boldsymbol{\eta}'')} \prod_{i=2}^{k} \theta_i^{\eta_{[k-i+1]} - \eta_i} f_{\boldsymbol{\eta}''} \geq P^\star,$$

$$\text{(Stop)} \qquad f_{\boldsymbol{\eta}''} - Y_{\boldsymbol{\eta}''} \leq 0, \qquad \forall \boldsymbol{\eta}'' \in \mathcal{N}''$$

$$\text{(Full Stop)} \qquad \Theta f^1_{\boldsymbol{\eta}''} + Y_{\boldsymbol{\eta}''} \leq 1, \qquad \forall \boldsymbol{\eta}'' \in \mathcal{N}'' \setminus \mathcal{B}'$$

$$\text{(Non-negativity)} \qquad f_{\boldsymbol{\eta}''}, f^1_{\boldsymbol{\eta}''} \geq 0, \qquad \forall \boldsymbol{\eta}'' \in \mathcal{N}''$$

$$\text{(Binary)} \qquad Y_{\boldsymbol{\eta}''} \quad \text{binary} \qquad \forall \boldsymbol{\eta}'' \in \mathcal{N}''$$

Figure 4.2: MIP Reformulation of the Constrained MSP

Table 4.1: Comparison of Original and Reformulated LPs

b	Characteristic	$k=2$		$k=3$		$k=4$		$k=5$	
		Old	New	Old	New	Old	New	Old	New
5	Nodes	20	9	55	12	125	14	251	15
	Variables	63	18	224	24	630	28	1.5K	30
	Constraints	47	10	168	13	446	15	991	16
10	Nodes	65	21	285	43	1.0K	62	3.0K	77
	Variables	198	42	1.1K	86	5.0K	124	18K	154
	Constraints	153	22	947	44	4.1K	63	14K	78
25	Nodes	350	104	3.3K	370	24K	882	143K	1.6K
	Variables	1.1K	208	13K	740	119K	1.8K	855K	3.2K
	Constraints	847	105	12K	371	107K	883	757K	1.6K
50	Nodes	1.3K	351	23K	2.3K	316K	9.0K	3.5M	25K
	Variables	4.0K	702	94K	4.6K	1.6M	18K	21M	51K
	Constraints	3.3K	352	87K	2.3K	1.5M	9.0K	20M	25K
100	Nodes	5.2K	1.3K	177K	16K	4.6M	113K	97M	543K
	Variables	15K	2.7K	707K	32K	23M	227K	579M	1.1M
	Constraints	13K	1.3K	666K	16K	22M	113K	464M	543K

4.2 Characteristics of the New Procedures

In this section, we discuss a few very important characteristics of Procedures $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{NR}}$. In §4.2.1 and §4.2.2, we derive two key properties of Procedure $\mathcal{M}_{\mathrm{R}}$ concerning the number of randomized stopping points and achieved $\mathsf{P}(\mathrm{CS})$, respectively.

4.2.1 Number of Randomized Stops in Procedure $\mathcal{M}_{\mathrm{R}}$

We begin by representing the reformulated LP in standard form, which consists only of equality and non-negativity constraints. Let m represent the number of nodes in the curtailed, left-lexicographic network, i.e., $m = |\mathcal{N}''|$. Let $\mathbf{A}$ denote the equality constraint matrix, with each column representing a variable and each row representing the left hand side of a constraint. The first row is the initialization constraint and the next $m-1$ rows are the flow conservation constraints. The last row is the $\mathsf{P}(\mathrm{CS})$ constraint; however, since it is a greater-than-or-equal-to constraint, we subtract a slack variable, f_s, from the left hand side to make it an equality constraint. We must add a column of zeros to $\mathbf{A}$ for the slack variable, but place -1 in that column for the last constraint. $\mathbf{A}$ then is a $(m+1) \times (2m+1)$ matrix. Let $\mathbf{f}$ be the $(2m+1) \times 1$ vector of variables (including the new slack variable) and $\mathbf{b}$ be the $(m+1) \times 1$ vector of constants from the right hand side of the constraints. The $(2m+1) \times 1$ vector $\mathbf{c}$ is the vector of coefficients in the objective function with an additional zero added for the slack variable.

For example, if $k = 2$ and $b = 3$, then, for general $\boldsymbol{p}$,

$$\mathbf{c} = \left(0,\ \Theta,\ 0,\ \Theta\left(1 + \frac{1}{\theta_2}\right),\ 0,\ \Theta\left(1 + \frac{1}{{\theta_2}^2}\right),\ 0,\ \Theta,\ 0,\ \Theta\left(1 + \frac{1}{\theta_2}\right),\ 0\right)^T,$$

$$\mathbf{f} = \left(f_{\mathbf{0}},\ f^1_{\mathbf{0}},\ f_{(1,0)},\ f^1_{(1,0)},\ f_{(2,0)},\ f^1_{(2,0)},\ f_{(1,1)},\ f^1_{(1,1)},\ f_{(2,1)},\ f^1_{(2,1)},\ f_s\right)^T,$$

$$\mathbf{A} = \begin{bmatrix} 1 & \Theta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -\Theta & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -\Theta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2/\theta_2 & 0 & 0 & -1 & -\Theta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/\theta_2 & 0 & -1 & -\Theta & 0 \\ 1/2 & 0 & 1 & 0 & 1 & 0 & 1/2 & 0 & 1 & 0 & -1 \end{bmatrix},$$

$$\mathbf{b} = (1,\ 0,\ 0,\ 0,\ 0,\ P^{\star})^T,$$

where $\Theta = \left(1 + \frac{1}{\theta_2}\right)$.

Since our non-negativity constraints are already in standard form, we can now represent the LP in standard form:

$$\begin{array}{lll} \text{minimize} & & \mathbf{c}^T\mathbf{f} \\ \text{subject to:} & & \mathbf{A}\mathbf{f} = \mathbf{b} \\ & & \mathbf{f} \geq \mathbf{0}. \end{array} \tag{4.12}$$

With our LP in standard form, we are now in a position to introduce and prove a sequence of lemmas that provide the foundation for our theorem regarding the optimal randomized procedure. But first, we show a proposition that is not strictly required to develop our theorem; however, it is an interesting result.

Proposition 4.1 *The $m+1$ rows of $\mathbf{A}$ are linearly independent.*

Proof: Let f_j and f_j^1 be the sink and outflow arc variables, respectively, for node j. To prove the proposition, we must show that none of the rows of $\mathbf{A}$ is a linear combination of the remaining rows. Each of the first m rows represents a flow conservation constraint for a particular node (including the initialization constraint as a special case of flow conservation). For any such row, say row i, there is a –1 (or a 1 in the case of the initialization constraint) in the column for f_i. This is evident in the reformulation and the previous example. No other flow conservation constraint has a nonzero entry in this column and so the remaining $m-1$ entries in the column for f_i cannot be linearly combined to equal –1 or 1, which implies that none of the remaining $m-1$ rows can be linearly combined to create row i.

The P(CS) constraint (row $m+1$), on the other hand, will have a nonzero entry in the column for f_i; however, it also has the only nonzero entry in the column for the slack variable f_s. Therefore, it cannot be used in any linear combination to make up one of the remaining rows. Nor can any linear combination of the remaining rows be used to create the P(CS) constraint, since they all have zeros in the column for the slack variable. Thus,

none of the rows of $\mathbf{A}$ are a linear combination of the remaining rows, and so the $m+1$ rows of $\mathbf{A}$ are linearly independent. $\square$

Lemma 4.1 *If the feasible set of solutions to the LP is nonempty, then there exists an optimal solution with, at most, $m+1$ basic variables, with the remaining m (nonbasic) variables equal to zero.*

Proof: The theory of linear optimization (for example, see Bertsimas and Tsitsiklis 1997) tells us that, for a problem in the standard form, the number of variables in an optimal basic feasible solution, $\mathbf{f}^\star$, is equal to the number of linearly independent rows of $\mathbf{A}$, i.e., the rank of $\mathbf{A}$. The remaining variables, called nonbasic variables, must be equal to zero. While Proposition 4.1 does in fact prove that the $m+1$ rows of $\mathbf{A}$ are linearly independent, we really only need the fact that there are, at most, $m+1$ basic variables since there are only $m+1$ constraints. Thus, there must exist an optimal solution with, at most, $m+1$ basic variables. Since there are a total of $2m+1$ variables, the remaining m or more variables must be equal to zero. $\square$

Let $\mathbf{f}^\star$ be an optimal solution to the randomized LP formulation, $\mathcal{LP}$, of a network consisting of the m nodes in set $\mathcal{N}''$. We use $\mathcal{N}'''$ to represent the subset of $\mathcal{N}''$ that consists only of the m' nodes through which there is nonzero flow in solution $\mathbf{f}^\star$, i.e., nodes j for which $f_j > 0$ or $f_j^1 > 0$. Now consider an LP formulation similar to $\mathcal{LP}$, but where we remove the flow conservation constraints for nodes not in $\mathcal{N}'''$, as well as all variables (and their corresponding non-negativity constraints) associated with nodes not in $\mathcal{N}'''$. We call the new formulation $\mathcal{LP}'$ and make the following claim.

Lemma 4.2 *If the $(2m+1)\times 1$ vector $\mathbf{f}^\star$ is an optimal solution to $\mathcal{LP}$, then the $(2m'+1)\times 1$ vector $\mathbf{f}^{\star\prime}$, formed by deleting the components associated with nodes not in set $\mathcal{N}'''$, is an optimal solution to $\mathcal{LP}'$.*

Proof: We prove this statement by contradiction. Suppose that $\mathbf{f}^{\star\prime}$ is not the optimal solution to $\mathcal{LP}'$. Then one of two cases must be true. Either $\mathbf{f}^{\star\prime}$ is infeasible or there must be another solution that is optimal.

We consider the former first. Clearly, solution $\mathbf{f}^{\star\prime}$ does not violate the non-negativity constraints in $\mathcal{LP}'$. Basic linear algebra shows us that since the $(2m+1)$-dimensional vector $\mathbf{f}^{\star}$ satisfies $\mathbf{A}\mathbf{f}^{\star} = \mathbf{b}$, then the $(2m'+1)$-dimensional vector $\mathbf{f}^{\star\prime}$, created by deleting components of $\mathbf{f}^{\star}$, must satisfy $\mathbf{A}'\mathbf{f}^{\star\prime} = \mathbf{b}'$, where $\mathbf{A}'$ is created by deleting columns corresponding to the components deleted from $\mathbf{f}^{\star}$, and $\mathbf{b}'$ is created by deleting the components corresponding to those deleted from $\mathbf{f}^{\star}$. But the equality constraint set in $\mathcal{LP}'$ is just $\mathbf{A}'\mathbf{f}^{\star\prime} = \mathbf{b}'$ with additional rows removed corresponding to the deleted nodes. We know that if a solution is feasible for a set of constraints, it must still be feasible when some of those constraints are removed. Thus, $\mathbf{f}^{\star\prime}$ must be feasible.

Now, let us assume that there exists an optimal solution $\mathbf{f}^{\dagger} \neq \mathbf{f}^{\star\prime}$ for which $\mathbf{c}'^{T}\mathbf{f}^{\dagger} < \mathbf{c}'^{T}\mathbf{f}^{\star\prime}$, where $\mathbf{c}'$ is the cost vector in the objective function of $\mathcal{LP}'$. We create a new vector, $\mathbf{f}^{\dagger\prime}$, in the $(2m+1)$-dimensional space of the original problem, $\mathcal{LP}$, by adding zeros to $\mathbf{f}^{\dagger}$ for the additional variables not in $\mathcal{LP}'$. Since we are just adding zeros, $\mathbf{c}^{T}\mathbf{f}^{\dagger\prime} = \mathbf{c}'^{T}\mathbf{f}^{\dagger}$; and since $\mathbf{f}^{\star}$ is just $\mathbf{f}^{\star\prime}$ with zeros for the variables they do not have in common, then $\mathbf{c}^{T}\mathbf{f}^{\star} = \mathbf{c}'^{T}\mathbf{f}^{\star\prime}$. But that implies that

$$\mathbf{c}^{T}\mathbf{f}^{\dagger\prime} = \mathbf{c}'^{T}\mathbf{f}^{\dagger} < \mathbf{c}'^{T}\mathbf{f}^{\star\prime} = \mathbf{c}^{T}\mathbf{f}^{\star},$$

which is a contradiction, since $\mathbf{f}^{\star}$ is the optimal solution to $\mathcal{LP}$. □

Consider an MSP with a large observation budget, b, and a large number of nodes, but a relatively modest $P^{\star}$ requirement. In such a problem, the maximum possible value of random variable N_{R} (i.e., the number of observations taken during Procedure $\mathcal{M}_{\mathrm{R}}$ at termination) in the optimal solution may be significantly less than the budget. In such a solution, nodes will be 'cut off' due to early termination and the pairs of $\mathbf{f}^{\star}$ entries corresponding to those nodal variables will both be equal to zero. Lemma 4.2 is important because it allows us to assume, for purposes of discussion and without loss of generality, that our optimal solution contains only variables representing the nodes through which there is nonzero flow, as well as the slack variable. If that is not the case, we could create a new network in a variable subspace with the same optimal solution.

Now, by our assumption enabled by Lemma 4.2, at least one of each pair of variables for a node in our optimal solution must be nonzero (and thus a basic variable). That accounts

for m of the (at most) $m+1$ basic variables which we know must exist by Lemma 4.1. For the particular optimal solution we are considering (there may be more than one), we know that we can have at most one other variable in the basis: the slack variable or a second variable from a particular node pair. A randomized node is one for which both variable entries are nonzero, i.e., there is some probability that flows to the sink and some probability that flows to other nodes. Thus, there can only be, at most, one randomized node. This leads us to the following theorem.

Theorem 4.1 *If the feasible set of solutions to $\mathcal{LP}$ is nonempty, then there exists an optimal solution with no more than one randomized node, that is, a node at which randomization will determine whether sampling stops or continues.*

Our theorem tells us that there exists an optimal solution with at most one randomized node, but it does not guarantee that we will find it. Fortunately, the following procedure guarantees that we will find, upon procedure termination, an optimal solution with no more than one randomized node, provided that the LP is feasible and that we are using a simplex-based method that searches for the optimal solution among the basic solutions.

Procedure

1. Solve the LP.
2. Determine the number of randomized nodes ξ in the solution.
   (a) If $\xi > 1$, identify the set of nodes for which both variables are equal to zero in the optimal solution. For each node in that set, eliminate its variables and its associated flow conservation constraint. Return to the first step.
   (b) If $\xi \leq 1$, terminate the procedure.

Since we require a simplex-based method for solving the LP, our optimal solution must be an optimal basic solution. Optimization theory, combined with the structure of our LP, tells us that an optimal basic solution has at most $m+1$ basic variables. But suppose there

is a node $\boldsymbol{\eta}^\dagger$ for which $f_{\boldsymbol{\eta}^\dagger} = f^1_{\boldsymbol{\eta}^\dagger} = 0$ (i.e., no flow). In that case, both $f_{\boldsymbol{\eta}^\dagger}$ and $f^1_{\boldsymbol{\eta}^\dagger}$ may be nonbasic variables (since they are equal to zero), freeing up a basic variable to create an additional randomized node. In general, let m' be the number of nodes through which there is some flow in the optimal solution (i.e., at least one variable is nonzero). We now have the possibility that the $m+1$ basic variables are distributed over the m' nodes, resulting in $\xi \leq m - m' + 1$ randomized nodes. By eliminating nodes through which there is no flow after each iteration, we must either arrive eventually at an optimal solution in which there is at most one randomized node or eliminate all of the nodes in the network. The latter is not possible by the initialization and non-negativity constraints.

The implications of this theorem and the subsequent procedure are significant. One of the main criticisms that could be leveled against a procedure like Procedure $\mathcal{M}_{\mathrm{R}}$ is that having to randomize at numerous potential stopping nodes might be burdensome and difficult to justify. Theorem 4.1 shows that there exists a solution with at most one randomized stopping node and the procedure provides us a way to find it. This simplifies implementation significantly and is easier to justify to a decision-maker.

4.2.2 Achieved P(CS) for Procedure $\mathcal{M}_{\mathrm{R}}$

We now let $\mathcal{LP}$ be our original problem formulation (i.e., no longer in standard form):

$$\begin{aligned} &\text{minimize} && \mathbf{c}^T\mathbf{f}, \\ &\text{subject to:} && \mathbf{A}\mathbf{f} = \mathbf{b}, \\ & && \mathbf{a}^T\mathbf{f} \geq P^\star, \\ & && \mathbf{f} \geq \mathbf{0}, \end{aligned}$$

In this formulation, we have separated the P(CS) constraint from $\mathbf{A}$ in the standard form and there is no longer a slack variable. Let $\mathcal{LP}''$ be the LP formulation with the P(CS) constraint removed. Now consider the vector $\mathbf{f}' = (1, 0, \ldots, 0)^T$, in which all variables are set to zero except for the variable representing flow along the arc from the root node, $\mathbf{0}$, to the sink, which equals 1.

Lemma 4.3 *$\mathbf{f}'$ is the unique optimal solution to $\mathcal{LP}''$.*

Proof: We must show that $\mathbf{f}'$ is a feasible solution. First, we note that $\mathbf{f}' \geq \mathbf{0}$. Additionally, none of the flow conservation constraints except for the initialization constraint involve the variable for the arc from the root node to the sink (the only nonzero variable). Thus, the left hand sides of all but the first equality constraint must equal zero and are therefore satisfied. The initialization constraint is $f'_1 + \Theta f'^1_1 = 1$, which simplifies to $1 = 1$. Thus $\mathbf{f}'$ is feasible.

All of our variables and our cost function are non-negative, so the value of the objective function cannot be less than 0. Since $\mathbf{c}^T\mathbf{f}' = 0$, $\mathbf{f}'$ is an optimal solution to $\mathcal{LP}''$. To see that $\mathbf{f}'$ is a unique optimal solution to $\mathcal{LP}''$, we suppose that there exists an optimal solution, $\mathbf{f}^\ddagger \neq \mathbf{f}'$. For the new solution to be unique, we must have one of three cases. In the first case, $f_j^{\ddagger 1} > 0$ for some $j \geq 1$. We know that the coefficients of all arc variables in the objective function are nonzero, so $f_j^{\ddagger 1} > 0$ implies $\mathbf{c}^T\mathbf{f}^\ddagger > 0$, and $\mathbf{f}^\ddagger$ is not optimal — a contradiction. In the second case, $f_1^\ddagger < 1$. That implies that $f_1^{\ddagger 1} > 0$ by the initialization constraint, reducing this case to the first. In the third case, $f_j^\ddagger > 0$ for some $j \geq 2$. Flow conservation implies that there must be nonzero flow along at least one arc into such a node j, which again reduces this case to the first. Therefore, $\mathbf{f}'$ is the unique optimal solution to $\mathcal{LP}''$. $\square$

Now consider the original problem, $\mathcal{LP}$. Solution $\mathbf{f}'$ leads to a $\mathsf{P}(\mathrm{CS}) = \mathbf{a}^T\mathbf{f}' = 1/k$. If $P^\star \leq 1/k$, then $\mathbf{f}'$ is a feasible solution to $\mathcal{LP}$ and thus an optimal solution since $\mathbf{c}^T\mathbf{f}' = 0$. In that case, $\mathsf{P}(\mathrm{CS}) = 1/k$ at the optimal solution. If $P^\star > 1/k$, then $\mathbf{f}'$ is no longer feasible, since it violates the $\mathsf{P}(\mathrm{CS})$ constraint. Optimization theory tells us that if the addition of a constraint changes the optimal solution to an LP, then it must be an active constraint at the optimal solution. In our problem, that means that $\mathsf{P}(\mathrm{CS}) = P^\star$ when $P^\star > 1/k$. We can now state the following theorem.

Theorem 4.2 *For the optimal solution to the linear program, if it exists,*

$$\mathsf{P}(\mathrm{CS}) = \begin{cases} \frac{1}{k} & \text{if } P^\star < \frac{1}{k} \\ P^\star & \text{otherwise.} \end{cases}$$

This theorem also has implications for our discussion in §4.2.1. A degenerate solution is one in which one or more of the basic variables is equal to zero. Trivially, that occurs whenever there is no flow through at least one node (i.e., both nodal variables are zero). More interestingly, if we have an optimal solution in which there is flow into every node (which we can reduce any problem to by Lemma 4.2), that solution is degenerate if one of each pair of nodal variables and the slack variable equal zero.

In our discussion for Theorem 4.1, we point out that there exists an optimal basic feasible solution in which *either* both variables for one node are basic *or* the slack variable is basic. Theorem 4.2 tells us that if $P^\star < 1/k$, then the slack variable, f_s, in the standard formulation must be positive. If $f_s > 0$, then it is a basic variable, none of the nodes are randomized, and the solution is not degenerate. If $P^\star \geq 1/k$, then $f_s = 0$. In this case, if none of the nodes are randomized in the optimal basic feasible solution, then it is a degenerate solution; otherwise, it is not degenerate.

4.3 Replicating Existing Procedures

There is tremendous flexibility in the optimization framework, allowing us to manipulate the objective function and constraints in order to replicate other procedures or specify additional requirements on the solution. This section describes one such set of capabilities.

Suppose we wish to choose the optimal parameter settings for an existing procedure using the optimization (in order to expand tables, etc.). As an example, consider Procedure $\mathcal{M}_{\mathrm{RA}}$, which has an inverse sampling parameter t and a difference parameter r. In order to find the values of the (r, t)-pair which minimize the expected number of observations for a particular combination of k, $\theta^\star$, and $P^\star$ under budget b, we must search over the two-dimensional grid of possible (r, t)-pairs, calculating $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ and $\mathsf{E}_{\mathrm{SC}}[N_{\mathrm{RA}}]$ for each. We would then select the (r, t)-pair with $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS}) \geq P^\star$ that minimizes the expected number of observations. This technique may be reasonable for creating tables for common choices of $k, \theta^\star, P^\star$, and b, but it is not at all convenient if we often need optimal settings for uncommon choices of the parameters. We can overcome this disadvantage by adding additional constraints to our optimization.

4.3.1 Notation

Consider any existing non-randomized procedure, such as those discussed in Chapter 2. We will use, as a starting point, our reformulated LP from §4.1.4, with the exception that our network will consist of the *uncurtailed* set of left-lexicographic nodes, $\mathcal{N}'$, instead of $\mathcal{N}''$, so that we can replicate procedures that do not use curtailment. Let Γ be the set of all stopping policies for that procedure, with γ, a single policy, being an element of Γ. The observation budget b may be explicitly stated, or implied by procedure parameters over which we choose to search. We use $\mathcal{S}_\gamma$ to denote the set of all nodes $\boldsymbol{\eta}' \in \mathcal{N}'$ such that $\boldsymbol{\eta}'$ is a stopping node under procedure policy $\gamma \in \Gamma$. We also define a new set of binary variables,

$$Z_\gamma \;\equiv\; I_{\left[\text{if } \gamma \text{ is the active policy}\right]}. \tag{4.13}$$

As an example, consider Procedure $\mathcal{M}_{\text{RA}}$. In that case,

$$\Gamma = \left\{ (r,t) : r \le t \text{ and } t \le \left\lceil \frac{b}{k} \right\rceil \right\},$$

where $\lceil x \rceil$ is the ceiling function (i.e., rounds x up to the nearest integer). Suppose $\gamma = (2,6)$. Then

$$\mathcal{S}_\gamma = \left\{ \eta' \in \mathcal{N}' : \eta_{[k]m} \ge 6 \quad \text{or} \quad \eta_{[k]m} - \eta_{[k-1]m} \ge 2 \right\}.$$

4.3.2 Formulation

The new formulation shown in Figure 4.3 has four additional sets of constraints, all involving the new binary variables, Z_γ. The first new set of constraints, labeled "Stops", activates stops by requiring that the outbound flow along arcs to other nodes be set to zero for all nodes that are stops when policy γ is active. The next set of constraints, labeled "Non-stops", forces all flows to the sink to be set to zero for nodes that are not stops when policy γ is active. The constraint labeled "Choose Policy" ensures that only one policy can be active at a time. The final set of new constraints, labeled "Binary", declares the binary variables.

We have successfully implemented Procedures $\mathcal{M}_{\text{RA}}$, $\mathcal{M}_{\text{RA}'}$, and $\mathcal{M}_{\text{C}}$ using this formulation. Procedures $\mathcal{M}_{\text{BEM}}$, $\mathcal{M}_{\text{BK}}$, and $\mathcal{M}_{\text{BG}}$ can also be modeled in this way; however,

$$\min_{f_{\boldsymbol{\eta}'}, f^1_{\boldsymbol{\eta}'}, Z_\gamma} \quad \Theta \sum_{\boldsymbol{\eta}' \in \mathcal{N}'} \sum_{\boldsymbol{\eta} \in \pi(\boldsymbol{\eta}')} \prod_{i=2}^{k} \theta_i^{\eta_{[k-i+1]} - \eta_i} f^1_{\boldsymbol{\eta}'}$$

Subject to:

$$\text{(Initialization)} \qquad f_{\mathbf{0}} + \Theta f^1_{\mathbf{0}} = 1,$$

$$\text{(Flow Conservation)} \qquad \sum_{\nu=1}^{k} \frac{1}{\theta_\nu} \prod_{i=2}^{k} \theta_i^{\eta^\nu_{[k-i+1]} - \eta^\nu_1} f^1_{L(\boldsymbol{\eta}^\nu)} - f_{\boldsymbol{\eta}'} - \Theta f^1_{\boldsymbol{\eta}'} = 0, \qquad \forall \boldsymbol{\eta}' \in \mathcal{N}' \setminus \mathbf{0}$$

$$\text{(Minimum P(CS))} \qquad \sum_{\boldsymbol{\eta}' \in \mathcal{N}'} \frac{1}{t(\boldsymbol{\eta}')} \sum_{\boldsymbol{\eta} \in \pi^\star(\boldsymbol{\eta}')} \prod_{i=2}^{k} \theta_i^{\eta_{[k-i+1]} - \eta_i} f_{\boldsymbol{\eta}'} \geq P^\star,$$

$$\text{(Non-negativity)} \qquad f_{\boldsymbol{\eta}'}, f^1_{\boldsymbol{\eta}'} \geq 0, \qquad \forall \boldsymbol{\eta}' \in \mathcal{N}'$$

$$\text{(Stops)} \qquad \sum_{s \in \mathcal{S}_\gamma} f^1_s \leq 1 - Z_\gamma, \qquad \forall \gamma \in \Gamma$$

$$\text{(Non-stops)} \qquad \sum_{s \notin \mathcal{S}_\gamma} f_s \leq 1 - Z_\gamma, \qquad \forall \gamma \in \Gamma$$

$$\text{(Choose Policy)} \qquad \sum_{\gamma \in \Gamma} Z_\gamma = 1,$$

$$\text{(Binary)} \qquad Z_\gamma \quad \text{binary} \qquad \forall \gamma \in \Gamma$$

Figure 4.3: MIP Formulation for Replicating Existing Procedures

since there is only one parameter, the truncation parameter n, to search over for these procedures, implementing the above MIP formulation would probably not be worth the effort versus simply searching over n. Nonetheless, if we choose to do so, Procedure $\mathcal{M}_{\text{BG}}$ requires special care in the above formulation. Recall that one of the Procedure $\mathcal{M}_{\text{BG}}$ stopping conditions in (2.3) is a function of $\theta^\star$ and $P^\star$. Such stops are fixed under all $\gamma \in \Gamma$ and should then either 1) be included in S_γ for all $\gamma \in \Gamma$, or 2) have their outbound arc, $f^1_{\boldsymbol{\eta}'}$, set to zero in an additional set of equality constraints. We have also successfully replicated Procedure $\mathcal{M}_{\text{BG}}$ in this way.

4.4 Summary

In this chapter, we have built upon the work of Chapter 3 by reformulating the MIP and LP mathematical programs to overcome some of the most significant drawbacks of the initial formulations. Problem symmetries and strong curtailment provide us the necessary relationships to represent the entire network of possible stopping nodes by a smaller network consisting only of left-lexicographic, non-curtailed nodes. Although they are less intuitive, the reformulations significantly reduce the computational requirements for solving the problem and have allowed us to consider substantially larger problems than could be solved initially.

Our reformulations have also made it easier to prove some key insights about the optimal randomized MSP procedures. First, we showed that an optimal procedure exists which has, at most, only one randomized node, and we developed a procedure to find such a solution. Second, we showed that $\mathsf{P}(\text{CS}) = P^\star$ when $P^\star \geq 1/k$.

Last, we showed that we can extend our formulations to replicate non-randomized MSP procedures, thereby facilitating the identification of the optimal procedure parameters for a specific problem. This can be a significant capability if an experimenter needs to identify such parameters often for problem characteristics that do not have tabulated results. Otherwise, for many procedures, we must conduct a detailed search of the parameter space, which can be quite time-consuming and computationally expensive.

CHAPTER V

INTEGRATION OF OBSERVATION COSTS

The implicit assumption of all existing MSP procedures, when developed to minimize the expected number of observations, is that all observations have equal cost. But what if observation costs are not constant? What if marginal observation costs increase, decrease, or both, as the experiment proceeds? Does purchasing the required observation supplies in batches affect procedure performance? Questions such as these are completely reasonable, particularly when we consider non-monetary resource expenditures. Here are some examples:

- Certain experimentation supplies can be purchased in batches of a predetermined size. Variable costs then may be incurred for each observation and at the beginning of each batch of observations.
- Costs are fixed for each observation; however, we wish to discount our observation costs over our experimental horizon and evaluate our procedure based upon current dollars.
- We wish to model the decision-maker's preference for the timeliness of decisions using some form of nonlinear utility or value function, instead of assuming that the decision delays inherent in taking additional observations are of equal utility or value.
- We are concerned with 'wear and tear' on a mechanical system required for an experiment, since it would be better to use the system for actual production than for testing. We wish to assign 'costs' to each observation taken based upon a known degradation function.

No existing MSP literature addresses such cost issues, largely because the tools required for such analysis have not heretofore been in place. Our formulations provide us the means to do so.

5.1 Methodology

This section describes our methodologies for incorporating variable observation costs and analyzing their effects on MSP procedure performance. In §5.1.1, we develop methodologies that guarantee optimal randomized and non-randomized MSP procedures under a very broad class of observation cost functions. We then describe the types of cost functions that we will consider in our analysis in §5.1.2. Finally, in §5.1.3, we describe the metrics that we will use to evaluate procedure performance.

5.1.1 Cost Integration

Consider Procedures $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{NR}}$. Theorems 3.1 and 3.2, combined with the reformulated mathematical programs in Chapter 4 guarantee that a randomized [non-randomized] procedure created from the solution to $\mathcal{LP}$ [$\mathcal{MIP}$], is an optimal randomized [non-randomized] solution (in terms of the expected number of observations) to the constrained, indifference zone MSP for a specified probability configuration $\boldsymbol{p}$. We wish to modify our procedures and the mathematical formulations in such a way that we can guarantee optimality for the more general set of total cost functions.

Now let us consider the general set of total costs. Recall that we are using the term *cost* in the broader sense to encompass any type of expenditure (e.g., effort, resources, penalties, opportunity cost, etc.). We will refer to the cost required to take one additional observation as the *marginal cost* (MC). The *total cost* (TC) at observation m, then, refers to the cumulative cost of taking m observations. We denote the MC of observation j as $c(j)$ and the TC of all observations taken through observation m as $TC(m)$, i.e., $TC(m) = \sum_{j=1}^{m} c(j)$. We assume that total cost is a non-decreasing function, i.e., $c(j) \geq 0 \ \forall j$. We also assume that when an observation is made, it is made for all alternatives in accordance with the problem definition of Chapter 1.

Because all alternatives must be observed each time, there is no need to consider different observation costs among the alternatives. Instead, we incur a single marginal cost for each observation. In the most general case, each cumulative success vector $\boldsymbol{\eta}$ has a unique cost, $c_{\boldsymbol{\eta}}$, but we require that all permutations of $\boldsymbol{\eta}$ have the same observation cost.

Cost Formulations In developing the objective function for the LP and MIP formulations, we showed that the expected number of observations is the sum of the flows along all of the success arcs (i.e., arcs connecting one node to another). Here, we integrate cost by following the same reasoning. Let C be a random variable representing the total cost at experiment termination and $\mathcal{N}$ be the set of all possible cumulative success vectors $\boldsymbol{\eta}$. Then

$$C \;=\; \sum_{\boldsymbol{\eta}\in\mathcal{N}} c_{\boldsymbol{\eta}}\, I_{\left[\text{observation taken at node } \boldsymbol{\eta}\right]}.$$

Thus,

$$\begin{aligned}
\mathsf{E}[C] &= \sum_{\boldsymbol{\eta}\in\mathcal{N}} c_{\boldsymbol{\eta}}\, \mathsf{P}(\text{observation taken at node } \boldsymbol{\eta}) \\
&= \sum_{\boldsymbol{\eta}\in\mathcal{N}} c_{\boldsymbol{\eta}}\, \mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and do not stop}) \\
&= \sum_{\boldsymbol{\eta}\in\mathcal{N}} c_{\boldsymbol{\eta}}\, (1-p_{\boldsymbol{\eta}})\, \mathsf{P}(\text{arrive at } \boldsymbol{\eta}) \\
&= \sum_{\boldsymbol{\eta}\in\mathcal{N}} c_{\boldsymbol{\eta}} \sum_{i=1}^{k} f_{\boldsymbol{\eta}}^{i}. \qquad (5.1)
\end{aligned}$$

Using the relationships (4.2) and (4.4) derived in Chapter 4, we can rewrite (5.1) as

$$\begin{aligned}
\mathsf{E}[C] &= \sum_{\boldsymbol{\eta}\in\mathcal{N}} c_{\boldsymbol{\eta}} \sum_{i=1}^{k} f_{\boldsymbol{\eta}}^{i} \\
&= \Theta \sum_{\boldsymbol{\eta}\in\mathcal{N}} c_{\boldsymbol{\eta}} f_{\boldsymbol{\eta}}^{1} \\
&= \Theta \sum_{\boldsymbol{\eta}'\in\mathcal{N}'} c_{\boldsymbol{\eta}'} \sum_{\boldsymbol{\eta}\in\pi(\boldsymbol{\eta}')} \prod_{i=2}^{k} \theta_i^{\eta_{[k-i+1]}-\eta_i}, \qquad (5.2)
\end{aligned}$$

where $\pi(\boldsymbol{\eta}')$ is the set of permutations of $\boldsymbol{\eta}'$ within which $\boldsymbol{\eta}$ is contained.

The objective functions for the randomized and non-randomized formulations both minimize the same function $\mathsf{E}[C]$, with the only difference being the variables over which the mathematical program is minimized. In fact, even the non-randomized formulation shown in §4.3 for replicating other non-randomized procedures uses the same objective function. We create $\mathcal{LP}_{\mathcal{C}}$ by replacing the objective function in $\mathcal{LP}$ with

$$\min_{f_{\boldsymbol{\eta}'},\, f^{1}_{\boldsymbol{\eta}'}} \; \mathsf{E}[C], \qquad (5.3)$$

where $\mathsf{E}[C]$ is defined by (5.2). Similarly, we create $\mathcal{MIP}_{\mathcal{C}}$ by replacing the objective function in $\mathcal{MIP}$ with

$$\min_{f_{\boldsymbol{\eta}'}, f^1_{\boldsymbol{\eta}'}, Y_{\boldsymbol{\eta}'}} \mathsf{E}[C]. \tag{5.4}$$

Optimal Cost Procedures We now modify Procedure $\mathcal{M}_{\mathrm{R}}$ slightly by identifying the ordered set $\mathcal{S}_{\mathrm{R}}$ based upon $\mathcal{LP}_{\mathcal{C}}$ instead of $\mathcal{LP}$. This new Procedure $\mathcal{M}_{\mathrm{RC}}$ is a more general one, with Procedure $\mathcal{M}_{\mathrm{R}}$ being a special case when $c_{\boldsymbol{\eta}'} = 1$ for all $\boldsymbol{\eta}' \in \mathcal{N}'$. We can now state the following theorem.

Theorem 5.1 *For a specified probability configuration $\boldsymbol{p}$, a randomized procedure created from the solution to the LP formulation $\mathcal{LP}_{\mathcal{C}}$ is an optimal solution, in terms of minimizing the total observation cost, to the constrained, indifference zone MSP under arbitrary observation costs $\{c_{\boldsymbol{\eta}'}\}$.*

Similarly, we create new Procedure $\mathcal{M}_{\mathrm{NRC}}$ by identifying the set $\mathcal{S}_{\mathrm{NR}}$ based upon $\mathcal{MIP}_{\mathcal{C}}$ instead of $\mathcal{MIP}$. The next theorem follows.

Theorem 5.2 *For a specified probability configuration $\boldsymbol{p}$, a non-randomized procedure created from the solution to the MIP formulation $\mathcal{MIP}_{\mathcal{C}}$ is an optimal* non-randomized *solution, in terms of minimizing the total observation cost, to the constrained, indifference zone MSP under arbitrary observation costs $\{c_{\boldsymbol{\eta}'}\}$.*

In execution, the changes to the objective functions are very simple to implement. We merely multiply the existing coefficient of each $f^1_{\boldsymbol{\eta}'}$ by the cost $c_{\boldsymbol{\eta}'}$ of making an additional observation at cumulative success vector $\boldsymbol{\eta}'$. We point out that the same changes to the objective function apply if we wish to modify our formulations of other MSP procedures, as described in §4.3. These modifications follow directly from our discussion; we do not describe them here.

5.1.2 Cost Functions

Our formulations allow us to consider unique observation costs for each possible left-lexicographic cumulative success vector. While there may be a need for this capability,

we focus on observation costs that are based upon the number of observations that have already been taken. That is, the jth observation will cost $c(j)$, regardless of $\boldsymbol{\eta}_{j-1}$ (the cumulative success vector just before observation j is made).

Our intent in selecting cost functions was to choose a representative set of possible functional types and 'shapes'. We were not as concerned with when such functions might arise, but with capturing a reasonable set of well-behaved, non-decreasing TC functions that would lead us to general insights concerning their effects on procedure performance. Using cost as a proxy for non-monetary expenditures gives us great flexibility for tailoring experiments to decision-maker needs; however, it also makes the selection of an all encompassing set of real world cost functions impractical. In the discussion that follows, our focus is on TC functions, from which we will derive the necessary MCs. Our TC functions have the same initial total cost, $TC(0) = 0$, and the same total cost, $TC(b) = \tau$, at the observation budget b. We make this assumption in order to ensure a 'fair' comparison between different cost functions that is based upon the type and shape of the function.

We consider five general types of total cost functions: linear, convex, concave, mixed, and batch functions. We describe each below.

Linear This type of cost function is one in which the MC of each subsequent observation is constant; therefore, it is the implicit TC function for all existing MSP procedures. The total cost through observation m under the linear cost function is:

$$TC_L(m) = \beta_1 m, \tag{5.5}$$

where $\beta_1 \equiv \tau/b$ is chosen to ensure $TC(b) = \tau$.

Convex Convex costs are those for which marginal costs are monotonically increasing, i.e., $c(j+1) - c(j) > 0 \ \forall j$. Figure 5.1 shows the convex TC functions [left] that we examined, as well as their associated MC functions [right]. For reference, we include the linear function as well. We chose seven convex functions in two categories of shapes: power and exponential. Our four convex power functions are of the following form:

$$TC_{Xp}(m) = \beta_2 m^d, \tag{5.6}$$

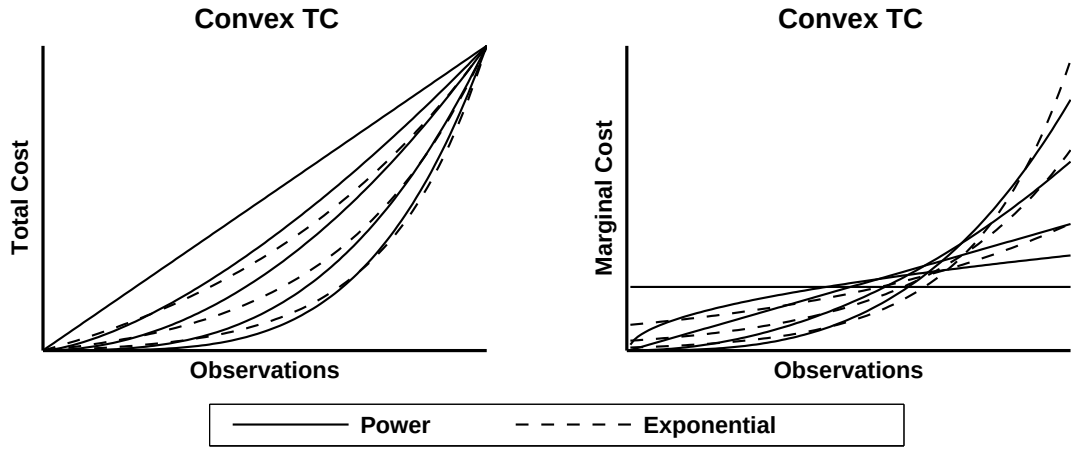

Figure 5.1: Convex Total Cost Functions

where $\beta_2 \equiv \tau/b^d$ and $d \in \{1.5, 2, 3, 4\}$. Note that linear costs are a special case of our power functions with $d = 1$; however, we keep linear cost as a separate type.

Our three convex exponential functions are of the form:

$$TC_{Xe}(m) = s\ {\beta_3}^m - s, \tag{5.7}$$

where $s \in \{1, 5, 25\}$ is a shift factor and $\beta_3 \equiv ((\tau + s)/s)^{1/b}$. We use the shift factor s to compensate for our starting cost of 0 and to create exponential curves with different degrees of convexity.

Concave Concave costs are those for which marginal costs are monotonically decreasing, i.e., $c(j+1) - c(j) < 0\ \forall j$. Figure 5.2 shows the seven concave TC functions (with associated MC functions) that we examined — one corresponding to each of our convex functions. Since we wished to make like comparisons between functions, our concave functions are generated directly from the convex functions as follows:

$$TC_V(m) = \tau - TC_X(b - m), \tag{5.8}$$

where $TC_X(b - m)$ is either the convex power TC function, $TC_{Xp}(b - m)$ or the convex exponential TC function, $TC_{Xe}(b - m)$.

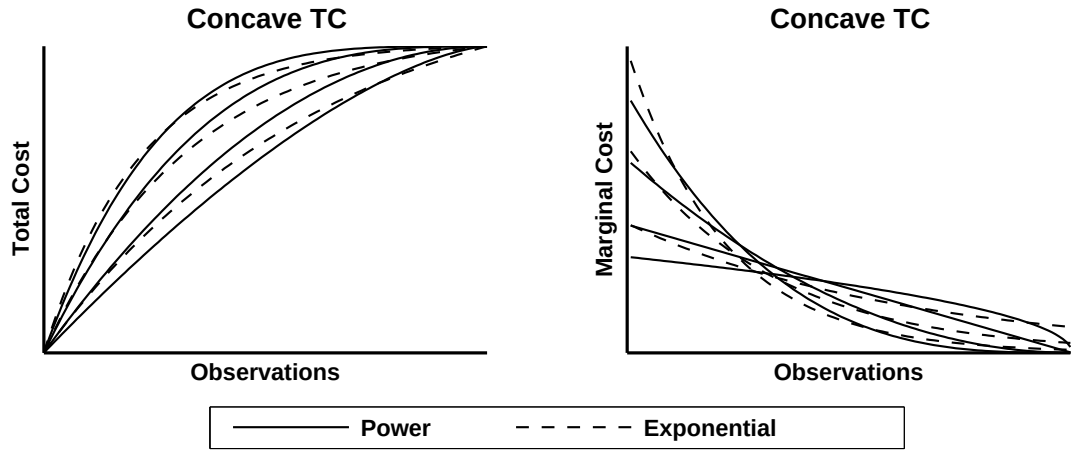

Figure 5.2: Concave Total Cost Functions

Mixed Mixed functions are those for which marginal costs are monotonic over a subinterval of the domain, but not over the entire domain. We chose 's-curves' to model these types of costs. In particular, our convex-concave mixed functions have monotonically increasing marginal costs for the first half of the observations and monotonically decreasing marginal costs for the second half of the observations. The TC and MC functions are the top two charts in Figure 5.3. We created four convex-concave TC functions using power functions:

$$TC_{XVp}(m) = \begin{cases} \beta_4 \, m^d & \text{if } m \leq b/2 \\ \tau - \beta_4 \, (b-m)^d & \text{otherwise,} \end{cases} \tag{5.9}$$

where $\beta_4 \equiv \frac{\tau/2}{(b/2)^d}$ and $d \in \{1.5, 2, 3, 4\}$. Similarly, we created three additional convex-concave functions using exponential functions:

$$TC_{XVe}(m) = \begin{cases} s \, {\beta_5}^m - s & \text{if } m \leq b/2 \\ \tau - s \, {\beta_5}^{b-m} + s & \text{otherwise,} \end{cases} \tag{5.10}$$

where $\beta_5 \equiv \left(\frac{\tau/2+s}{s}\right)^{(2/b)}$ and $s \in \{1, 5, 25\}$.

We also created concave-convex mixed TC functions with monotonically decreasing marginal costs for the first half of the observations and monotonically increasing marginal costs for the second half. They are derived in a similar manner to the functions already described and are shown graphically in the bottom two charts of Figure 5.3.

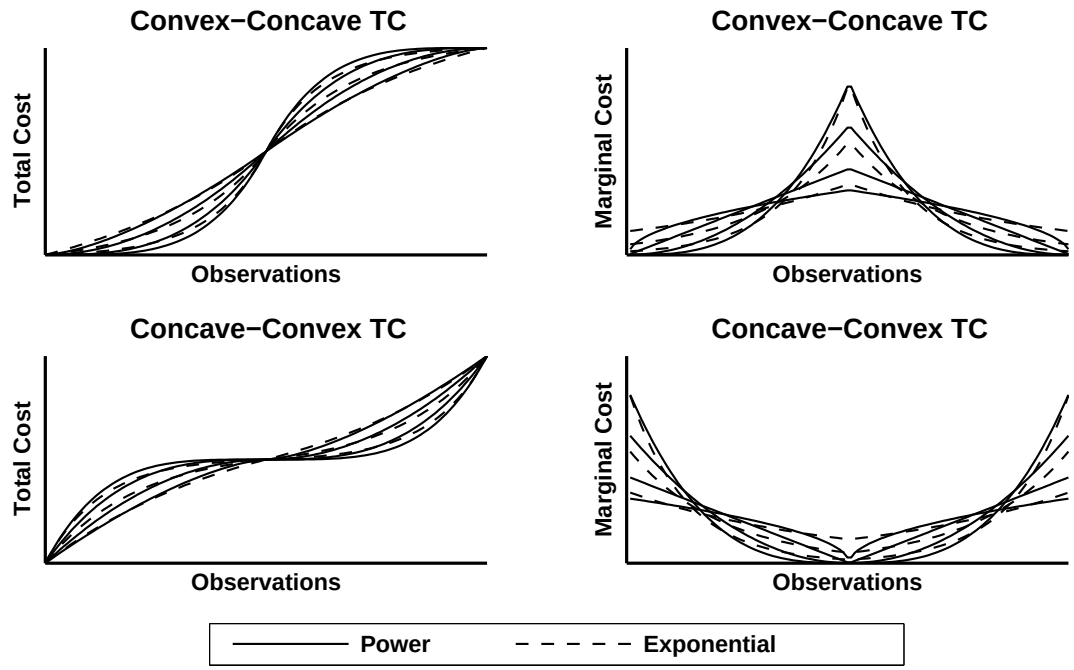

Figure 5.3: Mixed Total Cost Functions

Batch Batch functions represent periodic observation costs that lead to step functions (for TC). For such functions, there are two contributors to cost: periodic batch costs and the underlying continuous observation costs. For our analysis, we consider only linear underlying observation costs, so that we can focus on the effects of the periodic costs. For consistency, we characterize the costs due to batching by the fraction ρ of the total cost at budget b resulting from periodic costs. Let B represent the number of observations in each batch. We created our 16 batch TC functions by:

$$TC_B(m) = (1-\rho)\ \beta_1\ m + \rho\ \left(\frac{\tau}{\lceil b/B \rceil}\right) \left\lceil \frac{m}{B} \right\rceil, \tag{5.11}$$

where $\beta_1 \equiv \tau/b$, $\rho \in \{0.25, 0.5, 0.75, 1\}$ and $B \in \{2, 3, 4, 5\}$. The first term accounts for the constant marginal cost of each observation; the second term accounts for the periodic batch costs. When $\rho = 1$, observation costs are due entirely to batching.

Figure 5.4 shows two of the TC functions and their associated MC curves. Due to the requirement that $TC(b)$ be constant for all choices of B and ρ, there is significant overlapping of the functions, making it impractical to put all, or even many, TC curves on

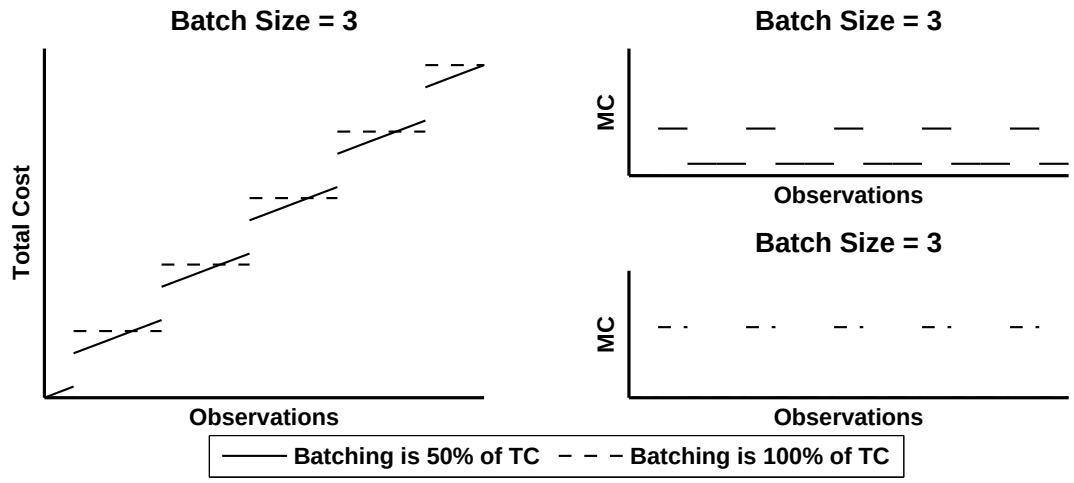

Figure 5.4: Example Batch Cost Functions with $B = 3$

one chart.

5.1.3 Measures of Performance

In this section, we briefly describe some of the measures that we can use in comparing procedure performance under different total cost functions. Our goal is to show that considering costs is important if the assumption of linear total costs is not valid.

Expected Total Cost The most common measure in the literature is the expected number of observations in the LFC. The assumption for all existing procedures — sometimes proven, sometimes conjectured — is that the SC is the LFC. Adapted for cost, the measure is the expected total cost in the SC, $\mathsf{E}_{\mathrm{SC}}[C]$. When comparing procedure performance under different cost functions, $\mathsf{E}_{\mathrm{SC}}[C]$ alone may not be very useful. For example, if we compare procedure performance under linear total costs to its performance under our convex TC functions, we will always find that $\mathsf{E}_{\mathrm{SC}}[C]$ is lower for convex TC functions than for linear functions. This is the direct result of the functions we chose — at every observation m, $TC_L(m) \geq TC_X(m)$. We are more interested in how the procedure performs under different cost functions *relative* to some benchmark.

Procedure Inefficiency Metric We develop a new metric to supplement $\mathsf{E}_{\mathrm{SC}}[C]$ and provide a relative measure for comparisons between cost functions and procedures. Procedure $\mathcal{M}_{\mathrm{RC}}$ is the ideal benchmark for relative comparisons, since it achieves the optimal $\mathsf{E}_{\mathrm{SC}}[C]$ for any TC function. Therefore, we use *procedure inefficiency*, W_{J}, as our primary metric for the performance of general procedure $\mathcal{M}_{\mathrm{J}}$. It is defined as:

$$W_{\mathrm{J}} \equiv \frac{\mathsf{E}_{\mathrm{SC}}[C_{\mathrm{J}}] - \mathsf{E}_{\mathrm{SC}}[C_{\mathrm{RC}}]}{\mathsf{E}_{\mathrm{SC}}[C_{\mathrm{RC}}]} = \frac{\mathsf{E}_{\mathrm{SC}}[C_{\mathrm{J}}]}{\mathsf{E}_{\mathrm{SC}}[C_{\mathrm{RC}}]} - 1, \tag{5.12}$$

where C is the random variable representing the total cost, $\mathsf{E}_{\mathrm{SC}}[C_{\mathrm{J}}]$ is the expected total cost in the SC using general procedure $\mathcal{M}_{\mathrm{J}}$, and $\mathsf{E}_{\mathrm{SC}}[C_{\mathrm{RC}}]$ is the expected total cost using Procedure $\mathcal{M}_{\mathrm{RC}}$. We can think of procedure inefficiency as the fractional increase in expected total cost incurred by using general procedure $\mathcal{M}_{\mathrm{J}}$ instead of the optimal Procedure $\mathcal{M}_{\mathrm{RC}}$.

For some procedures, the optimal procedure parameter settings (e.g., the (r, t)-pair for Procedure $\mathcal{M}_{\mathrm{RA}}$) may depend upon the TC function. However, for our analysis, we will calculate $\mathsf{E}_{\mathrm{SC}}[C_{\mathrm{J}}]$ for general procedure $\mathcal{M}_{\mathrm{J}}$ under its optimal parameter settings for minimizing the expected number of observations (i.e., the original case).

Remark 5.1 Procedures for which the only adjustable procedure parameter is the truncation parameter n (e.g., Procedures $\mathcal{M}_{\mathrm{BEM}}$, $\mathcal{M}_{\mathrm{BK}}$, and $\mathcal{M}_{\mathrm{BG}}$) will have the same optimal procedure parameter settings regardless of the cost function. For these procedures, the optimal n-values are chosen to meet the $P^\star$ requirement, which is not affected by the observation costs. Decreasing n will lead to a $\mathsf{P}(\mathrm{CS}) \leq P^\star$. Increasing n cannot lead to a lower expected cost, since it can only shift potential stops to points with equal or higher total costs (under the assumption of non-decreasing total costs).

5.2 Results

In this section, we examine the robustness of procedure performance under variable marginal observation costs, in order to show that considering such costs is important if the assumption of constant costs is invalid. In §5.2.1, we narrow our focus to the effects of continuous, nonlinear TC functions (e.g., convex, concave, and mixed). We then shift our focus to

periodic (batch) costs in §5.2.2. In §5.2.3, we introduce a tool for informing the decision of whether or not to purchase observations in batches. Finally, in §5.2.4, we briefly discuss additional insights that we can gain from the LP formulation of the optimal randomized procedure.

5.2.1 Effects of Continuous, Nonlinear Total Costs

A challenge when presenting results is that of deciding what cases to discuss. Each potential MSP is parameterized by k, $\theta^\star$, $P^\star$, and b. In much of the existing literature, common choices for the first three parameters are $k \in \{2, 3, 4\}$, $\theta^\star \in \{1.6, 2, 2.4, 3\}$, $P^\star \in \{0.75, 0.9, 0.95\}$. Explicit consideration of a budget, however, adds too many possible variations. In our examples, for a given k, $\theta^\star$, and $P^\star$, we set b equal to the optimal truncation procedure parameter n_{BG} for Procedure $\mathcal{M}_{\text{BG}}$, consistent with our methodology in Chapter 3.

The discussion regarding the effects of continuous, nonlinear TC functions on procedure performance is divided into two parts: the effects on optimal procedures in §5.2.1.1 and the effects on non-optimal procedures in §5.2.1.2.

5.2.1.1 Effects on Optimal Randomized Procedures

In this section, we examine W_{R} — the fractional increase in expected TC in the SC when using the optimal randomized Procedure $\mathcal{M}_{\text{R}}$ without considering the TC function. This is a situation in which the experimenter has the capability to create the optimal randomized procedure, but makes the incorrect assumption that MCs are constant. We analyzed 28 of the possible 36 combinations of common choices of k, $\theta^\star$, and $P^\star$; the remaining eight combinations were not run due to the relatively large MIP solver run times for Procedure $\mathcal{M}_{\text{NRC}}$ under large b. For our discussion, we will focus on a representative subset of nine cases with $\theta^\star = 2.4$, shown in Table 5.1.

Figure 5.5 shows W_{R} plotted for each case for the power-based functions (black squares) and exponential-based functions (blue triangles). Plotted shapes of increasing size indicate larger degree d [smaller shift s] for the power [exponential] TC functions. In other words, the larger the shape, the greater the TC function deviates from linear. There are four plots,

Table 5.1: Subset of Cases with $\theta^\star = 2.4$

Case	k	$P^\star$	b	Case	k	$P^\star$	b	Case	k	$P^\star$	b
1	2	0.75	3	**4**	3	0.75	8	**7**	4	0.75	15
2	2	0.9	11	**5**	3	0.9	22	**8**	4	0.9	31
3	2	0.95	17	**6**	3	0.95	31	**9**	4	0.95	44

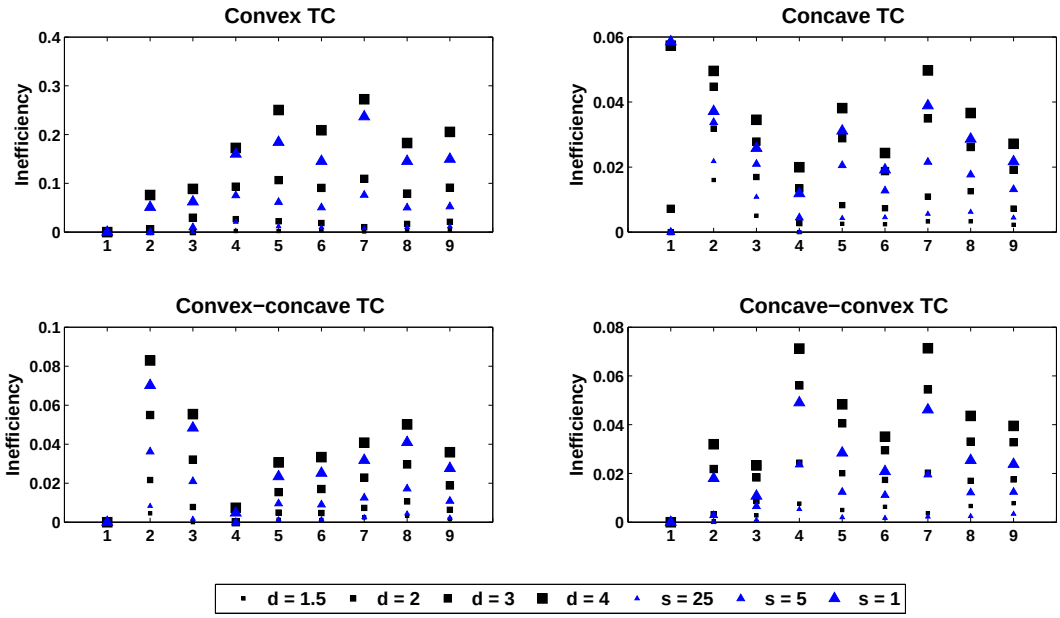

Figure 5.5: Procedure $\mathcal{M}_{\mathrm{R}}$ Inefficiency for Selected Cases with $\theta = 2.4$

one each for each type of total cost function — convex, concave, convex-concave (mixed), and concave-convex (mixed). Note that the scaling of the vertical axis is not consistent across the four plots.

We make a few observations based upon Figure 5.5.

1. Increasing the deviance of the TC function from linear tends to increase the inefficiency of Procedure $\mathcal{M}_{\mathrm{R}}$ for all types and shapes of functions. In the 28 cases that we examined, we found only one exception in which W_{R} for a concave power function with $d = 4$ was actually slightly lower than with $d = 3$.

2. Usually (though not always, see, for instance, Case 1), convex functions have the largest impact on W_{R}. For example, Case 3 has $W_{\mathrm{R}} = 0.273$ under a convex power

function with $d = 4$. The largest inefficiency for any of the other three types of functions is $W_{\mathrm{R}} = 0.083$ for Case 3 under a convex-concave power function with $d = 4$.

3. The particular *shape* of the total cost function matters (i.e., power versus exponential functions). Although we did not attempt to create exponential functions whose shift s corresponds in some way to each particular power function with degree d, Figure 5.5 provides the necessary insight anyway. For example, Cases 2–4 have greater W_{R} for concave power functions with $d = 3$ than for concave exponential functions with $s = 1$. On the other hand, for Cases 5–9, the reverse is true. If the particular shape did not matter, we would expect the same relative ordering of W_{R} for the different function shapes. The same insight applies across function types; see, for example, Case 3 across the four types of functions.

Thus, the type and shape of the cost function matters, as we expected. For many of the cases shown in Figure 5.5, the price for mistakenly assuming constant MCs is a non-negligible increase in expected total cost, up to nearly 30% for convex TC functions. The nine cases shown in Figure 5.5 are good representations of the results for the 28 cases that we examined for Procedure $\mathcal{M}_{\mathrm{R}}$.

For the sake of brevity, we do not show results for Procedure $\mathcal{M}_{\mathrm{NR}}$; however, the results are similar with a few minor differences. In five of the 28 cases, W_{NR} actually *decreases* with increasing deviance of the concave TC function from linear. The same behavior is evident for four of the 28 cases under concave-convex TC functions and one case under convex-concave TC functions. The exceptions mentioned above all occur when $b < 10$, i.e., when the observation budget is relatively low. Additionally, the magnitudes of W_{NR} are similar to those of W_{R}, except in a few cases with low b.

5.2.1.2 Effects on Selected Non-optimal Procedures

The general observations that we were able to make in §5.2.1.1 do not necessarily apply when examining the performance of the existing non-optimal procedures. The situation now is that either the experimenter does not have the capability to identify the optimal

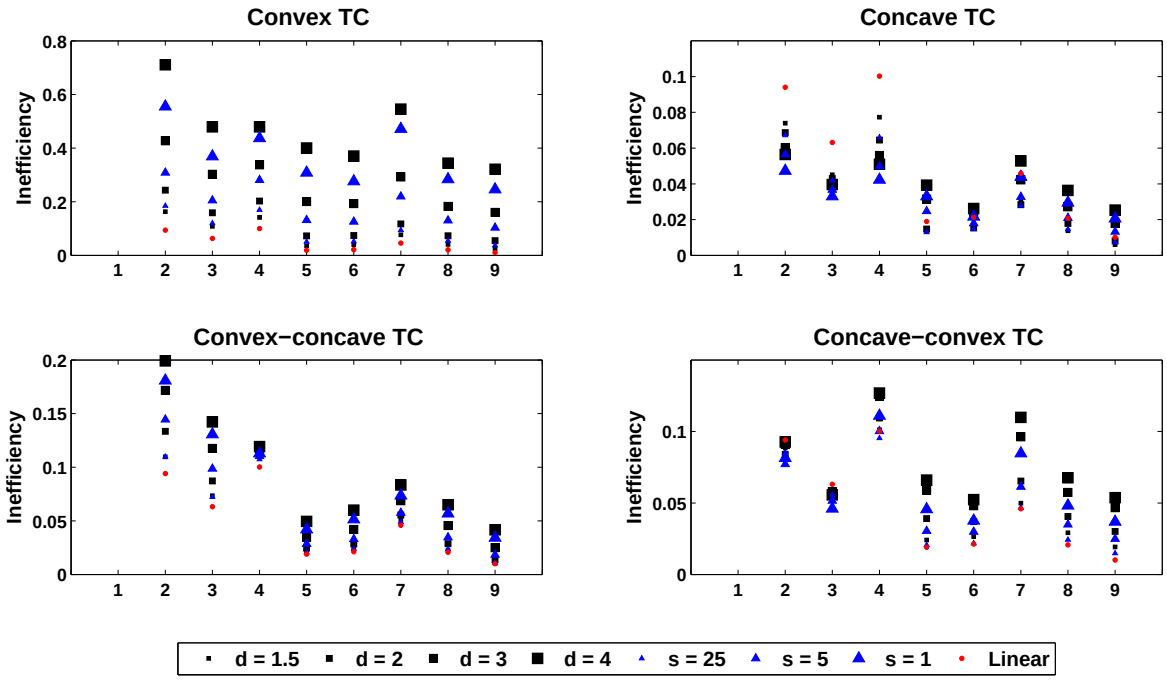

Figure 5.6: Procedure $\mathcal{M}_{\mathrm{BG}}$ Inefficiency for Selected Cases with $\theta = 2.4$

randomized procedure, or he prefers to use a non-optimal procedure. For our discussion here, we examine two procedures: Procedure $\mathcal{M}_{\mathrm{BG}}$ and Procedure $\mathcal{M}_{\mathrm{RA}'}$. We choose the latter since Procedures $\mathcal{M}_{\mathrm{RA}}$, $\mathcal{M}_{\mathrm{C}'}$, and $\mathcal{M}_{\mathrm{BK}}$ are all special cases of Procedure $\mathcal{M}_{\mathrm{RA}'}$.

Procedure $\mathcal{M}_{\mathrm{BG}}$ Figure 5.6 shows results for eight of the nine cases in Table 5.1. We do not include the results for Case 1 because its very high inefficiency skews the vertical axis limits. The plots are similar to those in Figure 5.5, except that W_{BG} is also shown for the linear TC function as a reference. For Procedure $\mathcal{M}_{\mathrm{R}}$ this was unnecessary since $W_{\mathrm{R}} = W_{\mathrm{RC}} = 0$ when the TC function is linear.

We make the following observations.

1. Convex functions usually have the largest impact on W_{BG}. The only exception that we saw in our 28 cases was when $b = 1$. In that special case, Procedure $\mathcal{M}_{\mathrm{RC}}$ is identical to all non-randomized procedures since taking exactly one observation achieves $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS}) = P^\star$.

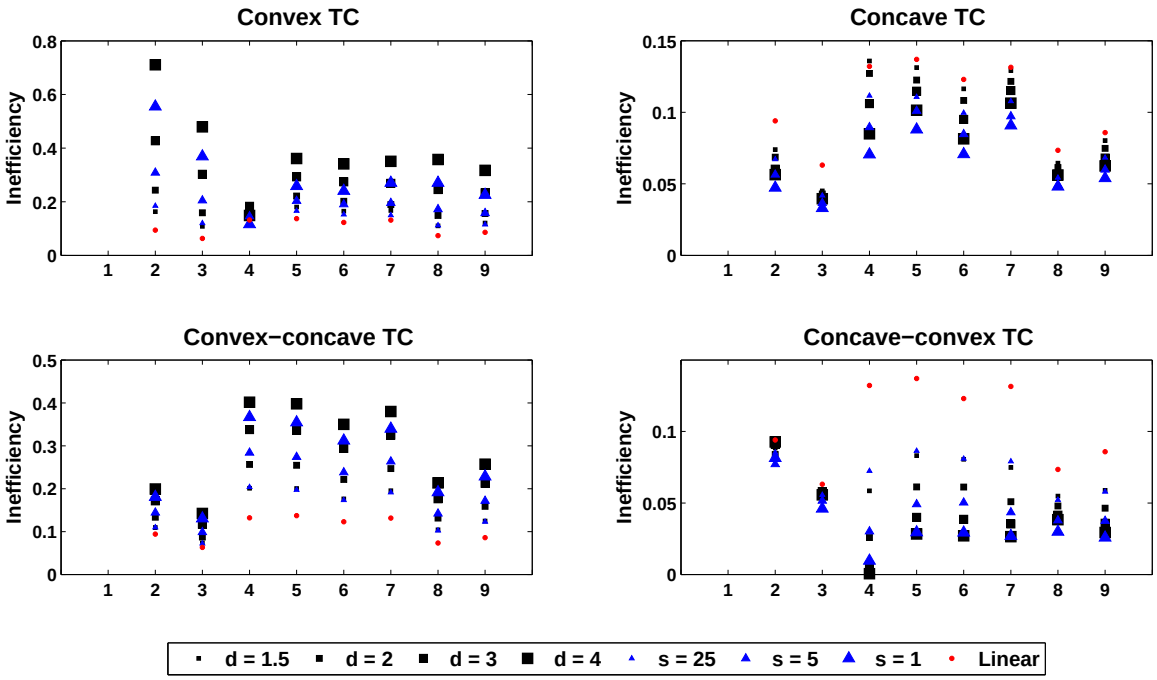

Figure 5.7: Procedure $\mathcal{M}_{\mathrm{RA}'}$ Inefficiency for Selected Cases with $\theta = 2.4$

2. For convex functions, increasing the deviance of the TC function from linear tends to increase the inefficiency of Procedure $\mathcal{M}_{\mathrm{BG}}$; such was the case for all 28 cases we examined. This implies that using Procedure $\mathcal{M}_{\mathrm{BG}}$ instead of the optimal randomized procedure when total costs are convex leads to an even greater inefficiency than when trying to minimize observations. The same is true for convex-concave mixed TC functions with only two exceptions in the 28 we examined.

3. For concave and concave-convex mixed TC functions, we see cases in which W_{BG} is increasing, decreasing, or neither as deviation from linear increases.

4. The particular *shape* of the total cost function matters as it does for Procedure $\mathcal{M}_{\mathrm{R}}$.

Procedure $\mathcal{M}_{\mathrm{RA}'}$ Figure 5.7 shows results for eight of the nine cases in Table 5.1. As for Procedure $\mathcal{M}_{\mathrm{BG}}$, we do not include the results for Case 1. The results are similar to those for Procedure $\mathcal{M}_{\mathrm{BG}}$.

We make the following observations concerning Procedure $\mathcal{M}_{\mathrm{RA}'}$.

1. Convex-concave mixed functions often have the largest impact on $W_{\mathrm{RA}'}$, as is the case for Cases 4–7 in the figure. This is in contrast to the other procedures we have analyzed for which convex functions almost always have the largest impact.

2. For convex functions, increasing the deviance of the TC function from linear tends to increase the inefficiency of Procedure $\mathcal{M}_{\mathrm{RA}'}$; such was the case for all but one of the 28 cases we examined. The same is true for convex-concave mixed TC functions with only one exception in the 28 we examined.

3. For concave functions, increasing the deviance of the TC function from linear actually *decreases* the inefficiency of Procedure $\mathcal{M}_{\mathrm{RA}'}$ for 19 of the 28 cases we examined. The same is true for concave-convex mixed TC functions for 15 of the 28 cases we examined.

4. The particular *shape* of the total cost function matters.

The results in this subsection again show that the costs associated with using non-optimal procedures can be significant. The trends associated with the different types and shapes of TC functions are often specific to the particular TC function and procedure. The only general result that applies across all non-optimal procedures is that using those procedures incurs unnecessary additional costs.

5.2.2 Effects of Periodic Costs

We first analyze the effect of periodic costs on Procedure $\mathcal{M}_{\mathrm{R}}$ performance. Figure 5.8 shows results for the nine cases in Table 5.1 for $B = 2, 5$ (the smallest and largest batch sizes we examined). A larger marker implies a larger value of ρ. As expected, W_{R} increases within each case as ρ increases. The same is true for all batch sizes and all 28 cases we examined. Figure 5.8 also indicates that the larger batch sizes have a greater effect on W_{R}; however, this is not necessarily the case in general.

In Figure 5.9, we plot results for $\rho = 0.5, 1$, and all four possible batch sizes. The figure shows that increasing the batch size does not necessarily lead to an increase in inefficiency. This result is likely due to the discrete nature of the periodic batch costs and when those

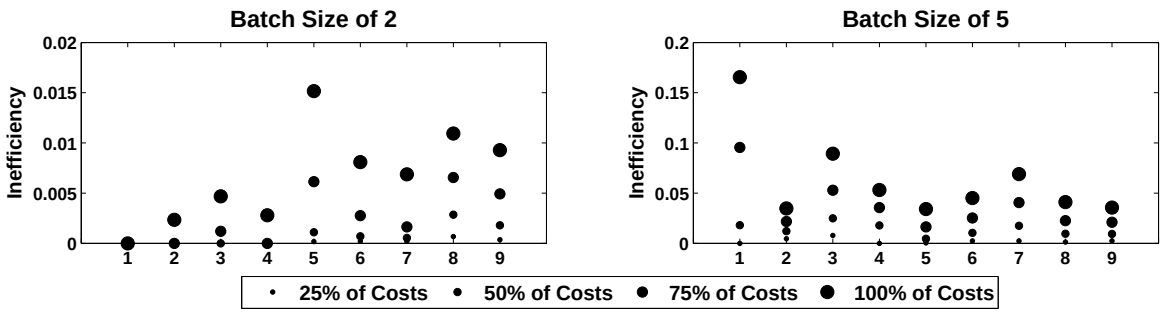

Figure 5.8: Procedure $\mathcal{M}_{\mathrm{R}}$ Inefficiency for Selected Cases with $B = 2, 5$ and $\theta = 2.4$

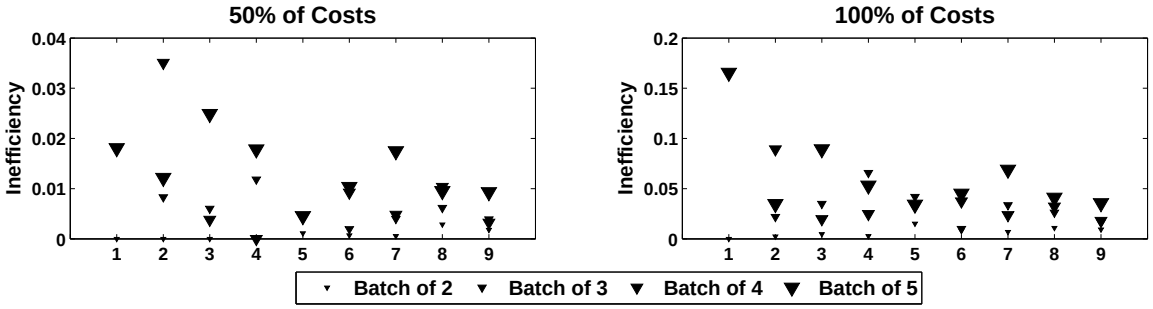

Figure 5.9: Procedure $\mathcal{M}_{\mathrm{R}}$ Inefficiency for Selected Cases with $\rho = 0.5, 1$ and $\theta = 2.4$

costs are incurred with respect to the stopping points in Procedure $\mathcal{M}_{\mathrm{R}}$ and the final budget b. We must also keep in mind that W_{R} is calculated with respect to the performance of the optimal Procedure $\mathcal{M}_{\mathrm{RC}}$. If Procedure $\mathcal{M}_{\mathrm{RC}}$ cannot achieve the same magnitude of improvement for a larger batch size as it can for a smaller batch size, then the metric W_{R} may be lower for the larger batch size.

For brevity, we do not show the results for Procedures $\mathcal{M}_{\mathrm{BG}}$ and $\mathcal{M}_{\mathrm{RA}'}$. For neither procedure does inefficiency increase for all cases as ρ increases, as it did for Procedure $\mathcal{M}_{\mathrm{R}}$; however, at larger budgets, the inefficiency of both procedures does tend to increase as ρ increases. Similar to Procedure $\mathcal{M}_{\mathrm{R}}$, increasing the batch size does not necessarily lead to an increase in inefficiency. Finally, W_{BG} and $W_{\mathrm{RA}'}$ are typically an order-of-magnitude larger (tenfold increase) for batches of size two than W_{R}. This, of course, translates to much higher increases in expected total cost.

5.2.3 Batch Pricing

All of the results we have discussed thus far give us insights into the penalty paid for not using Procedure $\mathcal{M}_{\text{RC}}$ when total costs are nonlinear. But for batch costs, we can gain some practical insight about how much we should be willing to pay for batches of a particular size. Armed with this information, decision-makers can negotiate beneficial batch prices or decide not to purchase batches at all.

Consider an experiment for which all costs are incurred through batch costs (i.e., $\rho = 1$). The experimenter has the choice to pay c_B for batches of observations of size B or to pay for each observation individually at cost c. The choice of policy must not change during an experiment, i.e., all purchases must be in batches of the same size or they must all be purchased individually. Initially, we let $c_B = B\,c$; in other words, the cost of the batch is equivalent to purchasing each observation individually. We also assume that B is a divisor of the budget b. The latter assumption is reasonable since an experimenter is likely to set a budget that is a multiple of the batch size, and is necessary to avoid penalizing batch purchases for observations that cannot be used. Let C^{B} be a random variable representing the total cost under the batch cost function and C^{L} be the random variable representing the total cost under a linear cost function.

We use optimization to identify the optimal Procedures $\mathcal{M}_{\text{RC}}$ for both cost functions and calculate $\mathsf{E}_{\text{SC}}[C^{\text{B}}_{\text{RC}}]$ and $\mathsf{E}_{\text{SC}}[C^{\text{L}}_{\text{RC}}]$. Consider a new metric D_B which we call the *break-even batch discount rate*, for batch size B. We calculate it as follows:

$$D_B \equiv 1 - \frac{\mathsf{E}_{\text{SC}}[C^{\text{L}}_{\text{RC}}]}{\mathsf{E}_{\text{SC}}[C^{\text{B}}_{\text{RC}}]}. \tag{5.13}$$

This is the minimum discount at which we should be willing to purchase batches of size B. In other words, we would be willing to pay at most $c'_B = (1 - D_B)\,c_B$ for each batch of size B.

To see that this is true, we recall that $\mathsf{E}_{\text{SC}}[C_{\text{RC}}]$ is the value of the objective function (5.3) at the solution to the mathematical program $\mathcal{LP}_{\mathcal{C}}$ in the SC. For the linear total cost

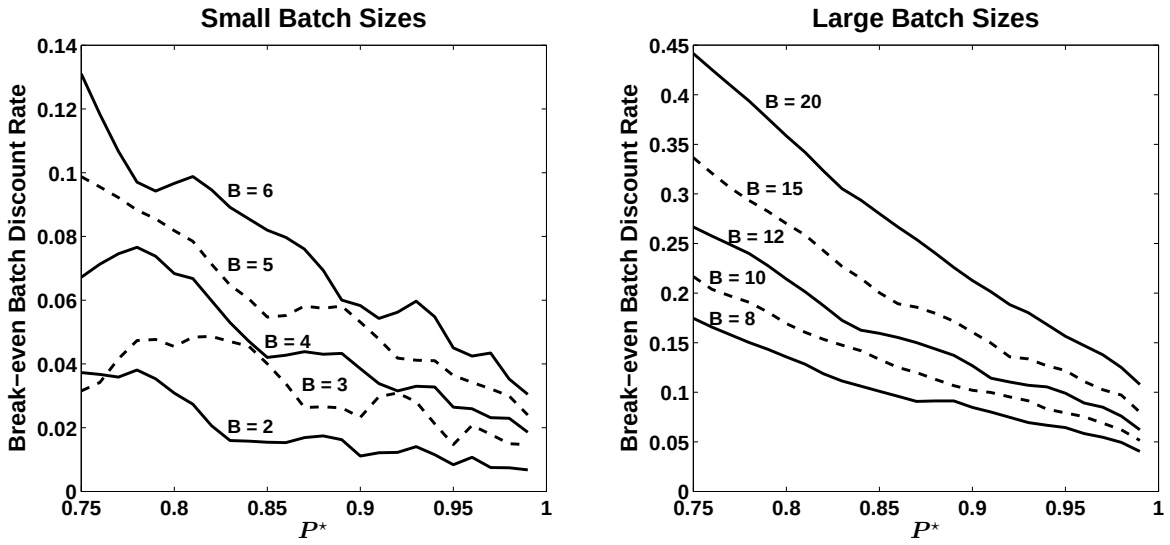

Figure 5.10: D_B versus $P^\star$ for $k = 3$, $\theta^\star = 1.8$, $b = 120$ and Selected Batch Sizes

function, $c_{\boldsymbol{\eta}'} = c$ for all $\boldsymbol{\eta}'$ in (5.2). For the batch cost function,

$$c_{\boldsymbol{\eta}'} = \begin{cases} B\,c & \text{if } B \text{ is a divisor of } \ m-1, \\ 0 & \text{otherwise.} \end{cases}$$

We know that the optimal solution to an LP is invariant to linear transformations of the objective function. Therefore, if our original value of the objective function for batches is $\mathsf{E}_{\mathrm{SC}}[C^{\mathrm{B}}_{\mathrm{RC}}]$, then if we multiply our objective function in $\mathcal{LP}_{\mathcal{C}}$ by $(1 - D_B)$, the new value of our objective function will be

$$\begin{aligned} &= (1 - D_B)\,\mathsf{E}_{\mathrm{SC}}[C^{\mathrm{B}}_{\mathrm{RC}}], \\ &= \left(1 - \left[1 - \frac{\mathsf{E}_{\mathrm{SC}}[C^{\mathrm{L}}_{\mathrm{RC}}]}{\mathsf{E}_{\mathrm{SC}}[C^{\mathrm{B}}_{\mathrm{RC}}]}\right]\right) \mathsf{E}_{\mathrm{SC}}[C^{\mathrm{B}}_{\mathrm{RC}}], \\ &= \mathsf{E}_{\mathrm{SC}}[C^{\mathrm{L}}_{\mathrm{RC}}], \end{aligned}$$

which is what we were trying to show.

Consider an example with $k = 3$, $b = 120$, and $\theta^\star = 1.8$. The value of b is chosen to maximize the number of batch sizes we can consider for purposes of the example (i.e., the possible divisors of b). The values of k and $\theta^\star$ were chosen so that the optimal solution for the largest $P^\star$ we will consider, $P^\star = 0.99$, will require most of the available observations. Figure 5.10 shows plots of D_B as a function of $P^\star$ for $B \in \{2, 3, 4, 5, 6\}$ on the left and

$B \in \{8, 10, 12, 15, 20\}$ on the right. As expected, the larger the batch size, the larger the discount D_B required to make batching an acceptable option. There are a couple of exceptions. At $P^\star = 0.75$ and $P^\star = 0.76$, D_B is actually larger for $B = 2$ than for $B = 3$. This result is likely due to the fact that lower $P^\star$-values can be achieved with relatively low numbers of observations. For these, it may be that stops at odd multiples of 3 give $\mathsf{P}(\text{CS})$ and $\mathsf{E}[C]$ flexibility when working with $B = 3$ but penalize $B = 2$ by requiring the purchase of an extra observation.

We have shown that we can leverage our mathematical programs to determine discount rates for batch price negotiation or to determine whether or not to purchase batches if the prices are already set. Curves such as those in Figure 5.10 can be developed for any MSP. Although we did not do so here, more complicated batching strategies can be examined by manipulating the constraints in $\mathcal{LP}_\mathcal{C}$ and/or adding variables. Our contribution is the development of mathematical programming formulations of MSPs that allow us to answer real world, cost-related questions that have heretofore been unanswerable.

5.2.4 Shadow Prices

Our mathematical programming formulation $\mathcal{LP}_\mathcal{C}$ provides another tool that can be quite useful for cost-related analysis. Any LP has a corresponding dual formulation that is also an LP. Each of the variables in the dual formulation (i.e., the dual variables) correspond to a constraint in the primal (i.e., original) formulation. A detailed discussion of the dual formulation and its properties would be too involved to include here; the reader may refer to any introductory linear optimization text, such as Bertsimas and Tsitsiklis (1997), for more information. Most optimization software packages determine the dual formulation automatically, and return the values of the dual variables (i.e., the dual solution) with the primal solution.

The values of the dual variables at the optimal solution to the LP can provide additional insight into cost. Let λ be the dual variable associated with the $\mathsf{P}(\text{CS})$ constraint, (4.11), in the LP. Let $\mu_{\boldsymbol{\eta}'}$ be the dual variable associated with the flow conservation (or initialization)

constraint for network node $\boldsymbol{\eta}'$. We know that since the $\mathsf{P}(\text{CS})$ constraint is a greater-than-or-equal-to constraint in a minimization problem, $\lambda \leq 0$ in the dual maximization problem. The $\mu_{\boldsymbol{\eta}'}$'s are free (unconstrained) variables in the dual problem. The dual variables are often called *shadow prices* of the resource constraints. For example, we can interpret λ as the instantaneous price that we should be willing to pay for an increase in one unit of $\mathsf{P}(\text{CS})$.

As an example, consider an MSP with $k = 3$, $P^\star = 0.9$, $\theta^\star = 3$, and $b = 30$, and linear costs such that each observation costs \$100. Then $\mathsf{E}_{\text{SC}}[C_{\text{RC}}] = \$622.37$ and $\mathsf{P}_{\text{SC}}(\text{CS}) = 0.9$. At this solution, $\lambda = 3967.58$. Since it is unrealistic to expect a unit increase in $\mathsf{P}_{\text{SC}}(\text{CS})$, we instead consider what the expected total cost would be if we required a 0.01 increase in $P^\star$, i.e., $P^\star = 0.91$. Our shadow price tells us that our expected total cost will increase \$39.68 to \$662.05. If we change the formulation slightly to $P^\star = 0.91$, we indeed see that $\mathsf{E}_{\text{SC}}[C_{\text{RC}}] = \$662.04$. This example is slightly misleading though. The interpretation of shadow price that we are using only applies for values 'close' to the original $P^\star$. And 'closeness' may vary depending upon the particular solution. For example, if we consider the original MSP, but increase $P^\star$ to 0.98, we get $\mathsf{E}_{\text{SC}}[C_{\text{RC}}] = \$1148.99$, $\mathsf{P}_{\text{SC}}(\text{CS}) = 0.98$, and $\lambda = 18{,}529.44$. This might imply that our expected total cost will increase \$185.29 to \$1334.28 if we increase our $P^\star$ requirement by 0.01. However, changing $P^\star$ to 0.99 and optimizing, we get $\mathsf{E}_{\text{SC}}[C_{\text{RC}}] = \$1517.99$. Thus, we must be careful when interpreting the implications of the dual variable, particularly when $P^\star$ is very large.

Interpretation of the dual variables associated with the equality constraints $\{\mu_{\boldsymbol{\eta}'}\}$ is less straightforward. It is difficult to conceptualize changes in the right hand side of our flow conservation constraints. By changing the right hand side, we either allow inbound probability flow to be greater than the outbound flow, or vice versa. Such changes violate the assumptions upon which our MSP formulation is built. Consider the original formulation but with the initialization constraint multiplied by –1. That puts the initialization constraint in the same form as the flow conservation constraints, with inbound flow having positive coefficients and outbound flow having negative coefficients. Thus, an increase to the right hand side would mean that there is less flow out of the node than flowed into it.

For all of the cases that we observed, $\mu_{\boldsymbol{\eta}'} \geq 0$. Thus, decreasing the right hand side (i.e., allowing more probability flow out of the node than flowed in), has the effect of decreasing the expected total cost. This makes intuitive sense. Allowing more outbound flow than flowed in actually creates additional probability at that node, making it a source node. In the simplest case, that node could send all of the additional probability directly to the sink, moving the solution closer to the required $P^\star$ without having increased total cost at all.

5.3 Summary

The main contributions in this chapter are LP and MIP formulations that, when solved, provide the optimal randomized and non-randomized procedures, respectively, for any given cost function. These formulations of general MSPs are built upon reasonable assumptions with respect to the observation costs, namely, that the marginal cost of taking an additional observation at $\boldsymbol{\eta}'$ can be specified and is the same for all permutations of $\boldsymbol{\eta}'$. Ours is the first methodology that integrates variable observation costs for MSPs.

Furthermore, using these formulations, we have been able to show that the type and shape of the total cost functions has an impact, often significant, on a procedure's efficiency with respect to the optimal solution. We examined a robust set of possible cost functions, including functions affected by periodic observation costs. In some cases, there were general insights that applied across function types. For example, the additional cost incurred by failing to use the optimal procedure is greatly amplified under convex (and sometimes convex-concave mixed) TC functions. Conversely, concave (and sometimes concave-convex mixed) TC functions tend to mitigate the additional costs incurred for using a non-optimal procedure. In many cases, procedure inefficiency due to ignoring the underlying cost function is unique to a particular total cost function. In all cases, incorrectly assuming that total costs are linear leads to unnecessary additional costs in expectation.

Finally, we provided two additional tools for addressing particular cost-related issues. One methodology allows the decision-maker to determine discount rates for batch price negotiation or to determine whether or not to purchase batches if the prices are already set. The other tool uses the information already embedded in the LP concerning the dual

variables to estimate the cost of increasing the $P^{\star}$ requirement.

CHAPTER VI

PROCEDURE COMPARISONS

In this chapter, we examine the performance of particular MSP procedures. We have already shown some comparative results in Chapter 3 and comparisons based upon expected total cost in Chapter 5. Here, we leverage our previous contributions to analyze the procedures more thoroughly, in order to develop deeper insights into performance in terms of not only the expected number of observations, but also other important metrics as well.

In §6.1, we demonstrate that for large b coupled with non-trivial $P^\star$ requirements, $\mathsf{E}_{\mathrm{SC}}[N]$ for Procedure $\mathcal{M}_{\mathrm{NR}}$ gets very close to that of Procedure $\mathcal{M}_{\mathrm{R}}$. Next, in §6.2, we conduct a detailed comparison of existing procedures. In §6.3, we look more closely at the performance of Procedure $\mathcal{M}_{\mathrm{RA}}$ and show its relationship to the classic gambler's ruin problem. We summarize our results in §6.4.

6.1 Performance Comparison of Procedures $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{NR}}$

In this section, we examine the performance (in terms of the expected number of observations) of optimal randomized Procedure $\mathcal{M}_{\mathrm{R}}$ and optimal non-randomized Procedure $\mathcal{M}_{\mathrm{NR}}$. Figure 6.1 is a plot of the expected number of observations for these procedures with $k = 2$, $\theta^\star = 2$ (left), and $\theta^\star = 3$ (right), as a function of the observation budget, b, for selected $P^\star$ requirements. For all but one particular combination of $P^\star$ and $\theta^\star$, the expected number of observations for Procedure $\mathcal{M}_{\mathrm{NR}}$ very quickly approaches, as a function of b, the expected number of observations for Procedure $\mathcal{M}_{\mathrm{R}}$.

For $P^\star = 0.75$ and $\theta^\star = 2$, the expected numbers of observations for the optimal non-randomized and randomized procedures do not converge to each other, but remain equidistant as b increases. In this case — not uncommon for lower values of $P^\star$ coupled with larger $\theta^\star$ — the required $P^\star$ is achieved very 'early' in the nodal network. At $b = 5$, Procedure $\mathcal{M}_{\mathrm{NR}}$ achieves a $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ of 0.7737 with 3.086 expected observations. Procedure $\mathcal{M}_{\mathrm{R}}$ achieves a $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ of 0.75 with 2.625 expected observations. Increasing b does not make

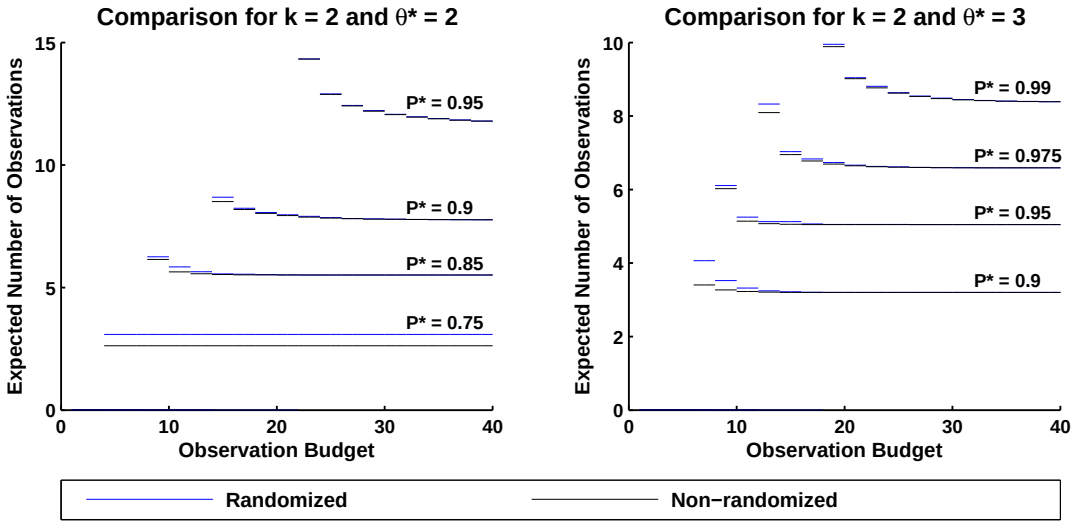

Figure 6.1: Randomized and Non-randomized Procedures for $k = 2$ and $\theta^\star = 2, 3$

available any nodes that can be taken in combination with existing nodes to improve upon the optimal value of either procedure. We discuss the intuition behind these observations later in the section.

Figure 6.1 also confirms the result peculiar to $k = 2$ that we pointed out in §4.1.2.3. When strong curtailment is used, the optimal expected number of observations for an even value of b is equivalent to that of $b - 1$.

Figure 6.2 shows similar results for $k = 3$. The plots for $P^\star = 0.75$ and $\theta^\star = 3$ (right chart) demonstrate that the non-convergent "delta" between optimal values may be very small. Additionally, both figures suggest that, for a given combination of k, $\theta^\star$, and $P^\star$, both procedures approach a nearly constant optimal expected number of observations as b increases. For such parameter combinations, increases to the budget reach a point beyond which the availability of additional observations has little impact on the minimum achievable expected number of observations.

We make the above observations without formal proofs; however, we can provide some insight using the standard multinomial probability distribution. Consider the multinomial distribution for three alternatives in probability configuration $\boldsymbol{p} = (0.5, 0.25, 0.25)$ after 30

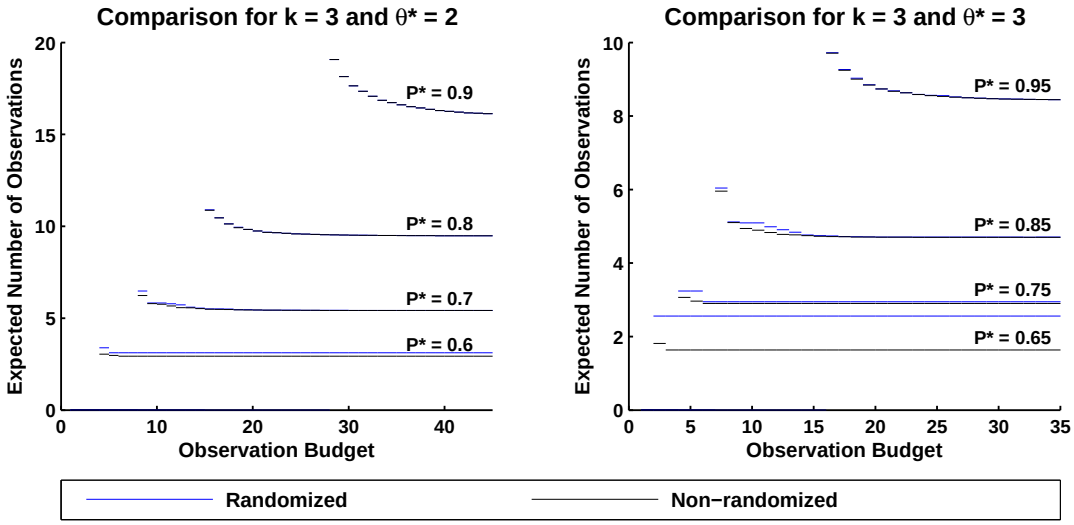

Figure 6.2: Randomized and Non-randomized Procedures for $k = 3$ and $\theta^\star = 2, 3$

observations. There are 496 possible outcomes, in terms of cumulative success vectors; but, due to symmetry, we are interested in the 91 unique left-lexicographic outcomes that the original 496 outcomes represent. The contribution of each left-lexicographic outcome (and its permutations) towards $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ ranges from 9.31×10^{-10} to 0.0567. The contribution towards the expected number of observations ranges from 2.79×10^{-8} to 1.7221. Figure 6.3 is a column graph of the contributed $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ (left) and the contributed expected number of observations (right) of each node sorted by $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$, so that the nodal order for the left graph is the same as that of the right graph. Figure 6.3 shows us that, at a given number of observations, nodes with lower $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ *generally* (but not always) also contribute the least towards the expected number of observations.

Now consider a non-randomized procedure with $b \geq 30$. In order to minimize the expected number of observations, the optimization can 'choose' between many potential stopping nodes to achieve $P^\star$ while minimizing the expected number of observations. Since nodes with a lower contribution towards $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ generally have a lower contribution towards the expected number of observations, it may choose from numerous sample paths with incrementally small contributions to $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ to create an optimal solution. Such a

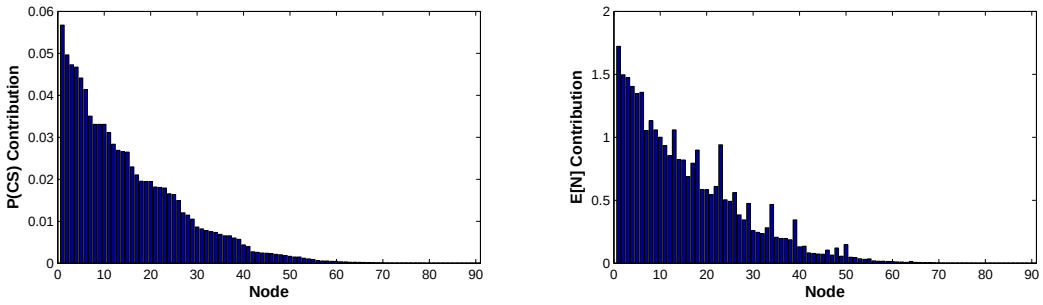

Figure 6.3: Multinomial Contribution to $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ and $\mathsf{E}_{\mathrm{SC}}[N]$ at the 30th Observation

large number of sample paths, with very small probabilities, gives the optimization flexibility for achieving $P^\star$ while pushing the expected number of observations very close to the optimal value for the randomized procedure.

In contrast, for the same multinomial distribution with $\boldsymbol{p} = (0.5, 0.25, 0.25)$ after only five observations, the $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ contributions due to the five possible left-lexicographic outcomes, {(2,2,1), (3,1,1), (3,2,0), (4,1,0), (5,0,0)}, are {0.1172, 0.1563, 0.1563, 0.1563, 0.0313}, respectively. Here the optimization does not have much flexibility. While increasing b may seem to provide such flexibility by lowering the probabilities at each added node, the optimization must still choose intermediate points through which the probability must flow to get to the further nodes in the network. These add relatively large expected numbers of observations (by going further in the network), making the effort costly if $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ has already been achieved.

Naturally, the above discussion is a simplification. In reality, any stop before 30 observations will affect the contributions at 30 observations by reducing the number of paths to (i.e., the multinomial coefficient of) particular nodes. Also, choices of stopping nodes across multiple numbers of observations make both the $\mathsf{P}(\mathrm{CS})$ and the expected number of observations complex functions of the choices themselves. Nonetheless, our example does demonstrate that the rapidly decreasing probabilities of reaching nodes as b increases give the optimization additional flexibility to choose between stopping nodes, provided $P^\star$ is not achieved very early in the network.

6.2 Performance Comparison of Non-optimal Procedures

In this section, we compare existing procedures from the literature by evaluating their performances (in terms of expected numbers of observations) relative to the optimal randomized Procedure $\mathcal{M}_{\mathrm{R}}$. In our comparisons, we do not include Procedure $\mathcal{M}_{\mathrm{NR}}$ for three reasons:

1. As we have shown, the optimal expected number of observations for Procedure $\mathcal{M}_{\mathrm{NR}}$ is very close to that of Procedure $\mathcal{M}_{\mathrm{R}}$ in most cases, particularly for large b.

2. The maximum size of the MIPs that we are able to solve is much smaller than the maximum size we are able to solve for the LPs. Considering only Procedure $\mathcal{M}_{\mathrm{R}}$ allows us to make more meaningful comparisons across a larger set of problems than we could if we considered Procedure $\mathcal{M}_{\mathrm{NR}}$.

3. Most importantly, Procedure $\mathcal{M}_{\mathrm{R}}$ is optimal. Should the decision-maker have the capability to determine the optimal randomized and non-randomized procedures, there would be no need to consider the optimal non-randomized procedure at all.

In §6.2.1, we identify the procedures that we will include in our comparisons. We then describe, in §6.2.2, the metrics we will use for our comparisons in §§6.2.3–6.2.5.

6.2.1 Procedures

All of the procedures that we consider are either single-stage or bounded sequential procedures, i.e., procedures that have a finite limit to the maximum number of observations. Bounded sequential procedures, unlike our optimal procedures, do not require the specification of a budget b; rather, their procedure parameters are chosen in order to satisfy the $P^\star$ requirement while minimizing $\mathsf{E}_{\mathrm{SC}}[N]$. Thus, for our comparisons, we will choose a b for each problem and then search only over the subset of possible procedure parameters that ensures that the maximum number of observations is less than b. For some problems, this may result in a particular procedure not being able to achieve $P^\star$ at all, depending upon our choice of b.

The following are the seven procedures that we will examine.

1. **Procedure $\mathcal{M}_{\mathrm{BEM}}$:** The single-stage procedure for which the truncation parameter $n_{\mathrm{BEM}} \leq b$.

2. **Procedure $\mathcal{M}_{\mathrm{BK}}$:** The bounded sequential procedure for which the truncation parameter $n_{\mathrm{BK}} \leq b$. We include updated tables for this procedure in Appendix A.1.

3. **Procedure $\mathcal{M}_{\mathrm{C}'}$:** A modified version of Chen's (1988a) inverse sampling Procedure $\mathcal{M}_{\mathrm{C}}$. In his original paper, Chen states that the strong curtailment stopping rule (2.1) of Procedure $\mathcal{M}_{\mathrm{BK}}$ could be used to reduce the expected number of observations for his procedure without affecting $\mathsf{P}(\mathrm{CS})$, but he does not implement the change. Therefore, we create Procedure $\mathcal{M}_{\mathrm{C}'}$ by modifying stopping rule (2.4) to include the strong curtailment stopping rule. In Appendix A.2, we tabulate appropriate (n, t)-pairs for the new procedure for selected choices of k, $\theta^\star$, and $P^\star$. Choices for this procedure include all parameter combinations with the truncation parameter $n_{\mathrm{C}'} \leq b$ and the inverse sampling parameter $t \leq n_{\mathrm{C}'}$.

4. **Procedure $\mathcal{M}_{\mathrm{RA}}$:** The bounded sequential procedure that includes all parameter combinations with the inverse sampling parameter $t \leq (b-1)/k + 1$ (which ensures that the procedure will stop at or before the budget b) and difference parameter $r \leq t$. We include updated tables for this procedure in Appendix A.3.

5. **Procedure $\mathcal{M}_{\mathrm{RA}'}$:** The bounded sequential procedure that includes all parameter combinations with the truncation parameter $n_{\mathrm{RA}'} \leq b$, inverse sampling parameter $t \leq n_{\mathrm{RA}'}/2$ (by strong curtailment), and difference parameter $r \leq t$. We include updated tables for this procedure in Appendix A.4.

6. **Procedure $\mathcal{M}_{\mathrm{BG}}$:** The bounded sequential procedure with truncation parameter $n_{\mathrm{BG}} \leq b$.

7. **Procedure $\mathcal{M}_{\mathrm{R}}$:** The optimal randomized constrained sequential procedure under budget b.

6.2.2 Metrics

In this section, we briefly describe some of the measures that we will use in comparing procedure performance. Some of these metrics align with those that we have already developed in Chapter 5.

6.2.2.1 Expected Number of Observations

The most common measure in the literature is $\mathsf{E}_{\mathrm{SC}}[N]$. Naturally, $\mathsf{E}_{\mathrm{SC}}[N]$ is quite important to the decision-maker when considering a procedure to use, since his primary goal is normally the minimization of this metric. In some cases, the decision-maker may be concerned with minimizing the maximum possible number of observations taken; however, we assume that in setting a budget, the decision-maker is more interested in the former than the latter. A decision-maker might also be interested in the expected number of observations in the EPC. This worst-case expectation, $\mathsf{E}_{\mathrm{EPC}}[N]$, gives the decision-maker insight into the possible impact if the alternatives are in a configuration for which it is most difficult to determine which is best.

6.2.2.2 Procedure Inefficiency Metric

We may also be interested in the deviation of procedure performance from what the optimal procedure can achieve, thereby using Procedure $\mathcal{M}_{\mathrm{R}}$ as a benchmark against which we compare other procedures as we did for total cost in Chapter 5. In order to facilitate an analysis across different problems with widely varying budgets, we use the same *procedure inefficiency* metric, W_{J}, for the performance of general procedure $\mathcal{M}_{\mathrm{J}}$, as we did for cost. The metric, redefined for expected number of observations, is:

$$W_{\mathrm{J}} \equiv \frac{\mathsf{E}_{\mathrm{SC}}[N_{\mathrm{J}}] - \mathsf{E}_{\mathrm{SC}}[N_{\mathrm{R}}]}{\mathsf{E}_{\mathrm{SC}}[N_{\mathrm{R}}]} = \frac{\mathsf{E}_{\mathrm{SC}}[N_{\mathrm{J}}]}{\mathsf{E}_{\mathrm{SC}}[N_{\mathrm{R}}]} - 1, \tag{6.1}$$

where $\mathsf{E}_{\mathrm{SC}}[N_{\mathrm{J}}]$ is the expected number of observations using general procedure $\mathcal{M}_{\mathrm{J}}$ in the SC, and $\mathsf{E}_{\mathrm{SC}}[N_{\mathrm{R}}]$ is the expected number of observations using Procedure $\mathcal{M}_{\mathrm{R}}$ in the SC. Recall that we can think of procedure inefficiency as the fractional increase in expected total observations for using general procedure $\mathcal{M}_{\mathrm{J}}$ instead of the optimal Procedure $\mathcal{M}_{\mathrm{R}}$.

Often, we may want to evaluate procedure performance over a range of potential problems. For that purpose, we extend our procedure inefficiency metric heuristically by calculating the *mean procedure inefficiency*, $\overline{W}_{\mathrm{J}}$, for a range of $P^\star$-values, $P^\star_{\mathrm{J},0}, P^\star_{\mathrm{J},1}, \ldots, P^\star_{\mathrm{J},m_{\mathrm{J}}}$, where m_{J} is the total number of $P^\star$-values at which we calculate $\mathsf{E}_{\mathrm{SC}}[N_{\mathrm{J}}]$. Since we only calculate the performance at each $P^\star$ increment, $\mathsf{E}^i_{\mathrm{SC}}[N_{\mathrm{J}}]$ calculated at $P^\star_{\mathrm{J},i}$ must be the assumed expected observations for the entire half-open interval $(P^\star_{\mathrm{J},i-1}, P^\star_{\mathrm{J},i}]$. Let I be the overall probability interval of $P^\star$-values we are considering. The mean procedure inefficiency, $\overline{W}^I_{\mathrm{J}}$, for procedure $\mathcal{M}_{\mathrm{J}}$ over interval I can be calculated via Riemann sums over the distribution of the number of observations:

$$\overline{W}^I_{\mathrm{J}} \equiv \frac{\displaystyle\sum_{i=1}^{m_{\mathrm{J}}} \mathsf{E}^i_{\mathrm{SC}}[N_{\mathrm{J}}] \ \left(P^\star_{\mathrm{J},i} - P^\star_{\mathrm{J},i-1}\right)}{\displaystyle\sum_{i=1}^{m_{\mathrm{R}}} \mathsf{E}^i_{\mathrm{SC}}[N_{\mathrm{R}}] \ \left(P^\star_{\mathrm{R},i} - P^\star_{\mathrm{R},i-1}\right)} - 1, \tag{6.2}$$

where m_{R} is the total number of $P^\star$-values at which we calculate $\mathsf{E}_{\mathrm{SC}}[N_{\mathrm{R}}]$. Note that our definition does not require constant increment size, nor do we need to use the same increment sizes for both procedures. It does, however, require the same overall $P^\star$-interval I:

$$P^\star_{\mathrm{J},m_{\mathrm{J}}} = P^\star_{\mathrm{R},m_{\mathrm{R}}} \quad \text{and} \quad P^\star_{\mathrm{J},0} = P^\star_{\mathrm{R},0}.$$

Keep in mind that the metric is specific to a particular combination of k, $\theta^\star$, $P^\star$, and b.

We must be careful here when comparing procedures, since $\overline{W}^I_{\mathrm{J}}$ compares each procedure with the optimum, based upon the $P^\star$-domain of the procedure, i.e., the range from $1/k$ to the maximum achievable $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ for that procedure. Procedures $\mathcal{M}_{\mathrm{BEM}}$, $\mathcal{M}_{\mathrm{BK}}$, $\mathcal{M}_{\mathrm{C}'}$, $\mathcal{M}_{\mathrm{RA}'}$, $\mathcal{M}_{\mathrm{NR}}$, and $\mathcal{M}_{\mathrm{R}}$ have the same domain. On the other hand, procedures $\mathcal{M}_{\mathrm{RA}}$ and $\mathcal{M}_{\mathrm{BG}}$ may have different domains from each other and the remaining procedures. Using the mean procedure inefficiency metric fails to recognize that the domains of the procedures are different. If we want to compare two procedures, say procedures $\mathcal{M}_{\mathrm{J}}$ and $\mathcal{M}_{\mathrm{L}}$, over a common domain, we create a new metric, which we call the *mean relative procedure*

performance, defined as follows:

$$\overline{V}^{I}_{\mathrm{J,L}} \equiv \frac{\sum_{i=1}^{m_{\mathrm{J}}} \mathsf{E}^{i}_{\mathrm{SC}}[N_{\mathrm{J}}] \ \left(P^{\star}_{\mathrm{J},i} - P^{\star}_{\mathrm{J},i-1}\right)}{\sum_{i=1}^{m_{\mathrm{L}}} \mathsf{E}^{i}_{\mathrm{SC}}[N_{\mathrm{L}}] \ \left(P^{\star}_{\mathrm{L},i} - P^{\star}_{\mathrm{L},i-1}\right)} - 1, \tag{6.3}$$

where I is the intersection of the domains of procedures $\mathcal{M}_{\mathrm{J}}$ and $\mathcal{M}_{\mathrm{L}}$. A positive value indicates that procedure $\mathcal{M}_{\mathrm{L}}$ performs better than procedure $\mathcal{M}_{\mathrm{J}}$ over the interval of interest; a negative value indicates the opposite.

6.2.2.3 Distributional Metrics

Since we can enumerate all of the possible stopping vectors for any MSP procedure, we can develop algorithms to determine the probability of arriving and stopping at each possible stopping vector. All MSP procedures under a finite budget have a finite number of stopping points; therefore, we have *complete* information about the probability distribution (i.e., the discrete probability mass function) of the number of observations required by the procedure. With this information, we can also calculate metrics such as the median, mode, variance, and quantiles of the random variable N.

6.2.3 Performance Comparison

Comparing procedures for a particular problem of interest is straightforward; comparisons across numerous potential problems are not. If we choose a few combinations of $\theta^{\star}$, $P^{\star}$, and b for each k, we may not get a good snapshot of performance. A procedure may be best for some problems but not for others. In Appendix B, we include comparison tables for the 36 possible combinations of $k \in \{2, 3, 4\}$, $\theta^{\star} \in \{1.6, 2, 2.4, 3\}$, and $P^{\star} \in \{0.75, 0.9, 0.95\}$, with a single budget b for each. The rationale for our choice of b is included in the appendix. The tables show results for both $\mathsf{E}_{\mathrm{SC}}[N]$ and $\mathsf{E}_{\mathrm{EPC}}[N]$. For those 36 cases, Procedure $\mathcal{M}_{\mathrm{BG}}$ usually performs better, in terms of $\mathsf{E}_{\mathrm{SC}}[N]$, than Procedure $\mathcal{M}_{\mathrm{RA}'}$; however, Tables B.4 and B.12 in Appendix B show that this is not always the case. While certain trends may be evident across the tables, it is hard to draw any completely general conclusions, particularly since our choice of b will affect the results and the relative performance between the procedures.

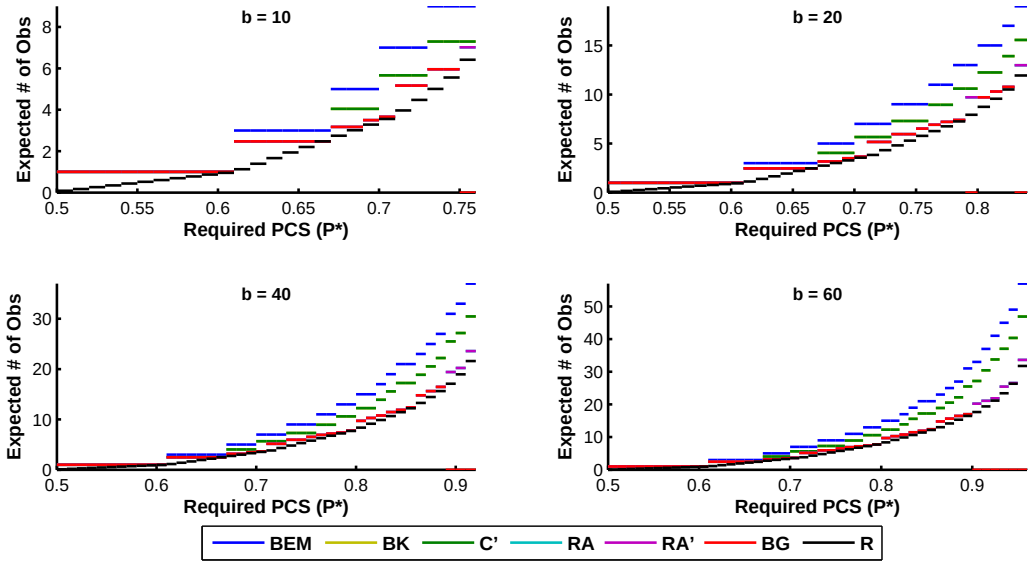

Figure 6.4: Procedure Comparison Plots for $k = 2$, $\theta^\star = 1.6$, $b \in \{10, 20, 40, 60\}$

We try to account for such issues by examining procedure performance across the 12 common combinations of $k \in \{2, 3, 4\}$ and $\theta^\star \in \{1.6, 2, 2.4, 3\}$. Within each combination, we choose four values for the budget b and then examine all possible $P^\star$ values between $1/k$ and 0.99, in increments of 0.01. We then plot our results to visualize relative procedure performance.

Figure 6.4 shows a series of four charts, one each for $b \in \{10, 20, 40, 60\}$, with $k = 2$, and $\theta^\star = 1.6$. The expected performance, $\mathsf{E}_{\mathrm{SC}}[N]$, of the seven procedures is plotted as a function of $P^\star$. Procedure $\mathcal{M}_{\mathrm{R}}$ is shown in black, always below the other procedures, as expected.

The plotted results for $k = 2$ illustrate some interesting findings. First, Procedures $\mathcal{M}_{\mathrm{BK}}$ and $\mathcal{M}_{\mathrm{C}'}$ perform identically when $k = 2$. As it turns out, for Procedure $\mathcal{M}_{\mathrm{C}'}$, stops due to t are identical to stops due to $n_{\mathrm{C}'}$ when $n_{\mathrm{C}'} = 2t - 1$. Choosing $n_{\mathrm{C}'} > 2t - 1$ or $t > \left\lfloor \frac{n_{\mathrm{C}'}+1}{2} \right\rfloor$, where $\lfloor x \rfloor$ is the floor function (i.e., rounds x down to the nearest integer), has no effect on the procedure. In other words, when $k = 2$, we can represent any two parameter Procedure $\mathcal{M}_{\mathrm{C}'}$ equivalently as the single parameter Procedure $\mathcal{M}_{\mathrm{BK}}$ with $n_{\mathrm{BK}} = \min\{n_{\mathrm{C}'}, 2t - 1\}$.

Similarly, Procedures $\mathcal{M}_{\mathrm{RA}}$ and $\mathcal{M}_{\mathrm{RA}'}$ also perform identically when $k = 2$. Again,

stops due to t are identical to stops due to $n_{\mathrm{RA}'}$ when $n_{\mathrm{RA}'} = 2t - 1$. Thus, when $k = 2$, Procedure $\mathcal{M}_{\mathrm{RA}'}$ with a particular $(n_{\mathrm{RA}'}, r', t')$-triplet is identical to Procedure $\mathcal{M}_{\mathrm{RA}}$ with a corresponding (r, t)-pair in which $r = r'$ and $t = \min\left\{t', \left\lfloor \frac{n_{\mathrm{RA}'}+1}{2} \right\rfloor\right\}$.

We also see that Procedure $\mathcal{M}_{\mathrm{BG}}$ has significant overlap with Procedures $\mathcal{M}_{\mathrm{RA}}$ and $\mathcal{M}_{\mathrm{RA}'}$ when $k = 2$. For Procedure $\mathcal{M}_{\mathrm{BG}}$, the parameter z_m in the stopping criteria (2.3) is based upon the differences between the alternative with the most successes and the other alternatives. When $k = 2$, there is only one difference to consider, which makes this parameter behave like the r parameter in Procedures $\mathcal{M}_{\mathrm{RA}}$ and $\mathcal{M}_{\mathrm{RA}'}$. To see this is true, let $r' \equiv \eta_{[2]m} - \eta_{[1]m}$. The Procedure $\mathcal{M}_{\mathrm{BG}}$ stopping condition becomes

$$z_m = \left(\frac{1}{\theta^\star}\right)^{r'} \leq \frac{1 - P^\star}{P^\star}. \tag{6.4}$$

Taking the natural logarithm and solving for r', we get

$$\begin{aligned} r' &\geq \frac{\ln P^\star - \ln(1 - P^\star)}{\ln \theta^\star} && (6.5) \\ &= \left\lceil \frac{\ln P^\star - \ln(1 - P^\star)}{\ln \theta^\star} \right\rceil, && (6.6) \end{aligned}$$

since r' must be an integer. As with the previous discussions, the parameter n_{BG} acts similarly to the t parameter when $k = 2$. The main aspect that makes Procedure $\mathcal{M}_{\mathrm{BG}}$ differ from Procedures $\mathcal{M}_{\mathrm{RA}}$ and $\mathcal{M}_{\mathrm{RA}'}$ in some cases is that parameter r can be chosen in the latter procedures to meet problem requirements, whereas r' is dictated by the problem parameters for Procedure $\mathcal{M}_{\mathrm{BG}}$.

In some cases, Procedure $\mathcal{M}_{\mathrm{BG}}$ cannot achieve a particular $P^\star$, but can achieve a $P^\star$ that is higher. Such an anomalous example can be seen in the chart for $b = 20$ in Figure 6.4. In that case, Procedure $\mathcal{M}_{\mathrm{BG}}$ cannot achieve $P^\star = 0.80$, but can achieve both the lower $P^\star = 0.79$ and the higher $P^\star = 0.81$. For this example, b must be increased to 25 before Procedure $\mathcal{M}_{\mathrm{BG}}$ can achieve $P^\star = 0.80$, whereas it can achieve $P^\star = 0.81$ in under 20 observations. This is a characteristic peculiar to Procedure $\mathcal{M}_{\mathrm{BG}}$ that is not shared by any of the other procedures we compare here. In some cases, such as for $k = 2$, $\theta^\star = 3$, $P^\star = 0.90$, Procedure $\mathcal{M}_{\mathrm{BG}}$ actually requires an infinite number of observations.

These anomalies are due to Procedure $\mathcal{M}_{\mathrm{BG}}$'s unique stopping rule. The stopping condition $z_m \leq (1 - P^\star)/P^\star$ was originally developed by BKS (1968) for an unbounded

stopping procedure. When BG (1985b, 1986) added the additional stopping parameter, n_{BG}, (thereby bounding the procedure) to save observations in expectation, the truncated procedure lost the ability to achieve $P^\star$-values for which the unboundedness of the observations was required. Furthermore, as our example shows, the minimum observation budget required to achieve $P^\star$ is not a non-decreasing function of $P^\star$. This is a result of the integer values of the differences calculated in the exponent of z_m, which make the set of possible values of z_m discrete instead of continuous.

For our example with $k = 2$, and $\theta^\star = 1.6$,

$$z_m = \begin{cases} 0.625 & \text{if } r' = 1 \\ 0.391 & \text{if } r' = 2 \\ 0.244 & \text{if } r' = 3 \\ 0.153 & \text{if } r' = 4. \end{cases} \tag{6.7}$$

For $P^\star = 0.8$, $(1 - P^\star)/P^\star = 0.25$, and we stop at difference $r' = 3$ when $z_m = 0.244$. For $P^\star = 0.79$, $(1 - P^\star)/P^\star = 0.266$, and we stop at difference $r' = 2$ when $z_m = 0.391$. As it turns out, stops from z_m values that are less than, but very close to, $(1 - P^\star)/P^\star$ require more observations to achieve $P^\star$, as is the case for $P^\star = 0.8$ above. For the extreme example when $k = 2$, $\theta^\star = 3$, and $P^\star = 0.9$, stopping condition $(1 - P^\star)/P^\star = 1/9$, which is exactly equal to z_m when $r' = 2$. In that case, the procedure requires an infinite observation budget to achieve $P^\star$.

We now examine similar plots when the number of alternatives is larger than $k = 2$. Figure 6.5 shows a series of charts for $b \in \{5, 10, 25, 40\}$ with $k = 4$ and $\theta^\star = 2.4$. In this figure, we see that the relationships between the procedures are more complex than they were for $k = 2$. None of the procedures perform identically, as some did for $k = 2$, but some do perform similarly when b is low.

We note some relationships between the procedures (regardless of k) that are reflected in Figure 6.5.

- Procedure $\mathcal{M}_{\mathrm{BK}}$ with parameter n_{BK} is a special case of Procedure $\mathcal{M}_{\mathrm{C}'}$ with parameter pair $(n_{\mathrm{C}'}, t)$, where $n_{\mathrm{C}'} = n_{\mathrm{BK}}$ and $t \geq \lceil n_{\mathrm{C}'}/2 \rceil$.

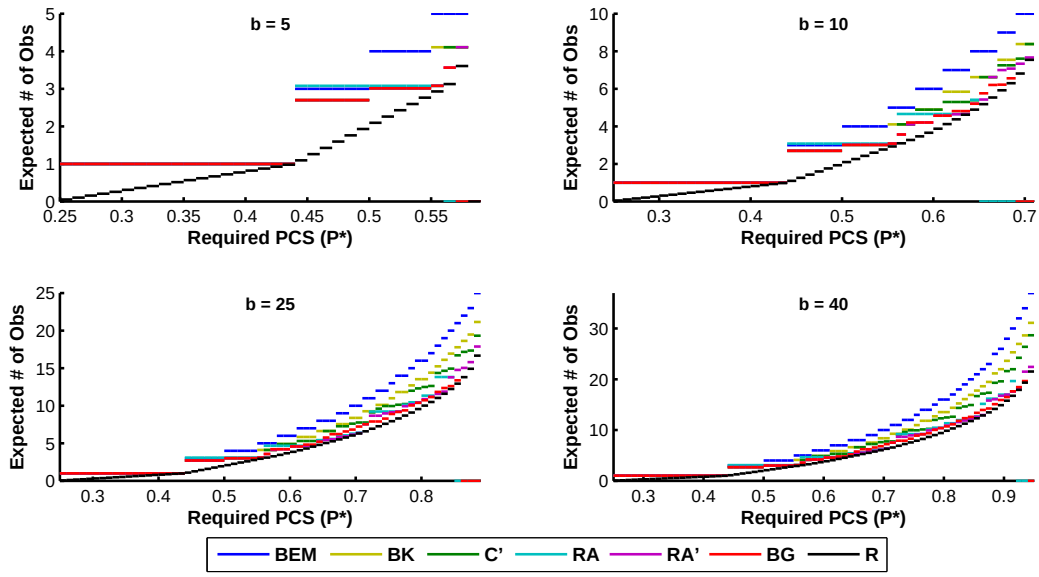

Figure 6.5: Procedure Comparison Plots for $k = 4$, $\theta^\star = 2.4$, $b \in \{5, 10, 25, 40\}$

- Procedure $\mathcal{M}_{\mathrm{C}'}$ with parameter pair $(n_{\mathrm{C}'}, t)$ is a special case of Procedure $\mathcal{M}_{\mathrm{RA}'}$ with parameter triplet $(n_{\mathrm{RA}'}, r', t')$ where $n_{\mathrm{RA}'} = n_{\mathrm{C}'}$, $r' \geq \lceil n_{\mathrm{C}'}/2 \rceil$, and $t' = t$.
- Procedure $\mathcal{M}_{\mathrm{RA}}$ with parameter pair (r, t) is a special case of Procedure $\mathcal{M}_{\mathrm{RA}'}$ with parameter triplet $(n_{\mathrm{RA}'}, r', t')$ where $n_{\mathrm{RA}'} \geq kt + 1$, $r' = r$, and $t' = t$.
- Procedure $\mathcal{M}_{\mathrm{BK}}$ will always perform better than Procedure $\mathcal{M}_{\mathrm{BEM}}$.

These relationships guarantee a relative ordering between Procedures $\mathcal{M}_{\mathrm{RA}}$ and $\mathcal{M}_{\mathrm{RA}'}$ and among Procedures $\mathcal{M}_{\mathrm{BEM}}$, $\mathcal{M}_{\mathrm{BK}}$, $\mathcal{M}_{\mathrm{C}'}$, and $\mathcal{M}_{\mathrm{RA}'}$, which are reflected in Figure 6.5, as well as in Figure 6.4. When considering the best performing procedure (not including the optimal procedures, of course), we need only compare Procedures $\mathcal{M}_{\mathrm{RA}'}$ and $\mathcal{M}_{\mathrm{BG}}$. Figure 6.5 shows that there are regions in which Procedure $\mathcal{M}_{\mathrm{RA}'}$ (and even Procedure $\mathcal{M}_{\mathrm{RA}}$) perform better than $\mathcal{M}_{\mathrm{BG}}$ and regions (seemingly more numerous) in which the opposite is true. We will address that in more detail later in this section.

Another insight from Figure 6.5 is the seemingly counterintuitive fact that Procedure $\mathcal{M}_{\mathrm{BK}}$ and even Procedure $\mathcal{M}_{\mathrm{BEM}}$ perform better than Procedure $\mathcal{M}_{\mathrm{RA}}$ for some $P^\star$-values. The reason for this phenomenon, which only occurs when b is low, is that the budget

Table 6.1: Procedures $\mathcal{M}_{\text{RA}}$ and $\mathcal{M}_{\text{BK}}$ Parameters for $k = 4$, $\theta^\star = 2.4$, and $b = 5$

Procedure $\mathcal{M}_{\text{RA}}$			Procedure $\mathcal{M}_{\text{BK}}$		
Parameters	$\mathsf{P}_{\text{SC}}(\text{CS})$	$\mathsf{E}_{\text{SC}}[N]$	Parameters	$\mathsf{P}_{\text{SC}}(\text{CS})$	$\mathsf{E}_{\text{SC}}[N]$
$(r = 1,\, t = 1)$	0.4444	1.000	$n_{\text{BK}} = 1$	0.4444	1.000
$(r = 2,\, t = 2)$	0.5690	3.080	$n_{\text{BK}} = 2$	0.4444	1.000
			$n_{\text{BK}} = 3$	0.5085	2.700
			$n_{\text{BK}} = 4$	0.5559	3.012
			$n_{\text{BK}} = 5$	0.5849	4.104

provides a more significant constraint on Procedure $\mathcal{M}_{\text{RA}}$ than it does for Procedures $\mathcal{M}_{\text{BEM}}$ and $\mathcal{M}_{\text{BK}}$. For a Procedure $\mathcal{M}_{\text{RA}}$ parameter pair (r, t) to be possible, we must have $b \geq k\,(t-1) + 1$. If $k = 4$ and $b = 5$, then we require that $t \leq 2$, resulting in the possible parameter pairs in Table 6.1. Note that when $r = 1$, the procedure stops after one observation, regardless of t; therefore, there is no need to include results for $(r = 1,\, t = 2)$. Consider $P^\star = 0.5$. Procedure $\mathcal{M}_{\text{BK}}$ with $n_{\text{BK}} = 3$ can achieve $P^\star$ with $\mathsf{E}_{\text{SC}}[N] = 2.7$, but Procedure $\mathcal{M}_{\text{RA}}$ with $(r = 2, t = 2)$, the only parameter pair that achieves $P^\star$, requires $\mathsf{E}_{\text{SC}}[N] = 3.08$. Even Procedure $\mathcal{M}_{\text{BEM}}$ with $n_{\text{BEM}} = 3$ achieves $P^\star$ with a lower $\mathsf{E}_{\text{SC}}[N] = 3$. These results agree with Figure 6.5, although the results for Procedure $\mathcal{M}_{\text{BK}}$ are masked by that of Procedure $\mathcal{M}_{\text{BG}}$ at $P^\star = 0.5$.

The anomalies that we noticed for Procedure $\mathcal{M}_{\text{BG}}$ do not appear at all in Figure 6.5. While common for $k = 2$, larger k allows for a greater number of possible z_m-values and thus fewer anomalies from large gaps between the discrete z_m-values. Nonetheless, we saw the phenomenon for $k > 2$, but much less frequently.

6.2.4 Mean Procedure Inefficiency

To supplement the visual insights provided by our charts, we also calculated our metrics, W_{J} and $\overline{W}_{\text{J}}^{I}$. The tables in Appendix B for our 36 procedure comparisons include values for W_{J} in the column labeled "% Incr", shown as a percentage (i.e., $100\,W_{\text{J}}$). For the 12 common combinations of $k \in \{2, 3, 4\}$ and $\theta^\star \in \{1.6, 2, 2.4, 3\}$ at four values of b, we also calculated $\overline{W}_{\text{J}}^{I}$. For those examples, we use a constant increment size of 0.01 for $P^\star$ (except for the interval between 1/3 and 0.34 when $k = 3$). We also use the same increment sizes

for each procedure $\mathcal{M}_{\mathrm{J}}$ as we do for Procedure $\mathcal{M}_{\mathrm{R}}$, against which procedure $\mathcal{M}_{\mathrm{J}}$ is being compared. The following are the four relevant intervals:

- $\overline{W}_{\mathrm{J}}^{I}$ is calculated from the entire interval from $1/k$ to the maximum achievable $P^\star$ by procedure $\mathcal{M}_{\mathrm{J}}$. For example, the maximum achievable $P^\star$ for Procedure $\mathcal{M}_{\mathrm{RA}}$ with $k = 2$, $\theta^\star = 2$, and $b = 20$ is 0.9313. The interval considered in this comparison is then (for both Procedures $\mathcal{M}_{\mathrm{RA}}$ and $\mathcal{M}_{\mathrm{R}}$) from 0.50 to 0.93, even though Procedure $\mathcal{M}_{\mathrm{R}}$ can achieve a higher $P^\star$ at $b = 20$. Thus, we should qualify the mean procedure inefficiency metric by calling it the *mean procedure inefficiency over its achievable $P^\star$-region* when that region is shorter than that of the optimal procedure. However, we omit the qualifier for the sake of brevity.

- $\overline{W}_{\mathrm{J}}^{75}$ is calculated from the interval from $1/k$ to the maximum achievable $P^\star$ or 0.75, whichever is less.

- $\overline{W}_{\mathrm{J}}^{90}$ is calculated from the interval from 0.75 to the maximum achievable $P^\star$ or 0.9, whichever is less. If procedure $\mathcal{M}_{\mathrm{J}}$ cannot achieve a $P^\star$ above 0.75, this metric is not defined.

- $\overline{W}_{\mathrm{J}}^{95}$ is calculated from the interval from 0.9 to the maximum achievable $P^\star$ or 0.95, whichever is less. If procedure $\mathcal{M}_{\mathrm{J}}$ cannot achieve a $P^\star$ above 0.9, this metric is not defined.

Figure 6.6 shows the mean procedure inefficiencies for each of the four $P^\star$-regions, with $b \in \{10, 20, 40, 60\}$, $k = 2$, and $\theta^\star = 1.6$ (i.e., corresponding to the charts in Figure 6.4). Here we see numerically what we noted in the plots of raw performance: Procedures $\mathcal{M}_{\mathrm{BK}}$ and $\mathcal{M}_{\mathrm{C}'}$ have the same performances, as do Procedures $\mathcal{M}_{\mathrm{RA}}$ and $\mathcal{M}_{\mathrm{RA}'}$. We also see that Procedure $\mathcal{M}_{\mathrm{BG}}$ performs similarly to Procedures $\mathcal{M}_{\mathrm{RA}}$ and $\mathcal{M}_{\mathrm{RA}'}$. The absence of a set of bars for any region means that none of the procedures can achieve $P^\star$ in that interval.

Figure 6.7 shows the mean procedure inefficiencies for each of the four regions, with $b \in \{5, 10, 25, 40\}$, $k = 4$, and $\theta^\star = 2.4$. Here, due to the larger k and lower numbers of observations, there are more regions within which none of the procedures can achieve a

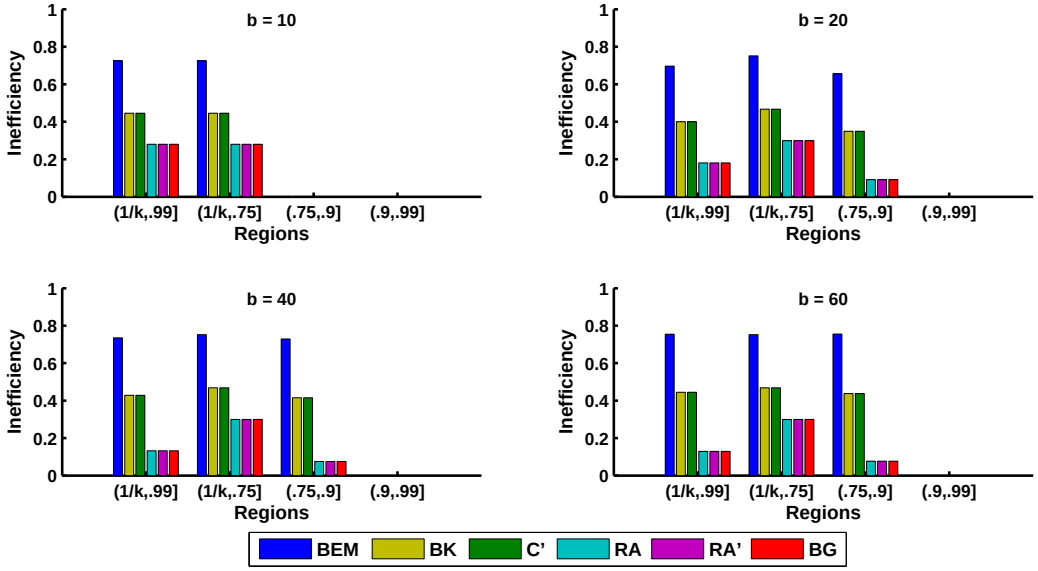

Figure 6.6: Mean Procedure Inefficiency for $k = 2$, $\theta^\star = 1.6$, $b \in \{10, 20, 40, 60\}$

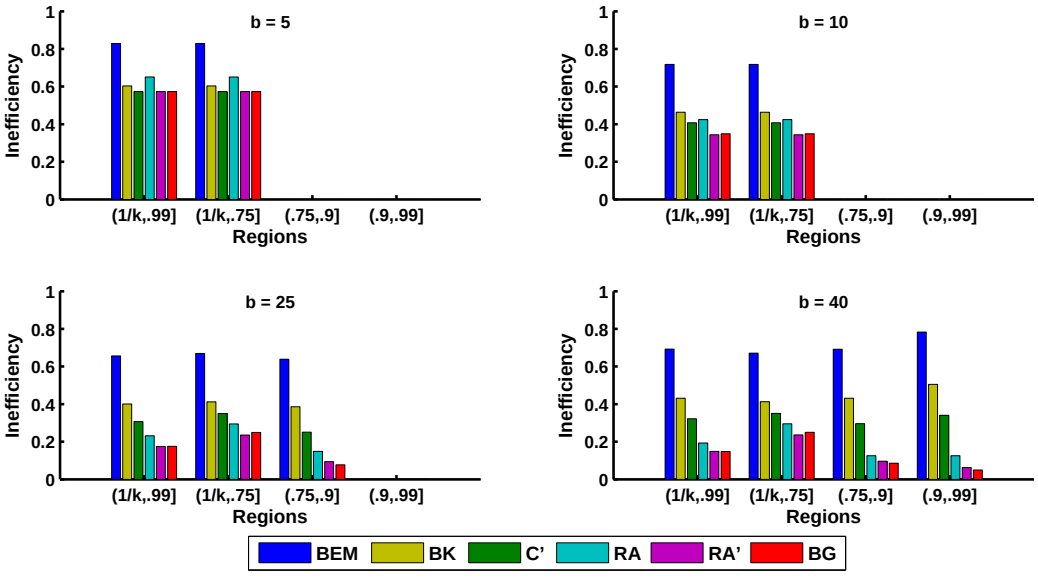

Figure 6.7: Mean Procedure Inefficiency for $k = 4$, $\theta^\star = 2.4$, $b \in \{5, 10, 25, 40\}$

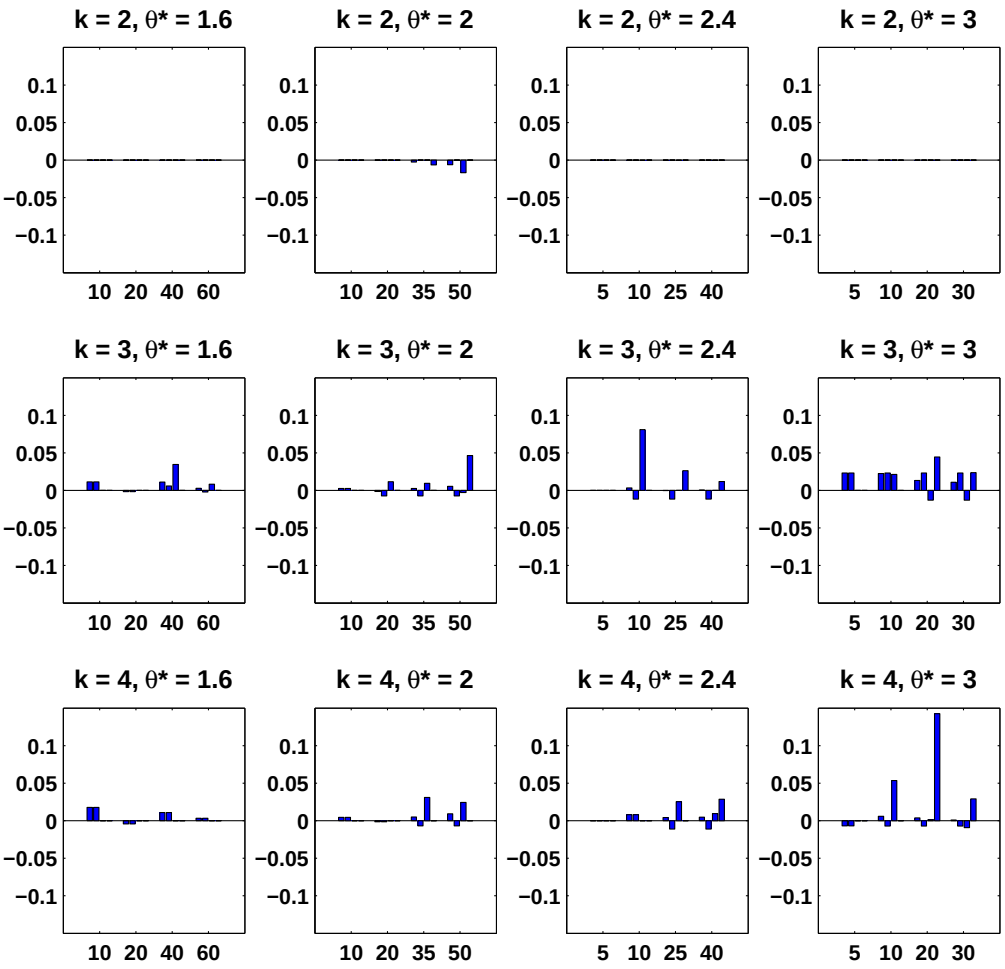

Figure 6.8: Mean Relative Procedure Performance: $\mathcal{M}_{\mathrm{RA}'}$ versus $\mathcal{M}_{\mathrm{BG}}$

particular $P^\star$. The relative ordering of procedure performance discussed in connection with Figure 6.5 is evident here, as is the poorer performance of Procedure $\mathcal{M}_{\mathrm{RA}}$ when b is low.

The results thus far prompted us to compare Procedures $\mathcal{M}_{\mathrm{RA}'}$ and $\mathcal{M}_{\mathrm{BG}}$. We have seen that the performance of Procedure $\mathcal{M}_{\mathrm{RA}'}$ dominates the performances of all other procedures except Procedure $\mathcal{M}_{\mathrm{BG}}$ (and the optimal procedures, of course). Therefore, we narrow our attention to just the two procedures by examining the metric $\overline{V}^I_{\mathrm{RA}',\mathrm{BG}}$.

Figure 6.8 shows the results for this comparison over the same sets of k, $\theta^\star$, and b we have analyzed thus far. Within each of the individual charts, we plot $\overline{V}^I_{\mathrm{RA}',\mathrm{BG}}$ for each of the four regions, grouped by the four values of b labeled on the horizontal axis. Bars

above the center line indicate regions within which Procedure $\mathcal{M}_{\mathrm{BG}}$ performs better than Procedure $\mathcal{M}_{\mathrm{RA}'}$. Bars below indicate regions within which the opposite is true. Points at which there are no bars indicate either identical or nearly identical performance, or a region within which the procedures cannot compete. The greater frequency of bars above versus below shows that for the regions and problems we examined, Procedure $\mathcal{M}_{\mathrm{BG}}$ performs better than $\mathcal{M}_{\mathrm{RA}'}$ more often than the reverse. However, we point out again that this comparison is over the intersection of their domains. In some cases, Procedure $\mathcal{M}_{\mathrm{RA}'}$ can attain a higher maximum $P^\star$ for a problem than can Procedure $\mathcal{M}_{\mathrm{BG}}$, which may provide a decisive advantage for particular situations.

Of course, we should not lose sight of the fact that Procedure $\mathcal{M}_{\mathrm{R}}$ (and Procedure $\mathcal{M}_{\mathrm{NR}}$) always perform as well as or better than all other existing procedures, and should be used if possible when minimization of the expected number of observations (or total cost) is the most important performance measure.

6.2.5 Distributional Comparisons

As we discussed in §6.2.2, we have complete distributional information for any procedure given the problem parameters (k and $\theta^\star$) and procedure parameters (e.g., n_{BK}, r, t, etc.). We can calculate the population variance of N in the SC, $\mathsf{Var}_{\mathrm{SC}}[N]$, and thus its standard deviation, $\mathsf{SD}_{\mathrm{SC}}[N]$, which we include in the tables in Appendix B. To look across the 36 cases, we use the coefficient of variation, CV, as our unitless measure of variability to compensate for the different values for b, where

$$CV = \frac{\mathsf{SD}_{\mathrm{SC}}[N]}{\mathsf{E}_{\mathrm{SC}}[N]]}. \tag{6.8}$$

Table 6.2 shows the results for the mean CV across all of the cases considered, as well as minimum and maximum values of CV for each procedure. We did not consider cases when $b = 1$ or when there is no entry for a procedure; therefore, the number of cases considered is less than 36. The relative ordering of the procedures, in terms of their mean CV, generally holds for each of the cases. The following lists the procedures in increasing order of variability for the cases we examined. This order was not necessarily intact for all cases, but summarizes the observed trend.

Table 6.2: Coefficient of Variation Results

Procedure	Cases	Mean CV	Min CV	Max CV
$\mathcal{M}_{\mathrm{R}}$	34	0.47	0.34	0.59
$\mathcal{M}_{\mathrm{NR}}$	29	0.45	0.20	0.60
$\mathcal{M}_{\mathrm{BG}}$	34	0.48	0.20	0.61
$\mathcal{M}_{\mathrm{RA}'}$	34	0.43	0.15	0.61
$\mathcal{M}_{\mathrm{RA}}$	25	0.41	0.20	0.61
$\mathcal{M}_{\mathrm{C}'}$	34	0.18	0.11	0.28
$\mathcal{M}_{\mathrm{BK}}$	35	0.13	0.05	0.20
$\mathcal{M}_{\mathrm{BEM}}$	35	0	0	0

1. Procedure $\mathcal{M}_{\mathrm{BEM}}$
2. Procedure $\mathcal{M}_{\mathrm{BK}}$
3. Procedure $\mathcal{M}_{\mathrm{C}'}$
4. Procedures $\mathcal{M}_{\mathrm{RA}}$ and $\mathcal{M}_{\mathrm{RA}'}$
5. Procedures $\mathcal{M}_{\mathrm{NR}}$ and $\mathcal{M}_{\mathrm{R}}$
6. Procedure $\mathcal{M}_{\mathrm{BG}}$

We may also be interested in more information about relative procedure performance. For example, a decision-maker might care about the minimum, maximum, or median of the observation distribution (N) as well. One tool we can use is a boxplot (or box-and-whisker plot). Figure 6.9 displays boxplots of the distribution of N for each procedure when $k = 3$, $\theta^\star = 2$, $P^\star = 0.9$, and $b = 34$, corresponding to Table B.6 in Appendix B. The bottom, middle, and top of the boxes represent the 25th, 50th (median), and 75th percentiles of the procedure distributions, respectively. The ends of the whiskers represent the minimum and maximum of the distributions. We have also added information about the mean and standard deviation in the blue triangular regions. The horizontal line in the center of the triangular region represents the mean ($\mathsf{E}_{\mathrm{SC}}[N]$); the triangles extend one standard deviation from the mean.

The figure confirms our relative ordering for procedure variability. We also see that the distributions of Procedures $\mathcal{M}_{\mathrm{R}}$, $\mathcal{M}_{\mathrm{NR}}$, and $\mathcal{M}_{\mathrm{BG}}$ are noticeably skewed towards lower

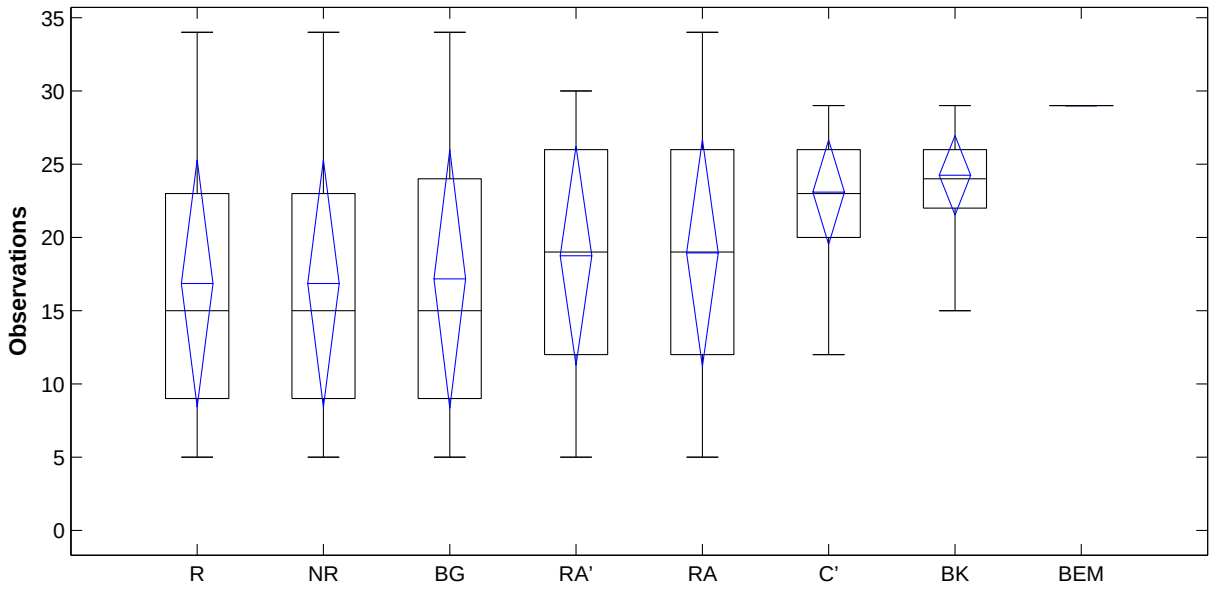

Figure 6.9: Procedure Distribution Boxplots for $k = 3$, $\theta^\star = 2$, $P^\star = 0.9$, and $b = 34$

numbers of observations, since their medians are below the centers of the rectangles. Plots such as these can provide decision-makers with the additional information necessary to compare other aspects of procedure performance in the SC or any other probability configuration. We could go a step further and plot the actual probability mass functions of each procedure; however, it would be rare that such detail would be necessary.

6.3 Procedure $\mathcal{M}_{\mathrm{RA}}$ as Gambler's Ruin when $k = 2$

One interesting result particular to Procedure $\mathcal{M}_{\mathrm{RA}}$ when $k = 2$ is its relationship to the gambler's ruin problem. Consider a game in which a gambler begins with r chips. At each turn, the gambler bets a single chip and wins with probability p (and thus loses with probability $q = 1 - p$). If he wins, he gets two chips; if he loses, he gets no chips. The gambler's goal is to obtain a total fortune of $2r$ chips before he is ruined (i.e., loses all of his original r chips). The characteristics of this game are the essence of the gambler's ruin problem.

Alam's (1971) unbounded MSP procedure, when $k = 2$, treats the MSP as a gambler's ruin problem. In his procedure, sampling stops when $\eta_{[2]m} - \eta_{[1]m} = r$, which is equivalent to the gambler either achieving his fortune of $2r$ chips or being ruined in the game we

described. Let p and q be the probabilities of success for the first and second alternatives, respectively, with $p \geq q$ and $p + q = 1$. Known results for the gambler's ruin problem (e.g., see Ross 2000) give the following:

$$\mathsf{P}(\text{CS}) = \frac{1}{1 + (q/p)^r}, \tag{6.9}$$

where $\mathsf{P}(\text{CS})$ can be interpreted as the probability that the gambler achieves his fortune, and

$$\mathsf{E}[N] = \frac{r}{p - q}\left(\frac{p^r - q^r}{p^r + q^r}\right), \tag{6.10}$$

where $\mathsf{E}[N]$ can be interpreted as the expected number of turns it will take for the game to end. As we mentioned in Chapter 2, Alam proved that the SC is the LFC when $k = 2$. That can be easily verified by noting that $\boldsymbol{p} = \left(\frac{\theta^\star}{\theta^\star + 1}, \frac{1}{\theta^\star + 1}\right)$ in the SC. Substituting for the values of p and q in (6.9), we get:

$$\mathsf{P}_{\text{SC}}(\text{CS}) = \frac{1}{1 + (1/\theta^\star)^r},$$

which decreases as $\theta^\star$ decreases.

Procedure $\mathcal{M}_{\text{RA}}$ adds an additional parameter t. The modification to the game associated with the gambler's ruin problem is that the game can now end under one of three conditions, whichever occurs first: 1) the gambler achieves his fortune, 2) the gambler is ruined, or 3) the gambler has won or lost t of the turns. Figure 6.10 shows the possible sample paths of the game when $r = 3$ and $t = 5$. Each level represents an observation [turn in the game]. Movement along an arc to the left represents a success by alternative 1 [win by the player]. Gray nodes represent potential stopping points and black nodes represent those points that cannot be reached given the procedure parameters. Figure 6.10 demonstrates that Procedure $\mathcal{M}_{\text{RA}}$ will never stop when the alternatives have the same number of successes [the game will never end with the players tied].

Initially, we consider Alam's (1971) unbounded MSP procedure with $r = 3$. Let $\mathsf{P}(i, j)$ be the probability of arriving at point (i, j). The first possible stops occur at the third observation (e.g., see Figure 6.10, which is equivalent to Alam's unbounded procedure through

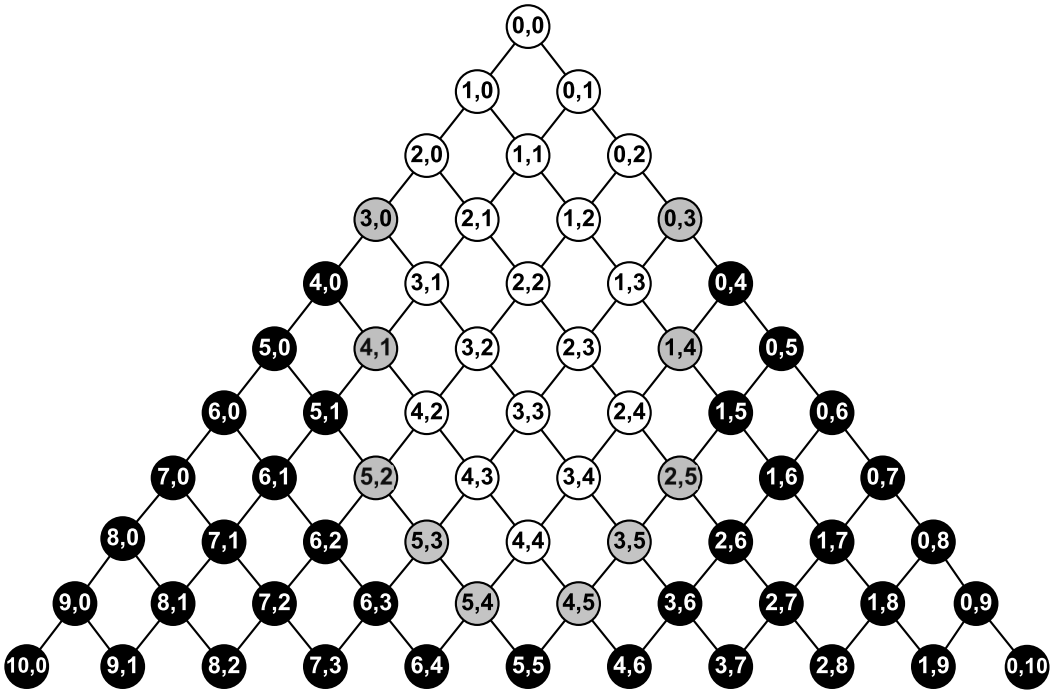

Figure 6.10: Procedure $\mathcal{M}_{\mathrm{RA}}$ Stopping Points for ($r = 3$, $t = 5$) when $k = 2$

the seventh observation). Then

$$\mathsf{P}(3,0) = p^3, \qquad \mathsf{P}(0,3) = q^3,$$
$$\mathsf{P}(2,1) = 3p^2q, \qquad \mathsf{P}(1,2) = 3pq^2.$$

The next possible stops occur at the fifth observation, where

$$\mathsf{P}(4,1) = p^2\mathsf{P}(2,1) = 3p^4q, \qquad \mathsf{P}(1,4) = q^2\mathsf{P}(1,2) = 3pq^4,$$
$$\mathsf{P}(3,2) = 2pq\mathsf{P}(2,1) + p^2\mathsf{P}(1,2) = 9p^3q^2, \qquad \mathsf{P}(2,3) = 2pq\mathsf{P}(1,2) + q^2\mathsf{P}(2,1) = 9p^2q^3.$$

At the seventh observation, we get

$$\mathsf{P}(5,2) = p^2\mathsf{P}(3,2) = 9p^5q^2, \qquad \mathsf{P}(2,5) = q^2\mathsf{P}(2,3) = 9p^2q^5,$$
$$\mathsf{P}(4,3) = 2pq\mathsf{P}(3,2) + p^2\mathsf{P}(2,3) = 27p^4q^3, \qquad \mathsf{P}(3,4) = 2pq\mathsf{P}(2,3) + q^2\mathsf{P}(3,2) = 27p^3q^4.$$

These recursive equations give us

$$\mathsf{P}(j+3,j) = p^3(3pq)^j, \qquad \mathsf{P}(j,j+3) = q^3(3pq)^j,$$
$$\mathsf{P}(j+1,j) = p(3pq)^j, \qquad \mathsf{P}(j,j+1) = q(3pq)^j,$$

for $j \geq 0$. In general, the $\mathsf{P}(\text{CS})$ contribution of the first ℓ stopping points is

$$\begin{aligned}
\mathsf{P}(\text{CS}) &= \sum_{j=0}^{\ell} \mathsf{P}(j+3, j) \\
&= \sum_{j=0}^{\ell} p^3 (3pq)^j \\
&= \frac{p^3 \left[1 - (3pq)^{\ell+1}\right]}{1 - 3pq}.
\end{aligned}$$

If we take the limit,

$$\lim_{\ell \to \infty} \frac{p^3 \left[1 - (3pq)^{\ell+1}\right]}{1 - 3pq} = \frac{p^3}{1 - 3pq},$$

it is equivalent to the gambler's ruin equation (6.9) with $r = 3$.

We have verified that Alam's unbounded procedure is equivalent to a gambler's ruin problem when $k = 2$. Now we return to Procedure $\mathcal{M}_{\text{RA}}$ and consider the specific problem when $r = 3$ and $t = 5$. Through the first seven observations, the stopping points coincide with Alam's procedure, but from that point we must consider the stopping points unique to the parameter t. For this example,

$$\begin{aligned}
\mathsf{P}(\text{CS}) &= \mathsf{P}(3,0) + \mathsf{P}(4,1) + \mathsf{P}(5,2) + \mathsf{P}(5,3) + \mathsf{P}(5,4) \\
&= p^3 \frac{1 - (3pq)^3}{1 - 3pq} + p\mathsf{P}(4,3) + pq\mathsf{P}(4,3) + q^2\mathsf{P}(3,4) \\
&= p^3 \frac{1 - (3pq)^3}{1 - 3pq} + (p + pq)[p(3pq)^3] + q^2[q(3pq)^3] \\
&= p^3 \frac{1 - (3pq)^3}{1 - 3pq} + (p^2 + 2p^2q)(3pq)^3 \qquad (6.11)
\end{aligned}$$

In our example, if $\theta^\star = 1.6$, then $\boldsymbol{p} = (8/13, 5/13)$ in the SC, and using (6.11), $\mathsf{P}_{\text{SC}}(\text{CS}) = 0.7559$, which agrees with Table A.3 in Appendix A.3.

We now turn our attention to $\mathsf{E}[N]$. For Alam's procedure,

$$\begin{aligned}
\mathsf{E}[N] &= \sum_{j=0}^{\infty}(j+3+j)[\mathsf{P}(j+3,j)+\mathsf{P}(j,j+3)] \\
&= \sum_{j=0}^{\infty}(2j+3)[p^3(3pq)^j+q^3(3pq)^j] \\
&= 2(p^3+q^3)\sum_{j=0}^{\infty}j(3pq)^j+3(p^3+q^3)\sum_{j=0}^{\infty}(3pq)^j \\
&= 2(p^3+q^3)\frac{3pq}{(1-3pq)^2}+\frac{3(p^3+q^3)}{1-3pq} \\
&= \frac{p^3+q^3}{1-3pq}\left(\frac{6pq}{1-3pq}+\frac{3-9pq}{1-3pq}\right) \\
&= \frac{3(1-pq)}{1-3pq}, \qquad (6.12)
\end{aligned}$$

where $p^3+q^3=1-3pq$ when $p+q=1$. Equation (6.12) is equivalent to the gambler's ruin equation (6.10) when $r=3$, although the algebra required to show that is quite involved. For our bounded Procedure $\mathcal{M}_{\mathrm{RA}}$ in the example:

$$\begin{aligned}
\mathsf{E}[N] &= 3[\mathsf{P}(3,0)+\mathsf{P}(0,3)]+5[\mathsf{P}(4,1)+\mathsf{P}(1,4)]+7[\mathsf{P}(5,2)+\mathsf{P}(2,5)]+ \\
&\qquad 8[\mathsf{P}(5,3)+\mathsf{P}(3,5)]+9[\mathsf{P}(5,4)+\mathsf{P}(4,5)] \\
&= 3(p^3+q^3)+5(3p^4q+3pq^4)+7(9p^5q^2+9p^2q^5)+ \\
&\qquad 8[p\mathsf{P}(4,3)+q\mathsf{P}(3,4)]+9[pq\mathsf{P}(4,3)+p^2\mathsf{P}(3,4)+pq\mathsf{P}(3,4)+q^2\mathsf{P}(4,3)] \\
&= 3(p^3+q^3)+15pq(p^3+q^3)+63p^2q^2(p^3+q^3)+ \\
&\qquad 8(27p^5q^3+27p^3q^5)+9(27p^5q^4+27p^5q^4+27p^4q^5+27p^4q^5) \\
&= (3+15pq+63p^2q^2)(p^3+q^3)+216p^3q^3(p^2+q^2)+243p^4q^4(p+p+q+q) \\
&= (3+15pq+63p^2q^2)(p^3+q^3)+216p^3q^3(p^2+q^2)+486p^4q^4. \qquad (6.13)
\end{aligned}$$

In our example, if $\theta^\star=1.6$ in the SC, and using (6.13), we find that $\mathsf{E}_{\mathrm{SC}}[N]$= 5.956, which agrees with Table A.3 in Appendix A.3.

We have shown that Alam's (1971) unbounded procedure is a gambler's ruin problem when $k=2$. We have also shown that Procedure $\mathcal{M}_{\mathrm{RA}}$ is a modified version of the gambler's ruin problem. It is possible to extend the gambler's ruin problem to more than two players (i.e., $k>2$); however, the computation of the recursive equations is very burdensome. For

example, see Sobel and Frankowski (2002). In their paper, they develop an MSP procedure based upon the multi-player gambler's ruin problem; however, their procedure is unbounded and allows the elimination of individual alternatives during sampling, both of which are violations of our problem assumptions.

6.4 Summary

In §6.1, we showed that, in many cases, the expected number of observations for Procedure $\mathcal{M}_{\mathrm{NR}}$ rapidly approaches, as a function of b, the expected number of observations for Procedure $\mathcal{M}_{\mathrm{R}}$. For lower values of $P^\star$ coupled with larger $\theta^\star$ (i.e., when few observations are required to achieve $P^\star$), the optimal Procedures $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{NR}}$ do not converge to each other, but remain equidistant as b increases. These results allow us to omit Procedure $\mathcal{M}_{\mathrm{NR}}$ in our subsequent comparisons.

We then developed a number of metrics in §6.2.2 to examine different aspects of procedure performance. We used those metrics and selected charts in §6.2.3 and §6.2.4 to demonstrate some important relationships between the procedures in terms of performance, particularly when $k = 2$, as well as some interesting anomalies in the performance of Procedure $\mathcal{M}_{\mathrm{BG}}$. We also focused on a more thorough comparison of Procedures $\mathcal{M}_{\mathrm{BG}}$ and $\mathcal{M}_{\mathrm{RA}'}$, showing that Procedure $\mathcal{M}_{\mathrm{BG}}$ usually performs better in terms of $\mathsf{E}_{\mathrm{SC}}[N]$, but that Procedure $\mathcal{M}_{\mathrm{RA}'}$ can sometimes attain a higher maximum $P^\star$. In §6.2.5, we looked at additional information provided by the distribution of N for each MSP procedure. In particular, we were able to examine and compare procedure variability.

Finally, in §6.3, we showed that Alam's (1971) procedure when $k = 2$ is identical to the classic gambler's ruin problem, and that Procedure $\mathcal{M}_{\mathrm{RA}}$ represents a modification to the gambler's ruin problem by changing the game slightly.

CHAPTER VII

PROCEDURE IMPLICATIONS

In this chapter, we address some additional questions that arose throughout the course of our research. Our goals are to gain initial insights into the issues, and, more importantly, to set the stage for continued research and future work. All three of the topics discussed here relate to fundamental assumptions regarding the application of MSP procedures normally taken for granted in the literature.

For example, we typically accept problem inputs, in particular, the IZ parameter, as given, without concern for their effects on the problem space. In §7.1, we show how the choice of the IZ parameter affects the size of the PZ when all alternative configurations are equally likely. We also examine the size of IZ subregions under the same conditions. In §7.2, instead of concerning ourselves with probability guarantees in the PZ, we define the concept of an "acceptable selection" for alternatives in the IZ and conduct experiments to characterize the new metric. Finally, we shift our focus in §7.3 from the *prior* $\mathsf{P}(\mathrm{CS})$ to *posterior* conditional $\mathsf{P}(\mathrm{CS})$ at procedure termination. In other words, what is the $\mathsf{P}(\mathrm{CS})$ after we have garnered additional information during the course of sampling? We summarize our results in §7.4.

7.1 Preference Zone

In this section, we wish to determine the probability that a configuration $\boldsymbol{p}$, randomly chosen from sample space Ω, is in the PZ. In order to do so, we must have some information about the distribution of $\boldsymbol{p}$. We will assume the simplest model in which all possible probability configurations, $\boldsymbol{p} \in \Omega$, are equally likely (i.e., probability configurations are uniformly distributed over the sample space). We use k-dimensional geometry to calculate the fraction of the probability configuration sample space that is considered the PZ. Under our model, that fraction is our desired probability.

Recall that the PZ is defined as $\Omega_{\mathrm{PZ}} \equiv \{\boldsymbol{p} \in \Omega : p_{[k]}/p_{[k-1]} \geq \theta^{\star}\}$. We wish to

calculate the ratio of the geometric volume of Ω_{PZ} to the geometric volume of Ω. By the uniformity assumption, all $k!$ permutations of a particular probability vector $\boldsymbol{p}$ are equally likely. Therefore, it is sufficient for us to consider the volume of one permutation of the input space, say $p_1 \geq p_2 \geq \cdots \geq p_k$. Multiplying by $k!$ will not be necessary, since that term will cancel in the ratio.

In §7.1.1, we identify the k vertices of the $(k-1)$-dimensional polytope representing the PZ. In §7.1.2, we apply a formula for the volume of a simplex to the PZ and the sample space to obtain our desired probability. We then briefly discuss the implications via numerical examples in §7.1.3. Finally, in §7.1.4 with reference to Appendix D, we apply similar techniques to the geometries of the IZ.

7.1.1 Vertices

We begin by identifying the vertices of the general k-dimensional polytope defined by the permutation of probability configurations in which $p_1 \geq p_2 \geq \cdots \geq p_k$. The resulting polytope is

$$\Omega'_{\text{PZ}} \equiv \left\{ (p_1, \ldots, p_k) \in \mathbb{R}^k : \begin{array}{l} p_1 \geq \theta^\star \, p_2;\; p_i \geq p_{i+1},\; 2 \leq i \leq k-1; \\ p_k \geq 0;\; \sum_{i=1}^{k} p_i = 1 \end{array} \right\}, \tag{7.1}$$

where the prime in Ω'_{PZ} indicates the permutation of the space.

The $k+1$ relationships defining the polytope are the minimum required to represent the region of interest. If we think of the region as a feasible region in the linear programming sense (i.e., defined by a set of constraints), we can use the following definition of a basic solution; for example, see Bertsimas and Tsitsiklis (1997). Given a bounded convex region in k-dimensional space, a basic solution is the unique solution to a set of k independent equations consisting of all m equalities that define its boundary and $k-m$ binding inequalities (i.e., inequalities represented as equalities). If the basic solution satisfies all of the constraints, we will call it a vertex.

We have only one equality, and must therefore choose among the k inequalities to determine the $k-1$ remaining equations that will define the basic solution. Our approach is to remove one inequality at a time, set the remaining $k-1$ inequalities to equalities, solve

the resulting set of equations to find each basic solution, and check to see if the solution satisfies the one inequality that we removed. If so, we will call it a vertex. Incidentally, if we have k unique vertices at the end of this process, we have also shown that our set of relationships defining $\Omega'_{\rm PZ}$ in (7.1) is the minimum set required, i.e., there are no redundant equations.

We only need to consider three cases. In the first, we remove the first inequality, $p_1 \geq \theta^\star\, p_2$, and make the rest into equalities. We now have

$$\begin{array}{rcl} p_i & = & p_{i+1},\ 2 \leq i \leq k-1, \\ p_k & = & 0, \\ \sum_{i=1}^k p_i & = & 1. \end{array}$$

Then $p_i = 0,\ 2 \leq i \leq k$, and by our total probability requirement, $p_1 = 1$. Clearly this satisfies $p_1 \geq \theta^\star\, p_2$, so our first vertex is

$$\boldsymbol{v}_1 = (1, 0, \cdots, 0)^T. \tag{7.2}$$

In the second case, we remove one of the inequalities in the set $p_i \geq p_{i+1},\ 2 \leq i \leq k-1$. Our new set of equations is:

$$\begin{array}{rcl} p_1 & = & \theta^\star\, p_2, \\ p_2 & = & p_3, \\ & \vdots & \\ p_{i-1} & = & p_i, \\ p_{i+1} & = & p_{i+2}, \\ & \vdots & \\ p_{k-1} & = & p_k, \\ p_k & = & 0, \\ \sum_{i=1}^k p_i & = & 1. \end{array}$$

When we remove inequality i, the new equations require that $p_{i+1} = p_{i+2} = \cdots = p_k = 0$. Solving the remaining equations we get

$$\theta^\star\, p_2 + (i-1)p_2 = 1, \quad \text{or} \quad p_2 = 1/(\theta^\star + i - 1).$$

Our ith feasible solution is then

$$\boldsymbol{v}_i = \left(\frac{\theta^\star}{\theta^\star + i - 1}, \frac{1}{\theta^\star + i - 1}, \cdots, \frac{1}{\theta^\star + i - 1}, 0, \cdots, 0 \right)^T, \tag{7.3}$$

in which there are i positive terms and $k-i$ zeros. It is a vertex because it satisfies the inequality, $p_i \geq p_{i+1}$, that we removed.

Our final case is the removal of $p_k \geq 0$. We now have

$$\begin{array}{rcl} p_1 & = & \theta^\star\, p_2, \\ p_i & = & p_{i+1},\ 2 \leq i \leq k-1, \\ \sum_{i=1}^k p_i & = & 1. \end{array}$$

These equations require that $\theta^\star\, p_2 + (k-1)p_2 = 1$. Solving and substituting back into our set of equations results in the final basic solution (and vertex since $p_k = 1/(\theta^\star + k - 1) \geq 0$):

$$\boldsymbol{v}_k = \left(\frac{\theta^\star}{\theta^\star + k - 1}, \frac{1}{\theta^\star + k - 1}, \cdots, \frac{1}{\theta^\star + k - 1} \right)^T, \tag{7.4}$$

which is the SC.

We now have k unique vertices in $\mathbb{R}^k$. Our polytope is a convex hull of the vertices, and is a $(k-1)$-simplex, a fact we will use in the next section to calculate the volume of the polytope.

7.1.2 Volume Calculations

Our objective is to calculate the ratio of the volume of the simplex for a general $\theta^\star$ over the volume of the simplex with $\theta^\star = 1$. That will provide the fraction of the overall volume of the sample space that is the PZ, and thus, the probability that a randomly drawn probability vector $\boldsymbol{p}$ is in the PZ, under the assumption that all $\boldsymbol{p} \in \Omega$ are distributed uniformly over the sample space.

Given its vertices, we can calculate the volume of any $(k-1)$-simplex in $\mathbb{R}^{k-1}$ via the following formula (Büeler, Enge, and Eukuda 2000, pg 133):

$$V = \frac{|\det[\boldsymbol{v}_2 - \boldsymbol{v}_1 \quad \boldsymbol{v}_3 - \boldsymbol{v}_1 \quad \cdots \quad \boldsymbol{v}_k - \boldsymbol{v}_1]|}{(k-1)!}, \tag{7.5}$$

where V is the volume of the simplex polytope with vertices $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$. Unfortunately, our vertices are in $\mathbb{R}^k$, since $\boldsymbol{v}_k$ has a nonzero component in the kth dimension, so we cannot apply (7.5) directly.

Let $V_{\theta^\star}$ be the volume of the $(k-1)$-simplex defined by vertices $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^k$. Let V_1 be the particular value of $V_{\theta^\star}$ when $\theta^\star = 1$. Now let $\boldsymbol{v}_i'$ be the projection of $\boldsymbol{v}_i$ onto the

standard basis in $\mathbb{R}^{k-1}$. Then $V'_{\theta^\star}$ is the volume of the $(k-1)$-simplex defined by vertices $\boldsymbol{v}'_1, \ldots, \boldsymbol{v}'_k \in \mathbb{R}^{k-1}$ and V'_1 is the particular value of $V'_{\theta^\star}$ when $\theta^\star = 1$. We state the following lemma.

Lemma 7.1 *The ratio $V_{\theta^\star}/V_1$ is equivalent to the ratio $V'_{\theta^\star}/V'_1$.*

We consider the shifted polytope (i.e., the polytope represented by $\mathbf{V} = [\boldsymbol{v}_2 - \boldsymbol{v}_1 \quad \boldsymbol{v}_3 - \boldsymbol{v}_1 \quad \cdots \quad \boldsymbol{v}_k - \boldsymbol{v}_1]$), since shifting the polytope does not affect the volume. After shifting, we see that

$$\begin{aligned} \boldsymbol{v}_i - \boldsymbol{v}_1 &= \left(\frac{-(i-1)}{\theta^\star + i - 1}, \frac{1}{\theta^\star + i - 1}, \ldots, \frac{1}{\theta^\star + i - 1}, 0, \ldots, 0 \right) \\ &= \frac{1}{\theta^\star + i - 1} \left(-(i-1), 1, \ldots, 1, 0, \ldots, 0 \right), \end{aligned} \tag{7.6}$$

for $i = 2, \ldots, k$, where there are $i-1$ ones and $k-i$ zeros. We are interested in the volume of two different simplexes. The first, Δ_1, is the $(k-1)$-simplex representing region Ω' — the permutation of the overall sample space in which we are interested. The second, $\Delta_{\theta^\star}$, is the $(k-1)$-simplex representing the region Ω'_{PZ}. Note that $\Delta_1 = \Delta_{\theta^\star}$ when $\theta^\star = 1$. By definition, $\Omega'_{\text{PZ}} \subset \Omega'$ when $\theta^\star > 1$, so $\Delta_{\theta^\star}$ is wholly contained within Δ_1. Furthermore, by (7.6), each vertex of $\Delta_{\theta^\star}$ falls along the same vector as each vertex of Δ_1, but is closer to the origin.

Let $\Delta'_{\theta^\star}$ be the projection of $\Delta_{\theta^\star}$ onto the standard basis in $\mathbb{R}^{k-1}$, $\boldsymbol{w}'_{\theta^\star}$ be the vertex $\boldsymbol{v}_k - \boldsymbol{v}_1$ of $\Delta'_{\theta^\star}$, and $\boldsymbol{w}_{\theta^\star}$ be the vertex $\boldsymbol{v}_k - \boldsymbol{v}_1$ of $\Delta_{\theta^\star}$. The only difference between $\Delta'_{\theta^\star}$ and $\Delta_{\theta^\star}$ is that $w'_{\theta^\star,k} = 0$ and $w_{\theta^\star,k} = 1/(\theta^\star + k - 1)$, respectively. For Ω', we use similar notation, but with $\theta^\star$ replaced by 1 in the subscript. Now consider the change from $w'_{1,k}$ to $w_{1,k}$, leaving all of the other vertices in Δ'_1 fixed. By (7.6), all of the vertices in $\Delta'_{\theta^\star}$ will also remain fixed, except for component $w_{\theta^\star,k}$ of $\boldsymbol{w}_{\theta^\star}$. The change in the magnitude of $\boldsymbol{w}_{\theta^\star}$ will maintain the same proportion as the change in the magnitude of $\boldsymbol{w}_1$ (since $\boldsymbol{w}_{\theta^\star}$ is a point along vector $\boldsymbol{w}_1$). But then the changes in the respective volumes must maintain the same proportion as well, since we are moving one vertex away from the other fixed vertices in the same proportion, and thus

$$V_{\theta^\star}/V_1 = V'_{\theta^\star}/V'_1,$$

concluding our proof. □

Lemma 7.1 allows us achieve our goal by calculating $V'_{\theta^\star}$ and V'_1 via (7.5) and using their ratio to determine the relative volume of the PZ to the sample space. We have already numbered our vertices in the manner we wish use them to create the matrix $\mathbf{V}' = [\boldsymbol{v}'_2 - \boldsymbol{v}'_1 \quad \boldsymbol{v}'_3 - \boldsymbol{v}'_1 \quad \cdots \quad \boldsymbol{v}'_k - \boldsymbol{v}'_1]$,

$$\mathbf{V}' = \begin{bmatrix} \frac{-1}{\theta^\star+1} & \frac{-2}{\theta^\star+2} & \frac{-3}{\theta^\star+3} & \cdots & \frac{-(k-3)}{\theta^\star+k-3} & \frac{-(k-2)}{\theta^\star+k-2} & \frac{-(k-1)}{\theta^\star+k-1} \\ \frac{1}{\theta^\star+1} & \frac{1}{\theta^\star+2} & \frac{1}{\theta^\star+3} & \cdots & \frac{1}{\theta^\star+k-3} & \frac{1}{\theta^\star+k-2} & \frac{1}{\theta^\star+k-1} \\ 0 & \frac{1}{\theta^\star+2} & \frac{1}{\theta^\star+3} & \cdots & \frac{1}{\theta^\star+k-3} & \frac{1}{\theta^\star+k-2} & \frac{1}{\theta^\star+k-1} \\ 0 & 0 & \frac{1}{\theta^\star+3} & \cdots & \frac{1}{\theta^\star+k-3} & \frac{1}{\theta^\star+k-2} & \frac{1}{\theta^\star+k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\theta^\star+k-3} & \frac{1}{\theta^\star+k-2} & \frac{1}{\theta^\star+k-1} \\ 0 & 0 & 0 & \ldots & 0 & \frac{1}{\theta^\star+k-2} & \frac{1}{\theta^\star+k-1} \end{bmatrix},$$

where $\mathbf{V}'$ is a $(k-1) \times (k-1)$ matrix.

The most straightforward way to calculate the determinant is to reduce $\mathbf{V}'$ to a triangular matrix, ensuring that the reduction methods used leave the determinant unchanged. Strang (1988) shows that the determinant of a triangular matrix is equal to the product of the diagonal elements of the matrix. He shows further that the operation of subtracting a multiple of one row from another row does not change the determinant. We indicate the jth element of the ith row by r_{ij}, with r_{ii} denoting the diagonal elements. The general form for the ith row, $i \geq 2$, is:

$$r_i = \left[0, \ldots, 0, \frac{1}{\theta^\star + i - 1}, \frac{1}{\theta^\star + i}, \ldots, \frac{1}{\theta^\star + k - 1}\right],$$

in which there are $i-2$ leading zeros.

Let $r_i^\dagger$ be r_i transformed by adding the previous $(i-1)$th row after transformation (i.e., $r_{i-1}^\dagger$). Note that $r_1^\dagger = r_1$ since the first row is already in a form required for an upper triangular matrix.

Lemma 7.2 *The general form of the ith row, $2 \geq i \geq k-1$, after reduction of $\mathbf{V}'$ to an upper triangular matrix, is*

$$r_i^\dagger = \left[0, \ldots, 0, \frac{-1}{\theta^\star + i}, \ldots, \frac{-(k-i)}{\theta^\star + k - 1}\right] \quad \forall i \geq 2,$$

in which there are $i-1$ leading zeros.

Our proof of Lemma 7.2 is by induction. For $i = 2$, we add $r_1^\dagger$ to r_2, giving us

$$r_2^\dagger = \left[0, \frac{-1}{\theta^\star + 2}, \frac{-2}{\theta^\star + 3}, \ldots, \frac{-(k-3)}{\theta^\star + k - 2}, \frac{-(k-2)}{\theta^\star + k - 1}\right].$$

We then assume that the induction hypothesis is true for i and consider $i+1$. We first note that

$$r_{i+1} = \left[0, \ldots, 0, \frac{1}{\theta^\star + i}, \frac{1}{\theta^\star + i + 1}, \ldots, \frac{1}{\theta^\star + k - 1}\right],$$

in which there are $i-1$ leading zeros. We then add $r_i^\dagger$ to r_{i+1} to get:

$$\begin{aligned} r_{i+i}^\dagger &= \left[0, \ldots, 0, 0, \frac{-1}{\theta^\star + i + 1}, \ldots, \frac{-(k-i)+1}{\theta^\star + k - 1}\right], \\ &= \left[0, \ldots, 0, 0, \frac{-1}{\theta^\star + (i+1)}, \ldots, \frac{-(k-(i+1))}{\theta^\star + k - 1}\right], \end{aligned}$$

in which there are $(i-1)+1 = i$ leading zeros. This completes the proof. □

We have used row reduction to transform $\mathbf{V}'$ into the triangular matrix

$$\mathbf{V}^\dagger = \begin{bmatrix} r_1^{\dagger T} & r_2^{\dagger T} & r_3^{\dagger T} & \cdots & r_{k-1}^{\dagger T} \end{bmatrix}^T,$$

where

$$r_{ii}^\dagger = \frac{-1}{\theta^\star + i} \quad \forall i = 1, \ldots, k-1.$$

Then

$$\begin{aligned} \det \mathbf{V}' = \det \mathbf{V}^\dagger &= \prod_{i=1}^{k-1} r_{ii}^\dagger \\ &= \prod_{i=1}^{k-1} \frac{-1}{\theta^\star + i} \\ &= (-1)^{k-1} \prod_{i=1}^{k-1} \frac{1}{\theta^\star + i}. \end{aligned} \tag{7.7}$$

Substituting (7.7) into (7.5), we get

$$\begin{aligned} V &= \frac{1}{(k-1)!} \left| (-1)^{k-1} \prod_{i=1}^{k-1} \frac{1}{\theta^\star + i} \right| \\ &= \frac{1}{(k-1)!} \prod_{i=1}^{k-1} \frac{1}{\theta^\star + i}. \end{aligned} \tag{7.8}$$

We are interested in the ratio

$$\begin{aligned}
\frac{V_{\theta^\star}}{V_1} &= \frac{V'_{\theta^\star}}{V'_1} \\
&= \frac{\dfrac{1}{(k-1)!}\displaystyle\prod_{i=1}^{k-1}\frac{1}{\theta^\star+i}}{\dfrac{1}{(k-1)!}\displaystyle\prod_{i=1}^{k-1}\frac{1}{1+i}} \\
&= \frac{\displaystyle\prod_{i=1}^{k-1}\frac{1}{\theta^\star+i}}{\displaystyle\prod_{i=2}^{k}\frac{1}{i}} \\
&= k!\ \prod_{i=1}^{k-1}\frac{1}{\theta^\star+i}. \qquad (7.9)
\end{aligned}$$

We now state the following theorem.

Theorem 7.1 *For an indifference zone MSP involving k alternatives and a given $\theta^\star$, and assuming that all possible probability configurations $\boldsymbol{p} \in \Omega$ are equally likely (i.e., uniformly distributed in Ω), the probability that $\boldsymbol{p} \in \Omega_{\mathrm{PZ}}$ when drawn randomly from Ω, denoted $\mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{PZ}})$, is given by*

$$\mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{PZ}}) = k!\ \prod_{i=1}^{k-1}\frac{1}{\theta^\star+i}.$$

7.1.3 Implications of PZ Size

Our assumption that all $\boldsymbol{p} \in \Omega$ are equally likely is a reasonable one when the experimenter has no prior knowledge concerning the relative performance of the alternatives. In such a situation, Theorem 7.1 gives the experimenter greater information about the implications of the choice of $\theta^\star$. For a number of choices of k and $P^\star$, Table 7.1 shows the probability that a randomly drawn $\boldsymbol{p}$ is in the PZ under the uniformity assumption. The table also includes the results of our Monte Carlo (MC) sampling experiment to verify our equations. For that experiment, we generated one million realizations of $\boldsymbol{p} \in \Omega$ under the uniformity assumption and calculated the fraction of $\boldsymbol{p} \in \Omega_{\mathrm{PZ}}$, or $\hat{p} \equiv \hat{\mathsf{P}}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{PZ}})$, as well as the standard error, s.e.$(\hat{p}) \equiv \sqrt{(1-\hat{p})\,\hat{p}/10^6}$. All MC results are within two standard errors of our calculations,

Table 7.1: MC Sampling Results for $\mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{PZ}})$

k	Result	$\theta^\star = 1.6$	$\theta^\star = 2$	$\theta^\star = 2.4$	$\theta^\star = 3$
2	p	0.7692	0.6667	0.5882	0.5000
	$\hat{p}$	0.7697	0.6668	0.5884	0.4998
	s.e.($\hat{p}$)	0.0004	0.0005	0.0005	0.0005
3	p	0.6410	0.5000	0.4011	0.3000
	$\hat{p}$	0.6412	0.5004	0.4012	0.2992
	s.e.($\hat{p}$)	0.0005	0.0005	0.0005	0.0005
4	p	0.5574	0.4000	0.2971	0.2000
	$\hat{p}$	0.5575	0.4002	0.2974	0.2000
	s.e.($\hat{p}$)	0.0005	0.0005	0.0005	0.0004
5	p	0.4977	0.3333	0.2321	0.1429
	$\hat{p}$	0.4984	0.3340	0.2317	0.1434
	s.e.($\hat{p}$)	0.0005	0.0005	0.0004	0.0004
10	p	0.3429	0.1818	0.1013	0.0455
	$\hat{p}$	0.3426	0.1822	0.1009	0.0451
	s.e.($\hat{p}$)	0.0005	0.0004	0.0003	0.0002

except that of $k = 10$ and $\theta^\star = 3$, the smallest PZ, which is within three standard errors. Table 7.1 shows, for an example with just three alternatives and $\theta^\star = 2$, that the probability of being in the PZ is $1/2$ — half of all possible alternative configurations. For the same example with five alternatives, only $1/3$ of the possible configurations would be in the PZ.

7.1.4 Size of the IZ Regions

Having solved for the size of the PZ for general k, it is natural to ask about the sizes of the IZ regions as well. Now we are interested in determining the probability that a randomly drawn configuration $\boldsymbol{p}$ is in the region of the IZ in which the decision-maker is indifferent to selection among the m best alternatives, where m is the largest integer for which $p_{[k]}/p_{[k-m+1]} < \theta^\star$, $m = 1, \ldots, k$. We denote that subregion $\Omega_{\mathrm{IZ}m}$, where $\Omega_{\mathrm{IZ1}} = \Omega_{\mathrm{PZ}}$. We make the same assumption that all probability configurations $\boldsymbol{p} \in \Omega$ are equally likely. Therefore, we can again use geometric volumes to calculate the probabilities we require, as we did in §7.1. We were able to do this when $m = 1$ for general k in §7.1, but here the complexities of the subregions do not allow us to generalize for k. Therefore, we derived the volumes directly for $k = 3$ and $k = 4$ in Appendix D. Table 7.2 shows similar information

Table 7.2: MC Sampling Results for $\mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{IZ}})$

k	Region	Result	$\theta^\star = 1.6$	$\theta^\star = 2$	$\theta^\star = 2.4$	$\theta^\star = 3$
3	Ω_{IZ2}	p	0.3114	0.4000	0.4453	0.4714
		$\hat{p}$	0.3113	0.3995	0.4446	0.4723
		s.e.$(\hat{p})$	0.0005	0.0005	0.0005	0.0005
	Ω_{IZ3}	p	0.0476	0.1000	0.1536	0.2286
		$\hat{p}$	0.0474	0.1001	0.1542	0.2284
		s.e.$(\hat{p})$	0.0002	0.0003	0.0004	0.0004
4	Ω_{IZ2}	p	0.3424	0.4000	0.4088	0.3857
		$\hat{p}$	0.3424	0.4001	0.4087	0.3860
		s.e.$(\hat{p})$	0.0005	0.0005	0.0005	0.0005
	Ω_{IZ3}	p	0.0908	0.1714	0.2395	0.3143
		$\hat{p}$	0.0909	0.1714	0.2394	0.3147
		s.e.$(\hat{p})$	0.0003	0.0004	0.0004	0.0005
	Ω_{IZ4}	p	0.0093	0.0286	0.0547	0.1000
		$\hat{p}$	0.0092	0.0283	0.0545	0.0994
		s.e.$(\hat{p})$	0.0001	0.0002	0.0002	0.0003

to that shown in Table 7.1; it includes the results of the same MC experiment of one million probability configurations, but for the IZ. Note that although the IZ grows with k and $\theta^\star$, individual regions may not be monotonically increasing. For example, Ω_{IZ2} when $k = 4$ increases in $\theta^\star$ initially but then decreases again.

7.2 Probability of Acceptable Selection

The probability guarantees of conditions (1.1) and (1.2) apply only when $\boldsymbol{p} \in \Omega_{\mathrm{PZ}}$. They are based upon the $\mathsf{P}(\mathrm{CS})$ for the procedures in the LFC, which by definition is really the LFC *in the PZ*. Clearly, the LFC for the whole of Ω is the configuration in which the best alternative has a probability of success very slightly greater than $1/k$, i.e., a configuration very close to the EPC. In that case, we would expect $\mathsf{P}(\mathrm{CS}) \approx 1/k$, regardless of the procedure.

Given the results in §7.1.3, one might ask what the probability guarantees are in general, or at least when the probability configuration is *not* in the PZ. For that, we must define a new concept. In our notation, m is the same as it was in the previous section. When $\boldsymbol{p} \in \Omega_{\mathrm{PZ}}$, our goal has been to select the alternative associated with $p_{[k]}$, and we say that a

correct selection is made if the goal is achieved. We will now say that an *acceptable selection* (AS) is made if we select any alternative associated with $p_{[k-i+1]}$, $1 \leq i \leq m$ (i.e., if we select among any of the m best alternatives). We use $\mathsf{P}_{\boldsymbol{p}}(\text{AS})$ to denote the probability of making an acceptable solution given probability configuration $\boldsymbol{p}$, or $\mathsf{P}(\text{AS})$ for short. In the PZ, $\mathsf{P}(\text{CS}) = \mathsf{P}(\text{AS})$.

Suppose we wish to extend our original requirement in (1.1) to the following. For user-specified constants $(\theta^\star, P^\star)$ with $\theta^\star > 1$ and $1/k < P^\star < 1$, we require

$$\mathsf{P}_{\boldsymbol{p}}(\text{AS}) \;\geq\; P^\star \;\text{ for all }\; \boldsymbol{p} \in \Omega. \tag{7.10}$$

Intuitively, it would seem that (7.10) follows from (1.1); otherwise, the $\mathsf{P}(\text{CS})$-requirement in (1.1) would be unsatisfying given no (realistic) guarantee that $\boldsymbol{p} \in \Omega_{\text{PZ}}$.

To help us understand the probability guarantee, we initially considered each procedure's LFC, in terms of $\mathsf{P}(\text{AS})$ *in the IZ*. In previous chapters, we have already discussed the complexities of proving that the SC is the LFC for $\mathsf{P}(\text{CS})$ in the PZ; nonetheless, all procedures have either proven or conjectured that it is so. In the IZ, the SC is a vertex for every subregion $\{\Omega_{\text{IZ}m} : m = 2, \ldots, k\}$. Our initial results led us to hypothesize that the LFC, in terms of $\mathsf{P}(\text{AS})$, for each subregion of the sample space, $\Omega_{\text{IZ}m}$, $m = 2, \ldots, k$, is the SC.

To test our hypothesis, we conducted the following set of MC sampling experiments. For each of the 36 cases shown in Appendix B and for each of the eight procedures compared there, we randomly sampled 100,000 realizations of $\boldsymbol{p}$ uniformly in Ω and calculated $\mathsf{P}(\text{AS})$, $\mathsf{P}(\text{CS})$, and $\mathsf{E}[N]$ for each $\boldsymbol{p}$. Thus, we conducted 268 experiments of 100,000 replications each. (The actual number of experiments is reduced from 288 since certain procedures were infeasible for particular problems and some problems were too large to obtain solutions.) We separated the $\boldsymbol{p}$-vectors into their appropriate regions, $\{\Omega_{\text{IZ}m} : m = 2, \ldots, k\}$, and analyzed the results. In all but one of the 268 experiments, our hypothesis was true; however, we discovered one case in which it was not. Further exploration and analysis led us to reject our hypothesis. Fortunately, our experiments did confirm that $\mathsf{P}_{\boldsymbol{p}}(\text{AS}) \geq \mathsf{P}_{\text{SC}}(\text{CS})$ for all $\Omega_{\text{IZ}m}, m = 1, 2, \ldots, k$, implying that probability requirement (1.1) guarantees probability

Table 7.3: Example of $\mathsf{P}_{\boldsymbol{p}}(\text{AS}) < \mathsf{P}_{\text{SC}}(\text{AS})$ for $k = 3$, $b = 3$, and $\theta^\star = 15$

$\boldsymbol{\eta}$	$\mathsf{P}_{\boldsymbol{p}}$(arrive at $\boldsymbol{\eta}$ and stop)	P_{SC}(arrive at $\boldsymbol{\eta}$ and stop)	P(AS | stop at $\boldsymbol{\eta}$)
$(0,0,3)$	0.00011	0.00020	0
$(0,1,2)$	0.00162	0.00061	0
$(0,2,1)$	0.00810	0.00061	1
$(0,3,0)$	0.01350	0.00020	1
$(1,0,2)$	0.00486	0.00916	0
$(1,1,1)$	0.04859	0.01832	2/3
$(1,2,0)$	0.12148	0.00916	1
$(2,0,1)$	0.07289	0.13739	1
$(2,1,0)$	0.36443	0.13739	1
$(3,0,0)$	0.36443	0.68695	1
	$\mathsf{P}_{\boldsymbol{p}}(\text{AS})$=0.9772	$\mathsf{P}_{\text{SC}}(\text{AS})$=0.9839	

Table 7.4: MC Sampling $\mathsf{P}(\text{AS})$ Results for Procedure $\mathcal{M}_{\text{R}}$

k	$P^\star$	$\theta^\star = 1.6$			$\theta^\star = 2.4$		
		$\overline{\mathsf{P}(\text{CS})}$	$\overline{\mathsf{P}(\text{AS})}$	$\min \mathsf{P}(\text{AS})$	$\overline{\mathsf{P}(\text{CS})}$	$\overline{\mathsf{P}(\text{AS})}$	$\min \mathsf{P}(\text{AS})$
2	0.75	0.9408	0.9545	1.0000	0.8908	0.9356	1.0000
	0.90	0.9881	0.9909	1.0000	0.9783	0.9873	1.0000
	0.95	0.9960	0.9969	1.0000	0.9921	0.9954	1.0000
3	0.75	0.9610	0.9710	0.8764	0.9310	0.9604	0.8759
	0.90	0.9901	0.9929	0.9511	0.9819	0.9902	0.9506
	0.95	0.9957	0.9969	0.9754	0.9932	0.9963	0.9752
4	0.75	0.9648	0.9751	0.8500	0.9371	0.9666	0.8399
	0.90	0.9904	0.9935	0.9433	0.9840	0.9920	0.9365
	0.95	N/A	N/A	N/A	0.9934	0.9968	0.9707

requirement (7.10), at least for the configurations sampled.

Table 7.3 is an example of a situation in which the SC is not the LFC in the IZ. It shows the results for an MSP with $k = 3$ and (very large) $\theta^\star = 15$ for Procedure $\mathcal{M}_{\text{BEM}}$ with $n_{\text{BEM}} = 3$. For that example, $\boldsymbol{p} = (15/21, 5/21, 1/21) \in \Omega_{\text{IZ2}}$, and we assume that the SC $= (15/17, 1/17, 1/17)$ is also in Ω_{IZ2} (for calculation purposes). Each row in the table represents a stopping point for the procedure. The last column, labeled '$\mathsf{P}(\text{AS} \mid \text{stop at } \boldsymbol{\eta})$' is the probability that an AS is made if the experiment stops at the node in the first column. The last row has the $\mathsf{P}(\text{AS})$ results, calculated as the vector product of the column above it and the last column.

We also examined expected $\mathsf{P}(\text{CS})$ and $\mathsf{P}(\text{AS})$. Such calculations require knowledge of the distribution of $\boldsymbol{p} \in \Omega_{\text{PZ}}$ for the former and $\boldsymbol{p} \in \Omega$ for the latter. To develop some insights, we assume our model in which the $\boldsymbol{p}$ are distributed uniformly in Ω. We derive our results from the same MC experiments previously discussed in this subsection. Table 7.4 shows results for the optimal randomized Procedure $\mathcal{M}_{\text{R}}$ under selected values for k and $\theta^{\star}$. It includes $\overline{\mathsf{P}(\text{CS})}$, the mean of $\mathsf{P}(\text{CS})$ across all MC samples $\boldsymbol{p} \in \Omega_{\text{PZ}}$; $\overline{\mathsf{P}(\text{AS})}$, the mean of $\mathsf{P}(\text{AS})$ across all MC samples $\boldsymbol{p} \in \Omega$; and $\min \mathsf{P}(\text{AS})$, the minimum $\mathsf{P}(\text{AS})$ across all MC samples *in the IZ*, since the minimum in the PZ equals $\mathsf{P}_{\text{SC}}(\text{CS})$. Our MC experiments also support our assumption that the expected number of observations is maximized in the EPC, although we do not show the results here.

7.3 Conditional (Posterior) $\mathsf{P}(\text{CS})$

As we noted in Chapter 1, our research focuses on prior probability requirements, which are also the primary concern in the MSP literature. In this section, we will examine posterior probabilities, in particular, the conditional probability of correct selection upon termination of the experiment.

7.3.1 Relationships

We will use $\mathsf{P}(\text{CS} \,|\, \boldsymbol{\eta})$ to denote the conditional probability of a CS given that we have arrived at cumulative success vector $\boldsymbol{\eta}$ and stopped. The conditional probability is then

$$\mathsf{P}(\text{CS} \,|\, \boldsymbol{\eta}) = \frac{\mathsf{P}(\text{arrive at } \boldsymbol{\eta}\text{, stop, and make a CS})}{\mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and stop})}. \tag{7.11}$$

We now let $\boldsymbol{y}$ be a permutation of $\boldsymbol{\eta}$. Borrowing the notation of Chapter 4, we let $t(\boldsymbol{y})$ represent the number of alternatives in cumulative success vector $\boldsymbol{y}$ that have the same number of successes as the best alternative, including the best. Then $1 \leq t(\boldsymbol{y}) \leq k$ with $t(\boldsymbol{y}) = 1$ when none are tied with the best. Additionally, let $\pi^{\star}(\boldsymbol{y})$ be the subset of the permutations of $\boldsymbol{y}$ for which the first alternative is greater than or equal to the remaining alternatives. In other words, $\pi^{\star}(\boldsymbol{y}) \subseteq \pi(\boldsymbol{y})$ is the subset of permutations that results in correct selections. Then

$$\mathsf{P}(\text{CS} \,|\, \boldsymbol{\eta}) = \frac{\sum_{\boldsymbol{y} \in \pi^{\star}(\boldsymbol{\eta})} \frac{1}{t(\boldsymbol{y})}\, p_{\boldsymbol{y}}\, \mathsf{P}(\text{arrive at } \boldsymbol{y})}{\sum_{\boldsymbol{y} \in \pi(\boldsymbol{\eta})} p_{\boldsymbol{y}}\, \mathsf{P}(\text{arrive at } \boldsymbol{y})}.$$

where $p_{\boldsymbol{y}} \equiv \mathsf{P}(\text{stop at } \boldsymbol{y} \,|\, \text{arrive at } \boldsymbol{y})$ and $\boldsymbol{y}$ is an index into the respective set of vectors. But $p_{\boldsymbol{y}} = p_{\boldsymbol{\eta}}$ for all $\boldsymbol{y} \in \pi(\boldsymbol{\eta})$, and $t(\boldsymbol{y}) = t(\boldsymbol{\eta})$ for all $\boldsymbol{y} \in \pi^\star(\boldsymbol{\eta})$, so

$$\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta}) = \frac{\frac{1}{t(\boldsymbol{\eta})}\, p_{\boldsymbol{\eta}} \sum_{\boldsymbol{y} \in \pi^\star(\boldsymbol{\eta})} \mathsf{P}(\text{arrive at } \boldsymbol{y})}{p_{\boldsymbol{\eta}} \sum_{\boldsymbol{y} \in \pi(\boldsymbol{\eta})} \mathsf{P}(\text{arrive at } \boldsymbol{y})}. \tag{7.12}$$

For simplicity and without loss of generality, we assume a $\boldsymbol{p}$ in which $p_1 \geq p_2 \geq \cdots \geq p_k$. Now consider a general probability configuration $\boldsymbol{p} = (q, \frac{q}{\theta_2}, \frac{q}{\theta_3}, \ldots, \frac{q}{\theta_k})$, where $\theta_i = p_1/p_i$, and let $S_{\boldsymbol{y}}$ be the number of sample paths terminating at cumulative success vector $\boldsymbol{y}$. After canceling the $p_{\boldsymbol{\eta}}$ and including $\boldsymbol{p}$, (7.12) becomes:

$$\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta}) = \frac{\frac{1}{t(\boldsymbol{\eta})} \sum_{\boldsymbol{y} \in \pi^\star(\boldsymbol{\eta})} S_{\boldsymbol{y}}\, q^{\sum_{i=1}^k y_i} \prod_{i=1}^k \left(\frac{1}{\theta_i}\right)^{y_i}}{\sum_{\boldsymbol{y} \in \pi(\boldsymbol{\eta})} S_{\boldsymbol{y}}\, q^{\sum_{i=1}^k y_i} \prod_{i=1}^k \left(\frac{1}{\theta_i}\right)^{y_i}}. \tag{7.13}$$

Since $S_{\boldsymbol{y}}$ is equivalent for all $\boldsymbol{y} \in \pi(\boldsymbol{\eta})$ and the term $q^{\sum_{i=1}^k y_i}$ cancels from all terms,

$$\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta}) = \frac{\frac{1}{t(\boldsymbol{\eta})} \sum_{\boldsymbol{y} \in \pi^\star(\boldsymbol{\eta})} \prod_{i=1}^k \left(\frac{1}{\theta_i}\right)^{y_i}}{\sum_{\boldsymbol{y} \in \pi(\boldsymbol{\eta})} \prod_{i=1}^k \left(\frac{1}{\theta_i}\right)^{y_i}}. \tag{7.14}$$

If $\boldsymbol{p}$ is the SC, then $\theta_i = \theta^\star$ for all $i > 1$, and $\boldsymbol{p} = (\theta^\star p, p, \ldots, p)$. Then (7.12) becomes

$$\begin{aligned} \mathsf{P}_{\mathrm{SC}}(\mathrm{CS} \,|\, \boldsymbol{\eta}) &= \frac{\frac{1}{t(\boldsymbol{\eta})} \sum_{\boldsymbol{y} \in \pi^\star(\boldsymbol{\eta})} S_{\boldsymbol{y}}\, p^{\sum_{i=1}^k y_i}\ (\theta^\star)^{y_1}}{\sum_{\boldsymbol{y} \in \pi(\boldsymbol{\eta})} S_{\boldsymbol{y}}\, p^{\sum_{i=1}^k y_i}\ (\theta^\star)^{y_1}} \\ &= \frac{\frac{1}{t(\boldsymbol{\eta})} \sum_{\boldsymbol{y} \in \pi^\star(\boldsymbol{\eta})} (\theta^\star)^{y_1}}{\sum_{\boldsymbol{y} \in \pi(\boldsymbol{\eta})} (\theta^\star)^{y_1}}. \end{aligned} \tag{7.15}$$

In Chapter 1, we state that $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$ is the expectation of the prior $\mathsf{P}(\mathrm{CS})$. We establish that relationship here. Recall from our development of the $\mathsf{P}(\mathrm{CS})$ constraint in Chapter 3 that

$$\begin{aligned} \mathsf{P}(\mathrm{CS}) &= \sum_{\substack{\boldsymbol{\eta} \in \mathcal{N} \\ \boldsymbol{\eta} \text{ is CS}}} \frac{1}{t(\boldsymbol{\eta})}\, \mathsf{P}(\text{arrive at } \boldsymbol{\eta} \text{ and stop at } \boldsymbol{\eta}) \\ &= \sum_{\substack{\boldsymbol{\eta} \in \mathcal{N} \\ \boldsymbol{\eta} \text{ is CS}}} \frac{1}{t(\boldsymbol{\eta})}\, p_{\boldsymbol{\eta}}\, \mathsf{P}(\text{arrive at } \boldsymbol{\eta}), \end{aligned}$$

where $\mathcal{N}$ is the set of all possible $\boldsymbol{\eta}$. We can use the left-lexicographic representation of $\boldsymbol{\eta}$, $\boldsymbol{\eta}'$, to rewrite $\mathsf{P}(\mathrm{CS})$ as

$$\mathsf{P}(\mathrm{CS}) \;=\; \sum_{\boldsymbol{\eta}' \in \mathcal{N}'} \frac{1}{t(\boldsymbol{\eta}')}\, p_{\boldsymbol{\eta}'} \sum_{\boldsymbol{y} \in \pi^\star(\boldsymbol{\eta}')} \mathsf{P}(\text{arrive at } \boldsymbol{y}).$$

By (7.12), the overall $\mathsf{P}(\mathrm{CS})$ of a procedure is the weighted sum of the conditional probabilities of CS:

$$\mathsf{P}(\mathrm{CS}) = \sum_{\boldsymbol{\eta}' \in \mathcal{N}'} w_{\boldsymbol{\eta}'}\, \mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta}'),$$

where $w_{\boldsymbol{\eta}'} \equiv p_{\boldsymbol{\eta}'} \sum_{\boldsymbol{y} \in \pi(\boldsymbol{\eta}')} \mathsf{P}(\text{arrive at } \boldsymbol{y})$.

7.3.2 Concerns

One criticism that can be leveled against all of the bounded procedures that we have discussed is the fact that we stated in Chapter 1 — the posterior conditional $\mathsf{P}(\mathrm{CS})$ may be less than $P^\star$. For example, consider Procedure $\mathcal{M}_{\mathrm{RA}}$ for an MSP with $k = 3$, $\theta^\star = 2$, and $P^\star = 0.75$. The parameter pair for that particular MSP is $r = 4$ and $t = 5$, which guarantees a $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS}) = 0.7556$. Suppose we stop at $\boldsymbol{\eta} = (5, 3, 3)$. In the SC, (7.15) gives us

$$\begin{aligned} \mathsf{P}_{\mathrm{SC}}(\mathrm{CS} \,|\, \boldsymbol{\eta}) &= \frac{2\,(\theta^\star)^5}{2\,(\theta^\star)^5 + 2\,(\theta^\star)^3 + 2\,(\theta^\star)^3} \\ &= 2/3, \end{aligned}$$

which is less than our desired $P^\star$. While MSP procedures guarantee a prior $\mathsf{P}(\mathrm{CS}) \geq P^\star$ (when $\boldsymbol{p}$ is in the PZ), there are no guarantees on posterior $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$. When presenting a result like that in our example to a decision-maker, the prior probability guarantees may no longer be important, and a low $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$ may erode confidence in a CS.

Given these concerns, we now formally state three necessary conditions for employing bounded MSP procedures.

1. **Repeated application:** The procedure should be repeatedly applied to the same application. While some experiments may result in a $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta}) < P^\star$, in the long term, we should make a correct selection in at least fraction $P^\star$ of the experiments.

2. **Tolerable loss:** The consequences of making an incorrect selection must not be catastrophic. In other words, the decision-maker must be willing to accept a selection even when $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta}) < P^\star$.

3. **Non-trivial observation costs:** There must be a cost associated with the observations such that a reduction in the expected number of observations is desirable.

We give two examples of possible applications that satisfy the above conditions. Consider a manufacturing process for which an important characteristic of product quality is qualitative (or even non-parametric) in nature, e.g., material composition and uniformity, impurity content, etc. Evaluation of the product for this particular characteristic is expensive because it requires the destruction of the product or use of a highly specialized evaluation process. Suppose that we only have access to machine-tuning technicians on a regularly scheduled interval. Their job is to tune the k machines responsible for the characteristic of interest; however, they do not have the time to tune all machines. Therefore, the quality engineer employs MSP procedures to determine which of the k machines is producing parts of the lowest quality just before the technicians' arrival. Thus, a success in this case is the production of the part with the lowest quality. Those experiments determine which machine the technicians will tune. If the wrong machine is occasionally identified as the worst, the consequences are not dire, since it will still benefit from the tuning.

Consider another example in the military domain. War gaming is a process in which military leaders and staff evaluate and refine potential courses of action within a realistic scenario. For the war-gaming process in a training environment, a scenario must be chosen in a way that maximizes the development of the skills of the war-gaming participants (e.g., maximizes participation, challenges participants, etc.). An overall set of k potential scenarios may be vetted with a small group of instructors, but, from that set, a single scenario must be chosen for the students. Unwilling to experiment on the class at large, the k scenarios are war gamed via smaller focus groups, and, at the end of each evaluation, one is chosen as best. An MSP procedure is used to minimize the number of these time-consuming focus groups required to select the best of the k scenarios. The original set of scenarios have

all been vetted by the instructors, so the occasional, incorrect selection of a scenario is not catastrophic.

7.3.3 Incorporation of $\mathsf{P}(\mathbf{CS}\,|\,\boldsymbol{\eta})$

Any bounded MSP procedure will have $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta}) < P^\star$ for a subset of its stopping vectors for non-trivial $P^\star$. For example, it is always possible to follow a sample path in which the largest difference between the number of successes of the best alternative and any other alternative is, at most, one. Since our procedures are bounded, such a sample path must terminate at a potential stopping point. The best case $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$ for such a sample path is when the best alternative has one more success than the remaining alternatives, or when $\boldsymbol{\eta} = (\eta_1, \eta_1 - 1, \ldots, \eta_1 - 1)$. In the SC,

$$\begin{aligned}\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta}) &= \frac{(\theta^\star)^{\eta_1}}{(\theta^\star)^{\eta_1} + (\theta^\star)^{\eta_1 - 1} + \cdots + (\theta^\star)^{\eta_1 - 1}} \\ &= \frac{\theta^\star}{\theta^\star + (k-1)}.\end{aligned}$$

Thus, at best, in the SC, such a sample path has $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta}) \leq \theta^\star/(\theta^\star + (k-1)) = p_1$. If $P^\star > p_1$, we are guaranteed that $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta}) < P^\star$ for at least one of the stopping points. This situation is caused by bounding the procedure, but removing bounds is rarely plausible. There must be some upper limit beyond which a decision-maker is unwilling to spend for additional observations.

If $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$ is a concern for the decision-maker, we can set a threshold on the minimum $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$ that we are willing to accept, short of the budget limit b. In other words, we can design optimal randomized and non-randomized procedures for which stops prior to the observation budget limit must meet a minimum $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$ threshold. Implementation of the LP and MIP formulations of such MSPs is very straightforward.

Consider the LP/MIP formulation in Chapter 4. We must start, however, with the entire set of uncurtailed nodes, $\mathcal{N}'$, instead of the curtailed set, $\mathcal{N}''$. Curtailment allows us to achieve $\mathsf{P}(\mathrm{CS})$ with less observations; however, it may entail a stop at a curtailment node for which $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta}')$ is less than our requirement and prevent us from reaching a node at a later observation that does meet our $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$ requirement. Recall that $\mathcal{B}$ is the set of nodes

at the budget, i.e., $\mathcal{B} = \{\boldsymbol{\eta}' : \sum_{i=1}^{k} \eta_{[i]} = b\}$, now a subset of $\mathcal{N}'$. Let $P_c^\star$ be our $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$ threshold. To adapt our LP/MIP for this new requirement, we do one of the following:

1. Remove all $f_{\boldsymbol{\eta}'},\ \boldsymbol{\eta}' \in \mathcal{N}' \setminus \mathcal{B}$, for which $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta}') < P_c^\star$, or

2. For each $f_{\boldsymbol{\eta}'},\ \boldsymbol{\eta}' \in \mathcal{N}' \setminus \mathcal{B}$, for which $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta}') < P_c^\star$, add the constraint $f_{\boldsymbol{\eta}'} = 0$.

Both choices are acceptable. In either case, we are removing, as potential stops, the non-budget nodes that fail to meet our $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$ threshold by removing the arc flowing from the node to the sink.

What happens if we set $P_c^\star = P^\star$? Consider any potential stop $\boldsymbol{\eta}'$ in the SC, and define a *group* as a set of vector $\boldsymbol{\eta}'$ components that are equal. Let m_i represent the number of elements in the ith group. For example, consider $\boldsymbol{\eta}' = (4, 4, 2, 1, 1)$ with $k = 5$. In that case, we have three groups with $m_1 = 2,\ m_2 = 1$, and $m_3 = 2$. For general k, we assume there are n groups; note that $m_1 = t(\boldsymbol{\eta}')$. We incorporate this new notation for counting permutations into (7.15), and get

$$\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta}') = \frac{\dfrac{1}{t(\boldsymbol{\eta}')} \left(\begin{array}{c} (k-1)! \\ \hline (t(\boldsymbol{\eta}')-1)! m_2! \cdots m_n! \end{array} \right) (\theta^\star)^{\eta_{[k]}}}{\displaystyle\sum_{j=1}^{n} \left(\begin{array}{c} (k-1)! \\ \hline m_1! m_2! \cdots m_n! \end{array} \right) m_j \left(\theta^\star\right)^{\eta_{[k-\sum_{\ell=1}^{j} m_\ell + 1]}}},$$

which simplifies to

$$\begin{aligned} \mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta}') &= \frac{(\theta^\star)^{\eta_{[k]}}}{\sum_{j=1}^{n} m_j \left(\theta^\star\right)^{\eta_{[k-\sum_{\ell=1}^{j} m_\ell + 1]}}}, \\ &= \frac{(\theta^\star)^{\eta_{[k]}}}{\sum_{i=1}^{k} (\theta^\star)^{\eta_{[k-i+1]}}}, \\ &= \frac{(\theta^\star)^{\eta_{[k]}}}{\sum_{i=1}^{k} (\theta^\star)^{\eta_{[i]}}}, \end{aligned} \tag{7.16}$$

regardless of the number of tied alternatives, or the size of the within-vector groups. Now

we include the $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$ threshold:

$$\begin{aligned}
\frac{(\theta^\star)^{\eta_{[k]}}}{\sum_{i=1}^{k}(\theta^\star)^{\eta_{[i]}}} &\geq P_c^\star, \\
\frac{1}{\sum_{i=1}^{k}(\theta^\star)^{\eta_{[i]}-\eta_{[k]}}} &\geq P_c^\star, \\
1+\sum_{i=1}^{k-1}(\theta^\star)^{\eta_{[i]}-\eta_{[k]}} &\leq \frac{1}{P_c^\star}, \\
\sum_{i=1}^{k-1}(\theta^\star)^{\eta_{[i]}-\eta_{[k]}} &\leq \frac{1-P_c^\star}{P_c^\star}.
\end{aligned}$$

If we substitute our original notation for the ordered cumulative success vector, we get

$$\sum_{i=1}^{k-1}\left(\frac{1}{\theta^\star}\right)^{\eta_{[k]m}-\eta_{[i]m}} \leq \frac{1-P_c^\star}{P_c^\star}, \tag{7.17}$$

which is the stopping condition, z_m, for Procedure $\mathcal{M}_{\mathrm{BG}}$. In fact, we see that the original unbounded procedure of BKS (1968) guaranteed $\mathsf{P}(\mathrm{CS})$ by ensuring that the procedure only stops when $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta}') \geq P^\star$. When BG (1985b) originally truncated the procedure of BKS, they did so without curtailment. Thus, if we set our $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$ threshold to $P^\star$, we will get the original procedure developed by BG just prior to their addition of curtailment in Procedure $\mathcal{M}_{\mathrm{BG}}$.

Before we conclude this discussion, we consider the concept of curtailment applied to $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$. It is possible to consider the same MSP, but with the following provision. We will allow stopping nodes $\boldsymbol{\eta}' \notin \mathcal{B}$ that do not meet the $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$ threshold, only if $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$ cannot be achieved even if we take the remaining observations in the budget. This $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$-curtailment is a plausible modification to our MSP and is easily implementable in the LP/MIP formulations. On the other hand, it is probably reasonable to assume that if the decision-maker is concerned at all with posterior $\mathsf{P}(\mathrm{CS})$, then he will want to continue sampling beyond such a node in order to maximize $\mathsf{P}(\mathrm{CS}\,|\,\boldsymbol{\eta})$, or to at least have more information about the alternatives. Therefore, we did not implement this type of curtailment.

7.4 Summary

Our initial insights here are ideal launch points for continued research that will deepen our understanding not only of MSP procedures, but also other types of RS procedures as well.

While we have a good understanding now about $\mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{PZ}})$ and $\mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{IZ}})$, when $\boldsymbol{p}$ are distributed uniformly in Ω, we might have an interest in other types of distributions for $\boldsymbol{p}$, say, for example, the Dirichlet distribution. Furthermore, these same questions concerning the PZ apply as well to Bernoulli selection problems that use the relative risk ratio IZ parameter, as well as other RS problems that use IZ methods.

Our concept of $\mathsf{P}(\mathrm{AS})$ allows us to look beyond the guarantees for $\boldsymbol{p}$ in the PZ to overall guarantees for $\boldsymbol{p} \in \Omega$. We have shown through MC sampling that $\mathsf{P}(\mathrm{AS}) \geq P^\star$ for all $\boldsymbol{p} \in \Omega$; however, it would be helpful to identify the LFC, in terms of $\mathsf{P}(\mathrm{AS})$, in each of the IZ regions. Such knowledge would give us stronger probabilistic guarantees. We have also given examples of the mean $\mathsf{P}(\mathrm{AS})$ based upon MC sampling under our uniformity assumption. The concept of $\mathsf{P}(\mathrm{AS})$ applies to all IZ-based RS procedures and could be extended to those as well.

Finally, we addressed posterior $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$. Based upon concerns that $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$ will be less than $P^\star$ for particular stopping points under bounded MSP procedures, we have identified three necessary conditions for applying the types of MSP procedures we have discussed in this thesis. We have also developed a capability to set a minimum threshold on $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$, and identify the optimal randomized [non-randomized] procedures via a modified LP [MIP] formulation. Concerns about $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$ extend well beyond MSP procedures to the broader field of RS procedures.

CHAPTER VIII

CONCLUSIONS

This research has advanced the existing state of the art by leveraging LP and MIP mathematical programming techniques to develop optimal procedures for MSPs with a fixed sampling budget, as we demonstrated in Chapter 3. By construction, our procedures always perform at least as well as existing MSP procedures. We also introduce a new type of procedure — the randomized procedure. Randomized procedures provide the same guarantees as non-randomized procedures, but are always more efficient when optimized for the problem of interest, and their LPs are much easier to solve than the corresponding MIP for the non-randomized procedures.

Our reformulations of the MIP and LP mathematical programs in Chapter 4 overcome some of the most significant drawbacks of our initial formulations. Problem symmetries and strong curtailment provide us the necessary relationships to represent the entire network of possible stopping nodes by a smaller network consisting only of left-lexicographic, non-curtailed nodes. Although they are less intuitive, the reformulations significantly reduce the computational requirements for solving the problem and have allowed us to consider substantially larger problems than could be solved initially.

Those reformulations allowed us to prove some key results about the optimal randomized MSP procedure, in particular, that an optimal procedure exists which has, at most, only one randomized node, and we provided an algorithm for finding it. We then showed that, for the optimal solution, $\mathsf{P}(\mathrm{CS}) = P^\star$ when $P^\star \geq 1/k$. We were also able to extend our formulations to replicate existing MSP procedures, thereby facilitating the identification of the optimal procedure parameters for a specific problem. This can be a significant capability if an experimenter often needs to identify such parameters for problem characteristics that do not have tabulated results.

In Chapter 5, to consider the realistic scenario of variable marginal observation costs, we

developed LP and MIP formulations that, when solved, provide the optimal randomized and non-randomized procedures, respectively, for a given cost function. These formulations are built upon reasonable assumptions with respect to the observation costs. Most importantly, ours is the first methodology in the field of RS that integrates such costs. Furthermore, using these formulations, we have been able to show that the type and shape of the total cost functions have an impact, often significant, on a procedure's efficiency with respect to the optimal solution. We examined a robust set of possible cost functions, including functions affected by periodic observation costs. In some cases, we gave general insights that applied across function types. Most importantly, we show that incorrectly assuming that total costs are linear leads to unnecessary additional costs in expectation. We were also able to provide two additional tools for addressing particular cost-related issues — one to inform observation batching decisions and the other to estimate the cost of changing the $P^\star$ requirement using dual variables.

In Chapter 6, we showed that, in many cases, the expected number of observations for Procedure $\mathcal{M}_{\mathrm{NR}}$ rapidly approaches, as a function of b, the expected number of observations for Procedure $\mathcal{M}_{\mathrm{R}}$. For lower values of $P^\star$ coupled with larger $\theta^\star$ (i.e., when few observations are required to achieve $P^\star$), the optimal Procedures $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{NR}}$ do not converge to each other, but remain equidistant as b increases.

We developed a number of metrics to examine select aspects of procedure performance, and used those metrics and representative charts to demonstrate some important relationships between the procedures. We looked further at the information provided by the distribution of N for each MSP procedure, in particular, the variability. We also demonstrated the relationships between Alam's (1971) procedure, Procedure $\mathcal{M}_{\mathrm{RA}}$, and the classic gambler's ruin problem.

Finally, in Chapter 7, we examined probabilities with respect to the PZ and IZ when the possible probability configurations are distributed uniformly in the sample space Ω, shedding light on the implications and interpretations of the indifference zone parameter, $\theta^\star$. We introduced the concept of $\mathsf{P}(\mathrm{AS})$, which allows us to consider probability guarantees for all $\boldsymbol{p} \in \Omega$. We have shown, through MC sampling, that $\mathsf{P}(\mathrm{AS}) \geq P^\star$ for all $\boldsymbol{p} \in \Omega$.

We also examined the posterior conditional probability of correct selection upon procedure termination, $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$. Based on concerns that $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$ will be less than $P^\star$ for particular stopping points in bounded MSP procedures, we have identified three necessary conditions for applying MSP procedures when we are concerned about $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$. We have also developed a capability to set a minimum threshold on $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$, and to identify the optimal randomized [non-randomized] procedures via a modified LP [MIP] formulation.

8.1 Contribution Summary

- Developed LP and MIP formulations of the MSP that guarantee the identification of the optimal stopping rules for any given MSP and probability configuration, and introduced the concept of a randomized procedure.

- Leveraged the characteristics of the MSP to refine the mathematical programs, thereby improving their algorithmic efficiency and facilitating the identification of key properties of the resulting optimal randomized procedures.

- Reformulated the mathematical programs to identify the optimal stopping rules under variable observation cost functions, and developed insights into the effects of variable observation cost functions on procedure performance.

- Conducted a thorough analysis of the performance of existing procedures with respect to the optimal randomized procedure, as well as the relative performance of the optimal randomized and non-randomized procedures.

- Examined key assumptions concerning the indifference zone parameter and the conditional probability of correct selection, resulting in novel insights and potential directions for future research.

8.2 Future Efforts

There are numerous potential extensions of our methodologies and formulations in the context of MSPs. For example, modifications to our mathematical programs could be used

to consider aspects of the subset selection problem or the nuisance alternative problem described by Aoshima and Chen (1999). Our formulations could also be modified to model Bernoulli selection problems; however, the problem of the rapidly increasing network size would have to be addressed for most realistic Bernoulli applications. In general, the concept of randomized stopping points has application beyond MSPs to other types of RS procedures.

The issue of variable observation costs merits further exploration. Our formulations provide the capability to assign unique costs to every possible cumulative success vector, if such an application exists. Furthermore, variable observation costs are possible for any type of experiment, not just MSPs or even RS problems. Future efforts should include the development of realistic case studies and the adaptation of existing procedures and experimental designs to account for such costs.

Our work in Chapter 7 concerning the PZ can be expanded to consider other types of probability configuration distributions, such as the Dirichlet distribution. Furthermore, the related concept of the probability of acceptable selection should be further studied to gain a better understanding of the general probabilistic guarantees of MSP procedures. These same questions extend beyond MSPs to other types of indifference zone RS problems as well.

The issue of prior $\mathsf{P}(\mathrm{CS})$ guarantees versus posterior $\mathsf{P}(\mathrm{CS} \,|\, \boldsymbol{\eta})$ results is not unique to MSP procedures. Such concerns, though very rarely addressed, if at all, in the MSP literature, have a significant impact on decision-maker acceptance of experimental design plans and results in the context of RS. Such concerns must be addressed to make RS procedures, particularly MSP procedures, more palatable for decision-makers.

Our sincere hope is that the ideas pursued in this research — formulating RS problems as mathematical programs, considering variable observation costs, challenging fundamental assumptions — are not just used to improve the application of MSP procedures, but are carried beyond MSPs to the broader fields of RS and experimental design.

APPENDIX A

UPDATED PROCEDURE TABLES

This appendix includes updated tables for Procedures $\mathcal{M}_{\mathrm{BK}}$, $\mathcal{M}_{\mathrm{C}'}$, $\mathcal{M}_{\mathrm{RA}}$, and $\mathcal{M}_{\mathrm{RA}'}$.

A.1 Updated Tables for Procedure $\mathcal{M}_{\mathrm{BK}}$

Table A.1 identifies the n_{BK}-values that minimize $\mathsf{E}_{\mathrm{SC}}[N]$ while still achieving $P^\star$. We searched all n_{BK}-values up to $n_{\mathrm{BK}} = 400$. Table entries with ">400" in the column for n_{BK} indicate $P^\star$ requirements that cannot be achieved within our search space for the given k and $\theta^\star$.

Bechhofer and Kulkarni (1984) focus on proving various theorems and lemmas associated with curtailment, not on providing tables for the user. Their tables only include results for $n_{\mathrm{BK}} \leq 20$ and are tabulated by k, n_{BK}, and $\theta^\star$. We supplement those tables by tabulating the data for common choices of $P^\star$, including a greater range of $\theta^\star$-values, and searching over a much larger search space for n_{BK}. We also provide the expected number of observations in the EPC (i.e., $\mathsf{E}_{\mathrm{EPC}}[N]$).

Table A.1: Updated Performance Characteristics for Procedure $\mathcal{M}_{\text{BK}}$

$P^\star$	$\theta^\star$	$k=2$				$k=3$				$k=4$			
		n_{BK}	$\mathsf{P}_{\text{SC}}(\text{CS})$	$\mathsf{E}_{\text{SC}}[N]$	$\mathsf{E}_{\text{EPC}}[N]$	n_{BK}	$\mathsf{P}_{\text{SC}}(\text{CS})$	$\mathsf{E}_{\text{SC}}[N]$	$\mathsf{E}_{\text{EPC}}[N]$	n_{BK}	$\mathsf{P}_{\text{SC}}(\text{CS})$	$\mathsf{E}_{\text{SC}}[N]$	$\mathsf{E}_{\text{EPC}}[N]$
0.95	3.0	9	0.9511	6.540	7.539	17	0.9554	12.958	15.079	23	0.9527	18.523	21.215
	2.8	11	0.9565	8.021	9.293	19	0.9535	14.687	16.946	26	0.9508	21.216	24.093
	2.6	13	0.9577	9.565	11.067	22	0.9527	17.273	19.784	31	0.9525	25.657	28.926
	2.4	15	0.9552	11.188	12.858	26	0.9511	20.781	23.596	37	0.9512	31.115	34.731
	2.2	19	0.9573	14.391	16.476	32	0.9502	26.091	29.321	46	0.9505	39.362	43.469
	2.0	23	0.9520	17.806	20.132	42	0.9509	35.023	38.921	61	0.9513	53.207	58.092
	1.8	33	0.9544	26.227	29.382	59	0.9508	50.489	55.325	86	0.9504	76.641	82.551
	1.6	49	0.9501	40.331	44.386	93	0.9502	82.011	88.365	138	0.9506	125.956	133.642
	1.4	97	0.9513	83.599	90.121	185	0.9503	168.952	178.424	277	0.9504	259.708	270.845
	1.2	327	0.9504	299.917	313.561	>400				>400			
0.9	3.0	7	0.9294	5.163	5.813	11	0.9014	8.460	9.482	16	0.9024	12.969	14.493
	2.8	7	0.9167	5.222	5.813	13	0.9073	10.103	11.312	19	0.9076	15.602	17.380
	2.6	7	0.9009	5.286	5.813	15	0.9054	11.833	13.222	22	0.9052	18.312	20.245
	2.4	9	0.9082	6.823	7.539	18	0.9056	14.436	16.035	26	0.9017	21.980	24.093
	2.2	11	0.9068	8.435	9.293	22	0.9034	17.985	19.784	33	0.9040	28.372	30.857
	2.0	15	0.9118	11.681	12.858	29	0.9044	24.242	26.455	43	0.9022	37.669	40.557
	1.8	19	0.9013	15.146	16.476	40	0.9018	34.299	36.984	61	0.9020	54.557	58.092
	1.6	31	0.9054	25.505	27.522	63	0.9007	55.643	59.203	97	0.9005	88.796	93.341
	1.4	59	0.9023	50.720	53.845	125	0.9004	114.286	119.609	195	0.9003	183.217	189.829
	1.2	199	0.9009	182.011	188.730	>400				>400			
0.75	3.0	1	0.7500	1.000	1.000	5	0.7690	3.950	4.111	8	0.7701	6.430	6.911
	2.8	3	0.8287	2.388	2.500	6	0.7803	4.493	4.926	9	0.7642	7.394	7.883
	2.6	3	0.8114	2.401	2.500	6	0.7536	4.549	4.926	10	0.7526	8.288	8.820
	2.4	3	0.7914	2.415	2.500	7	0.7502	5.559	5.786	12	0.7518	10.102	10.688
	2.2	3	0.7681	2.430	2.500	9	0.7545	7.286	7.675	15	0.7511	12.871	13.566
	2.0	5	0.7901	3.963	4.125	12	0.7577	9.902	10.431	20	0.7533	17.481	18.319
	1.8	5	0.7536	4.005	4.125	17	0.7580	14.412	15.079	29	0.7572	25.909	26.990
	1.6	9	0.7647	7.295	7.539	26	0.7517	22.732	23.596	46	0.7544	42.072	43.469
	1.4	17	0.7588	14.253	14.662	52	0.7529	47.206	48.550	91	0.7503	85.460	87.455
	1.2	55	0.7513	49.135	50.056	180	0.7510	170.794	173.516	323	0.7502	312.486	316.358

A.2 Updated Tables for Procedure $\mathcal{M}_{C'}$

Table A.2 identifies the $(n_{C'}, t)$-pairs that minimize $\mathsf{E}_{SC}[N]$ while still achieving $P^\star$. We do not include a table for $k = 2$, since, as shown in Chapter 6, Procedures $\mathcal{M}_{C'}$ and $\mathcal{M}_{BK}$ are identical in that case; and so we can consult the Procedure $\mathcal{M}_{BK}$ table in Appendix A.1.

We searched all possible $(n_{C'}, t)$-pairs up to $n_{C'} = 125$. Rows with no entries in the table are $\theta^\star$-values for which $P^\star$ cannot be achieved within the search space. These tables improve upon those in Chen (1988a), in which his values for $\mathsf{E}_{SC}[N]$ did not incorporate curtailment. Also, his tables only provided performance characteristics for $n_C \leq 30$. We also provide the expected number of observations in the EPC (i.e., $\mathsf{E}_{EPC}[N]$).

Table A.2: Updated Performance Characteristics for Procedure $\mathcal{M}_{C'}$

$P^\star$	$\theta^\star$	$k=3$					$k=4$				
		$n_{C'}$	t	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	$n_{C'}$	t	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$
0.95	3.0	17	7	0.9509	11.470	14.356	25	8	0.9508	15.773	20.953
	2.8	19	8	0.9508	13.451	16.486	28	9	0.9500	18.384	24.040
	2.6	23	9	0.9513	15.707	19.523	31	11	0.9505	23.151	28.453
	2.4	27	11	0.9527	19.840	23.981	43	12	0.9501	26.711	34.594
	2.2	33	13	0.9505	24.478	29.390	48	15	0.9501	35.040	43.523
	2.0	43	17	0.9516	33.488	39.228	64	19	0.9503	47.068	57.894
	1.8	60	23	0.9508	47.966	55.370	91	26	0.9504	68.832	82.933
	1.6	95	35	0.9503	78.097	88.608					
	1.4										
	1.2										
0.9	3.0	12	5	0.9066	8.026	9.669	17	6	0.9026	11.553	14.350
	2.8	13	6	0.9062	9.727	11.196	19	7	0.9032	13.891	16.779
	2.6	15	7	0.9049	11.583	13.162	22	8	0.9015	16.498	19.668
	2.4	18	8	0.9045	13.916	15.876	27	9	0.9004	19.607	23.688
	2.2	23	9	0.9025	16.691	19.523	34	11	0.9018	25.269	30.211
	2.0	29	12	0.9028	23.088	26.073	44	14	0.9006	34.069	39.933
	1.8	40	16	0.9001	32.667	36.449	63	19	0.9011	49.624	57.514
	1.6	64	24	0.9001	52.906	58.699	100	29	0.9003	82.010	92.877
	1.4	125	48	0.9001	112.579	119.348					
	1.2										
0.75	3.0	5	3	0.7690	3.950	4.111	8	3	0.7508	5.154	5.826
	2.8	6	3	0.7803	4.493	4.926	9	4	0.7627	7.051	7.733
	2.6	6	3	0.7536	4.549	4.926	11	4	0.7585	7.611	8.620
	2.4	7	4	0.7502	5.559	5.786	13	5	0.7627	9.952	11.065
	2.2	10	4	0.7581	6.744	7.348	16	6	0.7601	12.600	13.886
	2.0	12	5	0.7505	8.927	9.669	21	7	0.7527	15.900	17.641
	1.8	17	7	0.7533	13.374	14.356	29	10	0.7539	24.179	26.166
	1.6	26	11	0.7507	22.193	23.325	49	14	0.7511	37.863	41.335
	1.4	52	20	0.7509	45.454	47.565	94	27	0.7510	81.002	86.274
	1.2										

A.3 Updated Tables for Procedure $\mathcal{M}_{\mathrm{RA}}$

Table A.3 identifies the (r,t)-pairs that minimize $\mathsf{E}_{\mathrm{SC}}[N]$ while still achieving $P^\star$. We searched all possible (r,t)-pairs up to $t = 150$ for $k = 2$ and 3, and up to $t = 75$ for $k = 4$. Rows with no entries in the table are $\theta^\star$-values for which $P^\star$ cannot be achieved within the search space.

These tables improve upon those in BG (1985a) by including a greater range of $\theta^\star$-values (theirs included $\theta^\star = 2.0, 2.4, 3.0$ with some entries for $\theta^\star = 1.6$), as well as a few corrections to their original paper. In the table, the symbol † represents an entry in our table that is different from that in BG. For that particular instance, BG allow $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ to be slightly below $P^\star$; in our table, we do not. The symbol ‡ represents a value that is different than that in BG due to either our improved algorithm or our ability to calculate an exact result when BG estimated the result using MC sampling.

Table A.3: Updated Performance Characteristics for Procedure $\mathcal{M}_{\mathrm{RA}}$

$P^\star$	$\theta^\star$	$k=2$					$k=3$					$k=4$				
		r	t	P(CS)	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	r	t	P(CS)	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	r	t	P(CS)	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$
0.95	3.0	3	6	0.9522	5.251	6.943	4	7	0.9505	9.768	13.731	4	9	0.9577	13.747	22.193
	2.8	3	8	0.9515	5.650	7.843	4	9	0.9583	11.211	17.101	4	10	0.9527	15.006	24.349
	2.6	4	7	0.9536	7.541	9.658	4	10	0.9513	12.169	18.528	4	13	0.9533	17.009	29.714
	2.4	4	9	0.9548	8.465	11.380	4	14	0.9509	13.837	22.819	5	13	0.9531	22.297	34.652
	2.2	4	12	0.9505	9.426	13.127	5	14	0.9518	18.626	27.937	5	17	0.9518	26.322	43.276
	2.0	5	14	0.9537	13.091	17.898	5	22	0.9505	22.401	37.027	6	21	0.9534	36.255	57.562
	1.8	6	18	0.9508	18.034	24.301	6	28	0.9500	31.951	51.176	6	40	0.9502	46.412	86.080
	1.6	7	30	0.9502	26.559	37.094	7	80	0.9500	48.638	91.045	8	48	0.9503	78.184	128.640
	1.4	9	76	0.9501	48.312	72.356	10	99	0.9501	98.741	166.036					
	1.2															
0.9	3.0	$3^\dagger$	$4^\dagger$	$0.9261^\dagger$	$4.560^\dagger$	$5.344^\dagger$	3	5	0.9004	6.762	8.740	3	7	0.9091	9.869	14.709
	2.8	3	4	0.9132	4.633	5.344	3	6	0.9020	7.580	10.131	3	8	0.9008	10.750	16.168
	2.6	3	5	0.9168	5.234	6.258	3	8	0.9039	8.610	12.168	3	14	0.9002	12.220	20.826
	2.4	3	6	0.9113	5.718	6.943	4	8	0.9104	11.785	15.506	4	10	0.9107	16.980	24.349
	2.2	3	8	0.9037	6.326	7.843	4	10	0.9060	13.720	18.528	4	13	0.9055	19.976	29.714
	2.0	4	8	0.9033	8.899	$10.587^\ddagger$	4	15	0.9015	16.514	23.604	5	15	0.9060	27.928	39.218
	1.8	4	14	0.9021	11.040	13.907	5	18	0.9015	24.185	33.164	5	25	0.9017	36.201	55.197
	1.6	5	21	0.9006	17.001	21.482	6	30	0.9021	37.825	53.074	$6^\ddagger$	$45^\ddagger$	$0.9005^\ddagger$	$57.123^\ddagger$	$90.224^\ddagger$
	1.4	7	40	0.9000	32.924	41.835	8	64	0.9005	73.329	105.144	9	66	0.9001	121.152	173.668
	1.2	13	134	0.9003	112.282	143.502										
0.75	3.0	1	1	0.7500	1.000	1.000	2	3	0.7961	3.684	4.247	2	4	0.7718	5.148	6.386
	2.8	2	2	0.8287	2.388	2.500	2	3	0.7738	3.745	4.247	2	5	0.7639	5.668	7.187
	2.6	2	2	0.8114	2.401	2.500	2	4	0.7763	4.328	5.068	2	9	0.7506	6.226	8.243
	2.4	2	2	0.7914	2.415	2.500	2	5	0.7601	4.703	5.543	3	5	0.7731	9.221	10.900
	2.2	2	2	0.7681	2.430	2.500	3	4	0.7550	6.388	7.049	3	6	0.7629	10.895	12.963
	2.0	2	3	0.7737	3.086	3.250	4	5	0.7556	8.809	9.629	3	8	0.7525	13.371	16.168
	1.8	2	4	0.7524	3.437	3.625	3	9	0.7583	11.024	12.895	4	10	0.7578	20.687	24.349
	1.6	3	5	0.7559	5.956	6.258	4	12	0.7572	18.240	20.929	4	20	0.7514	30.051	37.290
	1.4	4	10	0.7555	11.354	12.056	5	26	0.7515	34.180	39.854	6	32	0.7535	64.368	76.981
	1.2	7	34	0.7514	36.748	39.283	9	90	0.7505	119.612	139.373					

A.4 Updated Tables for Procedure $\mathcal{M}_{\mathrm{RA}'}$

Table A.4 identifies the $(n_{\mathrm{RA}'}, r, t)$-triplets that minimize $\mathsf{E}_{\mathrm{SC}}[N]$ while still achieving $P^\star$. We do not include a table for $k = 2$, since, as shown in Chapter 6, Procedures $\mathcal{M}_{\mathrm{RA}'}$ and $\mathcal{M}_{\mathrm{RA}}$ are identical in that case; and so we can consult the Procedure $\mathcal{M}_{\mathrm{RA}}$ table in Appendix A.3 when $k = 2$.

We searched all possible $(n_{\mathrm{RA}'}, r, t)$-triplets up to $n_{\mathrm{RA}'} = 125$ for $k = 3$ and 4. Rows with no entries in the table are $\theta^\star$-values for which $P^\star$ cannot be achieved within the search space. These tables improve upon those in Chen (1992) by including a greater range of $\theta^\star$-values (his included $\theta^\star = 2.0, 2.4, 3.0$), as well as corrections to some numerical errors found in his original paper. We use a † to identify entries that are corrections to values found in Table 1 of Chen (1992).

Table A.4: Updated Performance Characteristics for Procedure $\mathcal{M}_{\mathrm{RA}'}$

$P^\star$	$\theta^\star$	$k=3$						$k=4$					
		$n_{\mathrm{RA}'}$	r	t	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	$n_{\mathrm{RA}'}$	r	t	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$
0.95	3.0	19	4	7	0.9505	9.768	13.731	25	4	9	0.9519	13.602	20.458
	2.8	20	4	9	0.9507	10.973	15.594	31	4	10	0.9510	14.976	23.712
	2.6	26	4	10	0.9503	12.157	18.351	39	4	13	0.9508	16.931	28.164
	2.4	36	4	14	0.9500	13.812	22.377	$41^\dagger$	5	13	0.9506	22.229	33.560
	2.2	37	5	14	0.9506	18.598	27.598	56	5	17	0.9505	26.279	42.318
	2.0	60	5	22	0.9501	22.389	36.742	68	6	21	$0.9505^\dagger$	$36.094^\dagger$	$55.301^\dagger$
	1.8	82	6	28	0.9500	31.951	51.176	101	7	28	0.9500	51.051	79.023
	1.6	111	8	39	0.9501	53.256	80.785						
	1.4												
	1.2												
0.9	3.0	13	3	5	0.9004	6.762	8.740	19	3	7	0.9016	9.739	13.762
	2.8	15	3	6	0.9004	7.574	10.089	26	3	8	0.9002	10.745	16.076
	2.6	19	3	8	0.9004	8.537	11.737	46	3	14	0.9000	12.217	20.626
	2.4	19	4	8	0.9038	11.637	14.857	29	4	10	0.9025	16.750	22.904
	2.2	24	4	11	0.9007	13.615	17.782	40	4	13	0.9012	19.832	28.528
	2.0	40	4	15	$0.9006^\dagger$	$16.489^\dagger$	23.396	47	5	15	0.9000	27.674	37.600
	1.8	49	5	18	0.9004	24.151	32.941	85	5	25	0.9003	36.125	54.247
	1.6	81	6	30	0.9003	37.691	52.270	111	7	32	0.9004	62.545	86.294
	1.4	125	14	49	0.9000	100.609	116.221						
	1.2												
0.75	3.0	5	2	3	0.7574	3.242	3.481	9	2	4	0.7513	4.895	5.740
	2.8	6	2	3	0.7623	3.699	4.148	12	2	5	0.7521	5.491	6.659
	2.6	7	2	4	0.7533	3.937	4.379	26	2	10	0.7500	6.217	8.154
	2.4	10	2	5	0.7507	4.534	5.187	13	3	5	0.7555	8.964	10.271
	2.2	10	3	4	0.7550	6.388	7.049	17	3	6	0.7529	10.733	12.534
	2.0	$13^\dagger$	$3^\dagger$	$6^\dagger$	$0.7572^\dagger$	$8.395^\dagger$	$9.360^\dagger$	26	3	8	0.7506	13.345	16.076
	1.8	21	3	9	0.7503	10.766	12.375	32	4	10	0.7528	20.551	23.974
	1.6	30	4	12	0.7505	17.927	20.327	70	4	20	0.7504	30.004	37.097
	1.4	72	5	26	0.7503	34.099	39.659	112	6	32	0.7503	64.011	75.963
	1.2												

APPENDIX B

PROCEDURE COMPARISON TABLES

This appendix includes tables for all possible combinations of $k \in \{2, 3, 4\}$, $P^\star \in \{0.75, 0.9, 0.95\}$, and $\theta^\star \in \{1.6, 2.0, 2.4, 3.0\}$. As discussed in Chapter 6, we require that all procedures be able to operate under a firm budget constraint, b, on the maximum number of observations, which sometimes results in a procedure not being able to achieve $P^\star$.

B.1 Explanation of Results

In all of the tables included here, the LP, or randomized, formulations solve to an optimal value. On the other hand, the MIP, or non-randomized, formulations, often solve to an optimal *integer tolerance* value, particularly for larger problems. That means that binary variables in the solution vector are not necessarily in the set $\{0, 1\}$ as we specified, but are within a preset tolerance of those values. The $\mathsf{E}_{\mathrm{SC}}[N]$ results for MIP integer tolerance solutions are marked with a † in the tables. In some cases for our integer tolerance solutions, $\mathsf{E}_{\mathrm{SC}}[N]$ for Procedure $\mathcal{M}_{\mathrm{NR}}$ is slightly lower than that of Procedure $\mathcal{M}_{\mathrm{R}}$, which should not be possible. This phenomenon occurs because the actual $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ achieved by the former is slightly lower than $P^\star$. We allow these results when $\mathsf{P}'_{\mathrm{SC}}(\mathrm{CS}) \geq P^\star$, where $\mathsf{P}'_{\mathrm{SC}}(\mathrm{CS})$ is $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ rounded to four decimal places. Such occurrences in the table are marked with a ‡. If $\mathsf{P}'_{\mathrm{SC}}(\mathrm{CS}) < P^\star$, we either tune the solver parameters until the condition is met or we do not report the results.

Upon conclusion of an optimization, we transform the solution vector into a set of stopping vectors and stopping probabilities, and input those into a numerical algorithm to recalculate the expected number of observations and achieved $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$. For all of the randomized results reported here, the recalculated metrics were equal to the optimization output (to four decimal places). For the non-randomized results, the integer tolerance issues led to some deviance between the expected number of observations returned by the

optimization and that returned by the algorithm, but the deviance never exceeded 10^{-4}. In all cases, the numerical results reported in the tables are from the algorithm, not the optimization. This removes the effects of integer tolerance on the expected number of observations and $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ and returns the actual values of those metrics for the particular set of stopping points (i.e., procedure) returned by the optimization.

We could have tuned the parameters of the solver in order to minimize or eliminate the deviations due to tolerance, and we did so in specific cases. However, the point here is to introduce the optimal procedures and show the potential savings in the expected number of observations, not necessarily to find the absolutely optimal solution for the non-randomized procedures. Anyone requiring more accuracy for a particular case could tune the parameters as necessary to meet their requirements. Note that the results for Procedure $\mathcal{M}_{\mathrm{R}}$ *are* optimal (under our problem assumptions) and therefore serve as a lower bound on the minimum expected number of observations for the remaining non-randomized procedures.

B.2 Result Verification

We generated all table entries, not just those of the optimal procedures, using our own algorithms. Thus, none of the entries are taken from published data. To verify our results, our first check was to compare our output with previously published data, when available. In all such cases, our values for $\mathsf{E}_{\mathrm{SC}}[N]$ and $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ are identical to published data when those data were calculated via numerical algorithms. When published data were estimated via Monte Carlo (MC) sampling, our results did differ slightly but within a few standard errors, as expected.

Our second check was to conduct MC sampling of *all* table entries. For each entry, we conducted 100,000 independent replications of the procedure. We then determined the distance, in number of standard errors, between our table entries and the sampling results. Of the 268 MC results, all but nine were within two standard errors of the tabulated expected number of observations and all but ten were within two standard errors of the tabulated $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$. None exceeded two standard errors for *both* $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ *and* the expected

number of observations, and in no cases were the results three or more standard errors from the tabulated entries.

For any MC result outside of two standard errors of the tabulated data, we first determined if the tabulated data could be verified via a published source. If so, we did not pursue those any further. If not, we took 100,000 more MC samples. In all ten of those cases, the MC results were within two standard errors of our tabulated data.

Thus, we have reasonable confidence that our results are accurate.

B.3 Tabulated Results

For all existing procedures, we report the parameters of the procedure that minimize $\mathsf{E}_{\mathrm{SC}}[N]$, while achieving the required $P^\star$ and remaining under the observation budget, b. Those parameters provide all that is necessary to implement the procedure. For Procedures $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{NR}}$, we only report the number of left-lexicographic stopping points, not the stopping points and stopping probabilities themselves. In addition to $\mathsf{E}_{\mathrm{SC}}[N]$ and $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$, we also report $\mathsf{E}_{\mathrm{EPC}}[N]$, $\mathsf{SD}_{\mathrm{SC}}(N)$, and W_{J} (as a percentage and labeled "% Incr"), i.e., the percent increase in the expected number of observations of general procedure $\mathcal{M}_{\mathrm{J}}$ over Procedure $\mathcal{M}_{\mathrm{R}}$.

Blank rows for a particular procedure in a table indicate one of two situations. First, the procedure may not be able to achieve the given $P^\star$ under the budget constraint. These are marked by an "N/A" in the Parameters column. Second, the computational time or requirements for calculating $\mathsf{E}_{\mathrm{SC}}[N]$ and $\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$ for a particular procedure may be beyond our current capabilities. These are marked by "??" in the Parameters column. The tables follow.

Table B.1: Comparative Results for $k = 2$ and $\theta^\star = 1.6$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 31$	0.9500	26.355	14.113	36.396	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 30$	0.9500	$26.356^{\ddagger}$	14.125	36.401	0.00
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 59$	0.9502	26.559	14.825	37.094	0.78
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 59, r = 7, t = 30$	0.9502	26.559	14.825	37.094	0.78
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 7, t = 30$	0.9502	26.559	14.825	37.094	0.78
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 49, t = 25$	0.9501	40.331	4.497	44.386	53.03
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 49$	0.9501	40.331	4.497	44.386	53.03
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 49$	0.9501	49.000	0.000	49.000	85.92
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 22$	0.9000	16.865	10.019	21.185	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 21$	0.9000	16.873	10.068	21.208	0.05
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 41$	0.9006	17.001	10.318	21.482	0.81
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 41, r = 5, t = 21$	0.9006	17.001	10.318	21.482	0.81
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 5, t = 21$	0.9006	17.001	10.318	21.482	0.81
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 31, t = 16$	0.9054	25.505	3.259	27.522	51.23
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 31$	0.9054	25.505	3.259	27.522	51.23
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 31$	0.9054	31.000	0.000	31.000	83.81
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 6$	0.7500	5.558	2.056	5.814	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 5$	0.7559	5.956	2.289	6.258	7.16
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 9$	0.7559	5.956	2.289	6.258	7.16
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 9, r = 3, t = 5$	0.7559	5.956	2.289	6.258	7.16
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 3, t = 5$	0.7559	5.956	2.289	6.258	7.16
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 9, t = 5$	0.7647	7.295	1.272	7.539	31.25
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 9$	0.7647	7.295	1.272	7.539	31.25
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 9$	0.7647	9.000	0.000	9.000	61.93

Table B.2: Comparative Results for $k = 2$ and $\theta^\star = 2$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 15$	0.9500	12.411	5.992	16.625	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 14$	0.9501	12.435	5.977	16.618	0.20
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 27$	0.9537	13.091	6.671	17.898	5.48
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 27, r = 5, t = 14$	0.9537	13.091	6.671	17.898	5.48
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 5, t = 14$	0.9537	13.091	6.671	17.898	5.48
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 23, t = 12$	0.9520	17.806	2.629	20.132	43.47
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 23$	0.9520	17.806	2.629	20.132	43.47
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 23$	0.9520	23.000	0.000	23.000	85.32
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 9$	0.9000	8.511	3.519	10.048	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 8$	0.9001	8.686	3.337	10.146	2.06
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 15$	0.9033	8.899	3.796	10.587	4.56
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 15, r = 4, t = 8$	0.9033	8.899	3.796	10.587	4.56
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 4, t = 8$	0.9033	8.899	3.796	10.587	4.56
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 15, t = 8$	0.9118	11.681	1.926	12.858	37.25
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 15$	0.9118	11.681	1.926	12.858	37.25
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 15$	0.9118	15.000	0.000	15.000	76.24
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 4$	0.7500	2.625	1.409	2.752	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 3$	0.7737	3.086	1.259	3.250	17.58
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 5$	0.7737	3.086	1.259	3.250	17.58
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 5, r = 2, t = 3$	0.7737	3.086	1.259	3.250	17.58
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 2, t = 3$	0.7737	3.086	1.259	3.250	17.58
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 5, t = 3$	0.7901	3.963	0.793	4.125	50.97
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 5$	0.7901	3.963	0.793	4.125	50.97
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 5$	0.7901	5.000	0.000	5.000	90.48

Table B.3: Comparative Results for $k = 2$ and $\theta^\star = 2.4$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 10$	0.9500	7.962	3.581	10.430	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 9$	0.9503	8.004	3.674	10.531	0.53
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 17$	0.9548	8.465	4.085	11.380	6.32
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 17, r = 4, t = 9$	0.9548	8.465	4.085	11.380	6.32
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 4, t = 9$	0.9548	8.465	4.085	11.380	6.32
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 15, t = 8$	0.9552	11.188	1.884	12.858	40.51
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 15$	0.9552	11.188	1.884	12.858	40.51
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 15$	0.9552	15.000	0.000	15.000	88.39
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 7$	0.9000	5.226	2.238	6.144	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 6$	0.9001	5.259	2.368	6.223	0.62
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 11$	0.9113	5.718	2.736	6.943	9.40
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 11, r = 3, t = 6$	0.9113	5.718	2.736	6.943	9.40
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 3, t = 6$	0.9113	5.718	2.736	6.943	9.40
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 9, t = 5$	0.9082	6.823	1.283	7.539	30.55
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 9$	0.9082	6.823	1.283	7.539	30.55
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 9$	0.9082	9.000	0.000	9.000	72.21
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 3$	0.7500	1.730	0.791	1.774	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 2$	0.7914	2.415	0.493	2.500	39.58
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 3$	0.7914	2.415	0.493	2.500	39.58
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 3, r = 2, t = 2$	0.7914	2.415	0.493	2.500	39.58
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 2, t = 2$	0.7914	2.415	0.493	2.500	39.58
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 3, t = 2$	0.7914	2.415	0.493	2.500	39.58
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 3$	0.7914	2.415	0.493	2.500	39.58
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 3$	0.7914	3.000	0.000	3.000	73.37

Table B.4: Comparative Results for $k = 2$ and $\theta^\star = 3$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 7$	0.9500	5.140	2.363	6.682	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 6$	0.9522	5.251	2.530	6.943	2.17
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 11$	0.9522	5.251	2.530	6.943	2.17
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 11, r = 3, t = 6$	0.9522	5.251	2.530	6.943	2.17
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 3, t = 6$	0.9522	5.251	2.530	6.943	2.17
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 9, t = 5$	0.9511	6.540	1.238	7.539	27.24
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 9$	0.9511	6.540	1.238	7.539	27.24
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 9$	0.9511	9.000	0.000	9.000	75.10
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 5$	0.9000	3.405	1.559	3.973	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 4$	0.9163	4.063	1.293	4.688	19.35
	Procedure $\mathcal{M}_{\mathrm{BG}}$	N/A					
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 7, r = 3, t = 4$	0.9261	4.560	1.493	5.344	33.94
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 3, t = 4$	0.9261	4.560	1.493	5.344	33.94
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 7, t = 4$	0.9294	5.163	1.020	5.813	51.65
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 7$	0.9294	5.163	1.020	5.813	51.65
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 7$	0.9294	7.000	0.000	7.000	105.61
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 1$	0.7500	1.000	0.000	1.000	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 1$	0.7500	1.000	0.000	1.000	0.00
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 1$	0.7500	1.000	0.000	1.000	0.00
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 1, r = 1, t = 1$	0.7500	1.000	0.000	1.000	0.00
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 1, t = 1$	0.7500	1.000	0.000	1.000	0.00
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 1, t = 1$	0.7500	1.000	0.000	1.000	0.00
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 1$	0.7500	1.000	0.000	1.000	0.00
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 1$	0.7500	1.000	0.000	1.000	0.00

Table B.5: Comparative Results for $k = 3$ and $\theta^\star = 1.6$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops = 2049	0.9500	50.049	27.313	79.216	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	??					
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 125$	0.9502	50.321	28.698	81.434	0.54
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 111, r = 8, t = 39$	0.9501	53.256	27.071	80.785	6.41
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 8, t = 39$	0.9504	53.270	27.102	81.005	6.44
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 95, t = 35$	0.9503	78.097	8.770	88.608	56.04
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 93$	0.9502	82.011	5.697	88.365	63.86
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 93$	0.9502	93.000	0.000	93.000	85.82
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops = 806	0.9000	36.949	19.345	50.761	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	??					
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 83$	0.9003	37.261	20.583	52.614	0.84
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 81, r = 6, t = 30$	0.9003	37.691	20.981	52.270	2.01
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 7, t = 25$	0.9010	41.243	17.791	53.841	11.62
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 64, t = 24$	0.9001	52.906	6.576	58.699	43.19
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 63$	0.9007	55.643	4.358	59.203	50.60
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 63$	0.9007	63.000	0.000	63.000	70.51
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops = 102	0.7500	17.242	8.101	19.548	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops = 101	0.7500	$17.243^\dagger$	8.111	19.548	0.01
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 32$	0.7517	17.597	8.823	20.254	2.06
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 30, r = 4, t = 12$	0.7505	17.927	8.409	20.327	3.97
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 5, t = 11$	0.7628	20.515	6.778	22.903	18.98
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 26, t = 11$	0.7507	22.193	2.844	23.325	28.72
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 26$	0.7517	22.732	2.275	23.596	31.84
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 26$	0.7517	26.000	0.000	26.000	50.80

Table B.6: Comparative Results for $k = 3$ and $\theta^\star = 2$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 252$	0.9500	22.750	11.975	35.038	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 234$	0.9500	$22.749^{\ddagger}$	11.982	35.035	
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 52$	0.9508	23.032	12.475	35.972	1.24
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 48, r = 6, t = 17$	0.9502	25.113	11.011	36.327	10.39
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 6, t = 17$	0.9504	25.114	11.013	36.343	10.39
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 43, t = 17$	0.9516	33.488	4.933	39.228	47.20
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 42$	0.9509	35.023	3.511	38.921	53.95
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 42$	0.9509	42.000	0.000	42.000	84.62
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 114$	0.9000	16.857	8.429	22.676	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 112$	0.9000	$16.859^{\dagger}$	8.440	22.662	0.01
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 34$	0.9016	17.165	8.813	23.296	1.83
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 30, r = 5, t = 12$	0.9001	18.749	7.470	23.902	11.22
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 5, t = 12$	0.9057	18.940	7.746	24.698	12.35
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 29, t = 12$	0.9028	23.088	3.563	26.073	36.96
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 29$	0.9044	24.242	2.716	26.455	43.80
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 29$	0.9044	29.000	0.000	29.000	72.03
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 22$	0.7500	7.831	3.200	8.765	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 19$	0.7500	7.837	3.197	8.775	0.08
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 13$	0.7512	7.966	3.315	8.934	1.72
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 13, r = 3, t = 6$	0.7572	8.395	3.399	9.360	7.21
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 4, t = 5$	0.7556	8.809	2.208	9.629	12.49
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 12, t = 5$	0.7505	8.927	1.912	9.669	13.99
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 12$	0.7577	9.902	1.453	10.431	26.45
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 12$	0.7577	12.000	0.000	12.000	53.24

Table B.7: Comparative Results for $k = 3$ and $\theta^\star = 2.4$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops = 97	0.9500	14.177	7.046	21.303	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops = 96	0.9500	14.181†	7.000	21.335	0.03
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 31$	0.9516	14.479	7.397	22.086	2.13
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 28, r = 5, t = 11$	0.9509	15.921	6.503	22.426	12.30
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 5, t = 11$	0.9530	15.969	6.590	22.903	12.64
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 27, t = 11$	0.9527	19.840	3.496	23.981	39.94
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 26$	0.9511	20.781	2.565	23.596	46.58
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 26$	0.9511	26.000	0.000	26.000	83.40
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops = 52	0.9000	10.235	5.160	13.784	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops = 50	0.9000	10.236†	5.163	13.792	0.02
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 22$	0.9021	10.429	5.397	14.247	1.90
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 19, r = 4, t = 8$	0.9038	11.637	4.638	14.857	13.71
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 4, t = 8$	0.9104	11.785	4.862	15.506	15.15
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 18, t = 8$	0.9045	13.916	2.405	15.876	35.97
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 18$	0.9056	14.436	1.994	16.035	41.05
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 18$	0.9056	18.000	0.000	18.000	75.87
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops = 10	0.7500	4.910	1.864	5.390	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops = 9	0.7569	5.172	1.673	5.658	5.32
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 8$	0.7602	5.403	1.770	5.938	10.02
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 7, r = 4, t = 4$	0.7502	5.559	0.857	5.786	13.22
	Procedure $\mathcal{M}_{\mathrm{RA}}$	N/A					
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 7, t = 4$	0.7502	5.559	0.857	5.786	13.22
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 7$	0.7502	5.559	0.857	5.786	13.22
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 7$	0.7502	7.000	0.000	7.000	42.56

Table B.8: Comparative Results for $k = 3$ and $\theta^\star = 3$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 44$	0.9500	8.844	4.399	13.383	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 44$	0.9500	$8.852^\dagger$	4.387	13.430	0.09
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 20$	0.9505	8.901	4.474	13.573	0.64
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 19, r = 4, t = 7$	0.9505	9.768	3.830	13.731	10.44
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 4, t = 7$	0.9505	9.768	3.830	13.731	10.44
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 17, t = 7$	0.9509	11.470	2.414	14.356	29.69
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 17$	0.9554	12.958	1.895	15.079	46.51
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 17$	0.9554	17.000	0.000	17.000	92.21
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 19$	0.9000	6.762	2.864	8.682	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 18$	0.9011	6.824	2.897	8.781	0.92
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 12$	0.9029	6.969	3.023	8.933	3.06
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 12, r = 3, t = 6$	0.9029	6.969	3.023	8.933	3.06
	Procedure $\mathcal{M}_{\mathrm{RA}}$	N/A					
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 12, t = 5$	0.9066	8.026	1.864	9.669	18.70
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 11$	0.9014	8.460	1.353	9.482	25.11
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 11$	0.9014	11.000	0.000	11.000	62.67
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 6$	0.7500	3.068	1.053	3.290	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 5$	0.7574	3.242	1.143	3.481	5.66
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 5$	0.7574	3.242	1.143	3.481	5.66
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 5, r = 2, t = 3$	0.7574	3.242	1.143	3.481	5.66
	Procedure $\mathcal{M}_{\mathrm{RA}}$	N/A					
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 5, t = 3$	0.7690	3.950	0.642	4.111	28.76
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 5$	0.7690	3.950	0.642	4.111	28.76
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 5$	0.7690	5.000	0.000	5.000	62.97

Table B.9: Comparative Results for $k = 4$ and $\theta^\star = 1.6$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	??					
	Procedure $\mathcal{M}_{\mathrm{NR}}$	??					
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 181$	0.9500	76.265	41.254	128.931	
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	??					
	Procedure $\mathcal{M}_{\mathrm{RA}}$	??					
	Procedure $\mathcal{M}_{\mathrm{C}'}$	??					
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 138$	0.9506	125.956	6.368	133.642	
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 138$	0.9506	138.000	0.000	138.000	
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops = 22031	0.9000	58.189	29.900	83.609	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	??					
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 126$	0.9004	58.715	31.160	86.889	0.90
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 111, r = 7, t = 32$	0.9004	62.545	29.113	86.294	7.49
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 7, t = 32$	0.9029	62.755	29.470	87.896	7.85
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 100, t = 29$	0.9003	82.010	10.439	92.877	40.94
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 97$	0.9005	88.796	4.993	93.341	52.60
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 97$	0.9005	97.000	0.000	97.000	66.70
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops = 2186	0.7500	30.549	14.559	36.195	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	??					
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 57$	0.7512	31.109	15.462	37.649	1.83
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 50, r = 5, t = 15$	0.7504	33.385	12.826	38.648	9.29
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 5, t = 15$	0.7551	33.601	13.115	39.218	9.99
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 49, t = 14$	0.7511	37.863	6.092	41.335	23.94
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 46$	0.7544	42.072	2.862	43.469	37.72
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 46$	0.7544	46.000	0.000	46.000	50.58

Table B.10: Comparative Results for $k = 4$ and $\theta^\star = 2$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops = 4914	0.9500	33.673	16.774	53.912	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	??					
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 74$	0.9500	33.824	17.611	55.326	0.45
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 68, r = 6, t = 21$	0.9505	36.094	16.872	55.301	7.19
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 8, t = 19$	0.9515	41.970	13.131	57.765	24.64
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 64, t = 19$	0.9503	47.068	7.664	57.894	39.78
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 61$	0.9513	53.207	4.034	58.092	58.01
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 61$	0.9513	61.000	0.000	61.000	81.16
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops = 1616	0.9000	25.566	12.699	36.378	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops = 1384	0.9000	$25.566^{\ddagger}$	12.700	36.396	
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 53$	0.9000	25.706	13.518	37.305	0.55
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 47, r = 5, t = 15$	0.9000	27.674	12.181	37.600	8.24
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 6, t = 14$	0.9049	30.293	10.316	39.759	18.49
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 44, t = 14$	0.9006	34.069	5.802	39.933	33.26
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 43$	0.9022	37.669	3.172	40.557	47.34
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 43$	0.9022	43.000	0.000	43.000	68.19
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops = 169	0.7500	13.342	6.098	15.712	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops = 163	0.7500	$13.342^{\ddagger}$	6.098	15.725	0.00
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 24$	0.7541	13.781	6.448	16.449	3.29
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 22, r = 4, t = 7$	0.7534	14.973	4.636	17.137	12.22
	Procedure $\mathcal{M}_{\mathrm{RA}}$	N/A					
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 21, t = 7$	0.7527	15.900	3.254	17.641	19.17
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 20$	0.7533	17.481	1.799	18.319	31.02
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 20$	0.7533	20.000	0.000	20.000	49.90

Table B.11: Comparative Results for $k = 4$ and $\theta^\star = 2.4$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 855$	0.9500	20.471	10.064	32.570	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 803$	0.9500	$20.474^{\dagger}$	10.165	32.662	0.01
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 44$	0.9506	20.679	10.484	33.382	1.01
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 41, r = 5, t = 13$	0.9506	22.229	9.744	33.560	8.58
	Procedure $\mathcal{M}_{\mathrm{RA}}$	N/A					
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 43, t = 12$	0.9501	26.711	5.313	34.594	30.48
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 37$	0.9512	31.115	2.998	34.731	51.99
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 37$	0.9512	37.000	0.000	37.000	80.74
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 322$	0.9000	15.604	7.378	21.869	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 310$	0.9000	$15.601^{\ddagger}$	7.375	21.874	
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 31$	0.9022	15.927	7.794	22.767	2.07
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 29, r = 4, t = 10$	0.9025	16.750	7.491	22.904	7.35
	Procedure $\mathcal{M}_{\mathrm{RA}}$	N/A					
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 27, t = 9$	0.9004	19.607	4.019	23.688	25.65
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 26$	0.9017	21.980	2.356	24.093	40.86
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 26$	0.9017	26.000	0.000	26.000	66.62
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 55$	0.7500	7.922	3.731	9.377	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 50$	0.7500	$7.927^{\ddagger}$	3.738	9.391	0.06
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 15$	0.7569	8.286	4.020	9.911	4.59
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 13, r = 3, t = 5$	0.7555	8.964	3.152	10.271	13.15
	Procedure $\mathcal{M}_{\mathrm{RA}}$	N/A					
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 13, t = 5$	0.7627	9.952	2.141	11.065	25.61
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 12$	0.7518	10.102	1.339	10.688	27.52
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 12$	0.7518	12.000	0.000	12.000	51.47

Table B.12: Comparative Results for $k = 4$ and $\theta^\star = 3$

$P^\star$	Procedure	Parameters	$\mathsf{P}_{\mathrm{SC}}(\mathrm{CS})$	$\mathsf{E}_{\mathrm{SC}}[N]$	$\mathsf{SD}_{\mathrm{SC}}[N]$	$\mathsf{E}_{\mathrm{EPC}}[N]$	% Incr
0.95	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 206$	0.9500	12.741	5.858	19.727	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 205$	0.9500	$12.743^\dagger$	5.793	19.688	0.02
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 26$	0.9513	12.968	6.114	20.341	1.78
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 25, r = 4, t = 9$	0.9519	13.602	6.070	20.458	6.76
	Procedure $\mathcal{M}_{\mathrm{RA}}$	N/A					
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 25, t = 8$	0.9508	15.773	3.590	20.953	23.80
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 23$	0.9527	18.523	2.222	21.215	45.38
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 23$	0.9527	23.000	0.000	23.000	80.53
0.9	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 95$	0.9000	9.534	4.690	13.361	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 93$	0.9000	$9.534^\dagger$	4.686	13.355	0.00
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 19$	0.9036	9.844	5.048	13.852	3.25
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 19, r = 3, t = 7$	0.9016	9.739	4.912	13.762	2.15
	Procedure $\mathcal{M}_{\mathrm{RA}}$	N/A					
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 17, t = 6$	0.9026	11.553	2.748	14.350	21.17
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 16$	0.9024	12.969	1.741	14.493	36.02
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 16$	0.9024	16.000	0.000	16.000	67.82
0.75	Procedure $\mathcal{M}_{\mathrm{R}}$	Stops $= 19$	0.7500	4.848	2.370	5.613	
	Procedure $\mathcal{M}_{\mathrm{NR}}$	Stops $= 16$	0.7513	$4.895^\dagger$	2.513	5.740	0.98
	Procedure $\mathcal{M}_{\mathrm{BG}}$	$n_{\mathrm{BG}} = 9$	0.7517	4.907	2.526	5.747	1.22
	Procedure $\mathcal{M}_{\mathrm{RA}'}$	$n_{\mathrm{RA}'} = 9, r = 2, t = 4$	0.7513	4.895	2.513	5.740	0.98
	Procedure $\mathcal{M}_{\mathrm{RA}}$	$r = 3, t = 3$	0.7541	5.167	1.468	5.864	6.60
	Procedure $\mathcal{M}_{\mathrm{C}'}$	$n_{\mathrm{C}'} = 8, t = 3$	0.7508	5.154	1.437	5.826	6.32
	Procedure $\mathcal{M}_{\mathrm{BK}}$	$n_{\mathrm{BK}} = 8$	0.7701	6.430	1.099	6.911	32.64
	Procedure $\mathcal{M}_{\mathrm{BEM}}$	$n_{\mathrm{BEM}} = 8$	0.7701	8.000	0.000	8.000	65.03

APPENDIX C

FLOW CONSERVATION CONSTRAINTS

C.1 Flow Conservation Reformulation Algorithm

This appendix describes an algorithmic approach for efficiently identifying the origination left-lexicographic nodes and determining the required coefficients for the flow conservation constraints. To develop such a method, we need to introduce a slightly different notation for our node representation, one that accounts for groupings in our left-lexicographic node vectors. A *group* is simply a set of vector components that are equal. Since all of our nodal vectors are left-lexicographic, groups must consist of adjacent vector components. Let $\mathcal{G} = \{\mathbf{g^1}, \mathbf{g^2}, \ldots, \mathbf{g^G}\}$ be the set of groups in a particular node and $G = |\mathcal{G}|$, the number of distinct groups, with $1 \leq G \leq k$. Within group $\mathbf{g^i}$, let g^i_j represent the jth element of the ith group. For example, consider the node $(4, 4, 2, 1, 1)$ with $k = 5$. Then

$$\mathbf{g^1} = (4, 4),\ \mathbf{g^2} = (2),\ \mathbf{g^3} = (1, 1),$$

with $G = 3$ and $g^1_1 = 4 = g^1_2$, etc.

Then we can use the following to describe a node:

$$(g^1_1, \ldots, g^1_{m_1}, g^2_1, \ldots, g^2_{m_2}, \ldots, g^G_1, \ldots, g^G_{m_G}),$$

where the commas still separate components of the vector (i.e., the number of successes for each alternative), and $g^i_1, \ldots, g^i_{m_i}$ are the m_i members of group $\mathbf{g^i}$ for which $g^i_1 = \cdots = g^i_{m_i}$. The last equality allows us to identify with any group $\mathbf{g^i}$ an integer, g^i, that represents the number of successes for each of the members of the group. Since the node is left-lexicographic and the groups are distinct, $g^1 > g^2 > \cdots > g^G$. Note that $\sum_{i=1}^{G} m_i = k$, the total number of alternatives.

Definition C.1 When we use the terms *coming into* and *terminating at* at group member g^i_j, when referring to an arc, we mean an arc representing the success of an alternative that previously had $g^i_j - 1$ successes and now has g^i_j successes.

Thus, an arc coming into group member g_2^2 corresponds to the arc representing a success of the $(m_1 + 2)$th alternative of the origination node, where m_1 is the number of members of the first group of the *terminating* node. Continuing the previous example with node $(4, 4, 2, 1, 1)$, an arc coming into group member $g_1^3 = 1$ corresponds to the arc representing a success of alternative $m_1 + m_2 + 1 = 2 + 1 + 1 = 4$ of $(4, 4, 2, 0, 1)$. This distinction is important because the grouping structures of the origination nodes will not be the same as that in the terminating nodes, so we are using the group structure of the terminating node as our frame of reference.

We are now in a position to introduce and prove a sequence of statements that will lead to our expression for the coefficients of the origination nodes.

For a left-lexicographic node, an incoming arc to a singleton group originates from a left-lexicographic node. To see that this is true, consider a singleton group, $\mathbf{g^i}$, in the left-lexicographic node under consideration. Assume that $g^i > 0$; otherwise, there is no origination node into group $\mathbf{g^i}$. Only one arc terminates at $\mathbf{g^i}$, and it originated at a node of the form $(\ldots, g_{m_{i-1}}^{i-1}, g^i - 1, g_1^{i+1}, \ldots)$. Since group elements are integers, and groups are distinct, then $g^{i-1} > g^i - 1 \geq g^{i+1}$. Thus, the originating node is left-lexicographic, although its group structure may be different from the terminating node. It is also clear that if $\mathbf{g^i}$ is the first or last group, this result still holds. Since $\mathbf{g^i}$ is a singleton, it cannot be the only group, as $k \geq 2$. $\square$

A direct result of this statement is that, for a left-lexicographic node in which all groups are singletons, all of its incoming arcs originate at left-lexicographic nodes.

For a left-lexicographic node, all arcs coming into the same non-singleton group in the node originated from a permutation of the same left-lexicographic node. To see that this is true, consider node $(\ldots, g_{m_{i-1}}^{i-1}, g_1^i, g_2^i, \ldots, g_{m_i}^i, g_1^{i+1}, \ldots)$. We again assume that $g^i > 0$. Any arc coming into group $\mathbf{g^i}$ must originate from a node with the same vector as the terminal node except for one component being decremented by one in a position corresponding to some $g_j^i \in \mathbf{g^i}$ in the terminal node. That implies that all origination nodes to group $\mathbf{g^i}$ have members from the set $\{g^i, ..., g^i, g^i - 1\}$ of cardinality m_i in the same positions as group $\mathbf{g^i}$ in the terminal node (but with m_i different orderings or permutations). Similar to the

reasoning for the previous statement, $g^i - 1 \geq g^{i+1}$. Therefore, all such origination nodes are permutations of the same left-lexicographic node,

$$(\ldots, g^{i-1}_{m_{i-1}}, g^i_1, \ldots, g^i_{m_i-1}, g^i - 1, g^{i+1}_1, \ldots, g^{i+1}_{m_{i+1}}, \ldots).$$

Again, it is clear that if $\mathbf{g^i}$ is the first, last, or only group, this result still applies. □

Let $\boldsymbol{\eta}'$ be a left-lexicographic node and $\pi(\boldsymbol{\eta}')$ be the set of permutations of $\boldsymbol{\eta}'$. Then, *for a left-lexicographic node $\boldsymbol{\eta}'$, only the arc terminating at the last member of a particular group originated at a left-lexicographic node. Arcs terminating at any other member of that group originated from a member of $\pi(\boldsymbol{\eta}')$, with no two arcs sharing the same origination node.* This statement is a direct result of the previous statements. Given the incoming arcs to group $\mathbf{g^i}$, the originating nodes must be from the ordered set:

$$\begin{aligned}\{ \quad & (\ldots, g^{i-1}_{m_{i-1}}, g^i_1 - 1, g^i_2, \ldots, g^i_{m_i}, g^{i+1}_1, \ldots), \\ & (\ldots, g^{i-1}_{m_{i-1}}, g^i_1, g^i_2 - 1, \ldots, g^i_{m_i}, g^{i+1}_1, \ldots), \\ & \qquad\qquad \vdots \\ & (\ldots, g^{i-1}_{m_{i-1}}, g^i_1, g^i_2, \ldots, g^i_{m_i} - 1, g^{i+1}_1, \ldots) \quad \}.\end{aligned}$$

Only the last node is left-lexicographic, since $g^i_j - 1 < g^i_\ell$, $j \neq \ell$, and it corresponds to the arc terminating at $g^i_{m_i}$. This reduces to our first statement when the group is a singleton. □

We are interested in the inbound flow along the arcs, $f^i_{\boldsymbol{\eta}^\nu}$, $1 \leq \nu \leq k$. Our previous discussion gives us a method to identify the incoming nodes, but we must still represent all incoming flow from each left-lexicographic node in terms of the alternative 1 success arc only.

Let us again consider the incoming arcs to a particular group $\mathbf{g^i}$ within node

$$(\ldots, g^{i-1}_{m_{i-1}}, g^i_1, \ldots, g^i_j, \ldots, g^i_{m_i}, g^{i+1}_1, \ldots).$$

Specifically, consider group member g^i_j. We will call the position in the nodal vector represented by this group member by its arc index, ν^i_j, with the indexing determined by the

terminal node, not the origination node. We can calculate this index as follows:

$$\nu_j^i = j + \sum_{\ell=1}^{i-1} m_\ell.$$

The node from which this arc originates has the following form:

$$(\ldots, g_{m_{i-1}}^{i-1}, g_1^i, \ldots, g_j^i - 1, \ldots, g_{m_i}^i, g_1^{i+1}, \ldots).$$

From (4.4), we calculate the flow along this arc by:

$$\begin{aligned} f_{\boldsymbol{\eta}^\nu}^{\nu_j^i} &= \theta_{\nu_j^i}{}^{g_{m_i}^i - (g_j^i - 1)}\, \theta_{\nu_{m_i}^i}{}^{g_j^i - 1 - g_{m_i}^i} f_{L(\boldsymbol{\eta}^\nu)}^{\nu_j^i}, \\ &= \frac{\theta_{\nu_j^i}}{\theta_{\nu_{m_i}^i}} f_{L(\boldsymbol{\eta}^\nu)}^{\nu_j^i}, \end{aligned} \tag{C.1}$$

since the only components being 'swapped' to make the node left-lexicographic are $g_j^i - 1$ and $g_{m_i}^i$, and since $g_j^i = g_{m_i}^i$. Note that we are slightly simplifying notation by replacing $\boldsymbol{\eta}^{\nu_j^i}$ by $\boldsymbol{\eta}^\nu$.

Now, we need the inbound flow to be converted to $\nu = 1$. Using (4.1), we get:

$$f_{L(\boldsymbol{\eta}^\nu)}^{\nu_j^i} = \frac{1}{\theta_{\nu_j^i}} f_{L(\boldsymbol{\eta}'_\nu)}^1,$$

which we substitute into (C.1) to get

$$\begin{aligned} f_{\boldsymbol{\eta}^\nu}^{\nu_j^i} &= \left(\frac{\theta_{\nu_j^i}}{\theta_{\nu_{m_i}^i}}\right)\left(\frac{1}{\theta_{\nu_j^i}}\right) f_{L(\boldsymbol{\eta}^\nu)}^1, \\ &= \frac{1}{\theta_{\nu_{m_i}^i}} f_{L(\boldsymbol{\eta}^\nu)}^1. \end{aligned} \tag{C.2}$$

Note that if group member g_j^i is the last member of the group (i.e., if its originating node is already left-lexicographic), (C.2) still applies. Then, for any group $\mathbf{g^i}$, all m_i originating nodal arc variables will have the same coefficient so that the total incoming flow into group $\mathbf{g^i}$ is

$$\frac{m_i}{\theta_{\nu_{m_i}^i}} f_{L(\boldsymbol{\eta}^\nu)}^1. \tag{C.3}$$

Across the groups, the overall flow into the left-lexicographic node $\boldsymbol{\eta}'$ is

$$\sum_{i=1}^{G} \frac{m_i}{\theta_{\nu_{m_i}^i}} f_{L(\boldsymbol{\eta}^{\nu_{m_i}^i})}^1. \tag{C.4}$$

While the notation is admittedly (but necessarily) complicated, the final equation is not difficult to understand. The index $\nu^i_{m_i}$ is merely an integer representing the position of the last member of the ith group. Then $L(\boldsymbol{\eta}^{\nu^i_{m_i}})$ is the left-lexicographic permutation of the ith (of G) origination node, which is created by subtracting 1 from vector component $\nu^i_{m_i}$ of node $\boldsymbol{\eta}'$. This completes the identification of the origination nodes. The coefficient in the summation is simply the number of members in the ith group divided by the probability ratio (a constant) of the last member of that group, i.e., alternative $\nu^i_{m_i}$.

We have shown that the knowledge of the grouping structure of a node allows us to identify all of the $G \leq k$ incoming left-lexicographic nodes rapidly and to calculate the coefficients of the G incoming arc variables efficiently. This result leads to a greatly simplified algorithm for constructing the flow conservation constraint matrix.

C.2 Information State Model

As we noted in Remark 4.2, we could have approached the problem of representing our network with only left-lexicographic nodes in a different way — this one involving information states. In this section, we show that alternate method applied to the derivation of the flow conservation constraints. These techniques can be similarly applied to the objective function and the other constraints to derive the complete LP and MIP reformulations. We will use the same notation as the main paper, with exceptions noted. Also, we show the derivations in terms of the p_i, $i = 1, \ldots, k$. Conversion to a form with $\theta_i = p_1/p_i$, $i = 1, \ldots, k$, like that in the main portion of this thesis, is straightforward.

Let $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k) \in \mathcal{N}'$ be an information state, a nonnegative integer vector such that $\eta_i \geq \eta_{i+1}$ and $\sum_{i=1}^k \eta_i \leq b$ where b is the budget limit. To be in state $\boldsymbol{\eta}$ means that the components of $\boldsymbol{\eta}$ are the numbers of successes of the alternatives, ordered from largest to smallest. Corresponding to an information state $\boldsymbol{\eta}$ is a set of actual state vectors $\boldsymbol{v}$, each a permutation of $\boldsymbol{\eta}$. The component v_i is the number of alternative i successes. (Aside: Unless $\eta_i > \eta_{i+1}$ $\forall i$ there are fewer than $k!$ actual state vectors associated with the information vector $\boldsymbol{\eta}$. The number is actually $k!/\prod_{j \in J} G_j!$ where G_j is the cardinality of group j of vector $\boldsymbol{\eta}$.) Denote the set of state vectors associated with $\boldsymbol{\eta}$ as $V(\boldsymbol{\eta})$.

Symmetry dictates that given that we are in information state $\boldsymbol{\eta}$, the conditional probability of being in state $\boldsymbol{v}$ is proportional to $\prod_{i=1}^{k} {p_i}^{v_i}$. All $\boldsymbol{v} \in V(\boldsymbol{\eta})$ have the same proportionality constant. Hence the conditional probability of being in actual state $\boldsymbol{v}$ given that we are in information state $\boldsymbol{\eta}$ is

$$\prod_{i=1}^{k} {p_i}^{v_i} / \sum_{\boldsymbol{v} \in V(\boldsymbol{\eta})} \prod_{i=1}^{k} {p_i}^{v_i}.$$

The flow conservation equation at node $\boldsymbol{\eta}$ is written with variables $f_{\boldsymbol{\eta}}$ and $f^1_{\boldsymbol{\eta}}$.

- $f_{\boldsymbol{\eta}}$ is the flow to the sink from $\boldsymbol{\eta}$, i.e. the absolute probability of terminating sampling at $\boldsymbol{\eta}$.
- $f^1_{\boldsymbol{\eta}}$ is the flow to actual state $\boldsymbol{\eta}+\boldsymbol{e}_1$ from actual state $\boldsymbol{\eta}$ (in the network from Chapter 3), where $\boldsymbol{e}_1$ is the first unit vector $(1, 0, 0, \ldots, 0)$.

Therefore the flow to actual state $\boldsymbol{\eta}+\boldsymbol{e}_j$ from actual state $\boldsymbol{\eta}$ is $f^1_{\boldsymbol{\eta}} p_j / p_1$, by the definition of $\boldsymbol{p}$. The flow to actual state $\boldsymbol{v} + \boldsymbol{e}_1$ from actual state $\boldsymbol{v}$ is

$$f^1_{\boldsymbol{\eta}} \prod_{i=1}^{k} {p_i}^{v_i} / \prod_{i=1}^{k} {p_i}^{\eta_i}$$

because you must multiply by the ratio of the conditional probability of being in actual state $\boldsymbol{v}$ to the conditional probability of being in actual state $\boldsymbol{\eta}$.

The flow to actual state $\boldsymbol{v} + \boldsymbol{e}_j$ from actual state $\boldsymbol{v}$ is therefore

$$\frac{p_j}{p_1} f^1_{\eta} \prod_{i=1}^{k} {p_i}^{v_i} / \prod_{i=1}^{k} {p_i}^{\eta_i}.$$

The flow conservation equation at information node $\boldsymbol{\eta}$ requires that flow in equals flow out. Flow out is

$$f_{\boldsymbol{\eta}} + \frac{1}{p_1} f^1_{\boldsymbol{\eta}} \left(\sum_{\boldsymbol{v} \in V(\boldsymbol{\eta})} \sum_{j=1}^{k} p_j \prod_{i=1}^{k} {p_i}^{v_i - \eta_i} \right).$$

To write the flow in we need a little more notation. For any actual state vector $\boldsymbol{v}$ let $L(\boldsymbol{v})$ denote the left-lexicographic permutation of $\boldsymbol{v}$. For any actual state vector $\boldsymbol{v}$ and integer $j : 1 \leq j \leq k$ let $d(\boldsymbol{v}, j)$ be the ordinate at which $L(\boldsymbol{v})$ and $L(\boldsymbol{v} - \boldsymbol{e}_j)$ differ. Then

flow in

$$
\begin{aligned}
&= \sum_{\boldsymbol{v}\in V(\boldsymbol{\eta})} \sum_{j=1}^{k} \text{flow to actual state } \boldsymbol{v} \text{ from actual state } \boldsymbol{v}-\boldsymbol{e}_j \\
&= \sum_{\boldsymbol{v}\in V(\boldsymbol{\eta})} \sum_{j=1}^{k} f^1_{L(\boldsymbol{v}-\boldsymbol{e}_j)} \frac{p_j}{p_1} \prod_{i=1}^{k} {p_i}^{(\boldsymbol{v}-\boldsymbol{e}_j)_i - L(\boldsymbol{v}-\boldsymbol{e}_j)_i} \\
&= \sum_{\boldsymbol{v}\in V(\boldsymbol{\eta})} \sum_{j=1}^{k} f^1_{L(\boldsymbol{v}-\boldsymbol{e}_j)} \frac{p_{d(\boldsymbol{v},j)}}{p_1} \prod_{i=1}^{k} {p_i}^{v_i - L(\boldsymbol{v})_i}.
\end{aligned}
$$

APPENDIX D

IZ REGION PROBABILITIES

In §D.1 and §D.2, we calculate the probability that $\boldsymbol{p} \in \Omega_{\text{IZ}3}$ and $\boldsymbol{p} \in \Omega_{\text{IZ}2}$, respectively. In §D.3, we briefly describe (but do not show) the corresponding derivations for $k = 4$ and present the resulting equations.

D.1 $\mathsf{P}_{\text{u}}(\boldsymbol{p} \in \Omega_{\text{IZ}3})$ *for* $k = 3$

Our process for finding the volume of each $\Omega_{\text{IZ}m}$ polytope is similar to the methodology we followed for the volume of the PZ in §7.1, in that we begin by identifying the vertices of the polytope and concern ourselves with only the volume of one permutation of the input space, $p_1 \geq p_2 \geq p_3$. In that permutation of the space, the $\Omega'_{\text{IZ}3}$ polytope is defined as follows:

$$\Omega'_{\text{IZ}3} \equiv \left\{ (p_1, p_2, p_3) \in \mathbb{R}^3 : \; p_1 \leq \theta^\star \, p_3, \; p_1 \geq p_2, \; p_2 \geq p_3, \; p_3 \geq 0, \; \textstyle\sum_{i=1}^{3} p_i = 1 \right\}, \quad \text{(D.1)}$$

where the prime in $\Omega'_{\text{IZ}3}$ indicates the permutation of the space.

We have only one equality and must therefore choose among the four inequalities to determine the two remaining equations that will define the feasible solution. We remove two inequalities at a time, set the remaining two inequalities to equalities, solve the resulting set of equations to find each basic solution, and check to see if the solution satisfies the two inequalities that we removed. Following that procedure for the $\binom{4}{2}$ possible sets of equations, we identify three vertices:

$$\begin{aligned}
\boldsymbol{v}_1 &= \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right), \\
\boldsymbol{v}_2 &= \left(\frac{\theta^\star}{2\theta^\star + 1}, \frac{\theta^\star}{2\theta^\star + 1}, \frac{1}{2\theta^\star + 1} \right), \\
\boldsymbol{v}_3 &= \left(\frac{\theta^\star}{\theta^\star + 2}, \frac{1}{\theta^\star + 2}, \frac{1}{\theta^\star + 2} \right).
\end{aligned}$$

The three vertices define a two-dimensional triangle, or simplex, in $\mathbb{R}^3$. Therefore, we can use the following relationship to determine the volume (a term we use generally even

when dealing with area) of the simplex:

$$V = \frac{1}{2}\sqrt{\begin{vmatrix} \boldsymbol{v}_{1,1} & \boldsymbol{v}_{1,2} & 1 \\ \boldsymbol{v}_{2,1} & \boldsymbol{v}_{2,2} & 1 \\ \boldsymbol{v}_{3,1} & \boldsymbol{v}_{3,2} & 1 \end{vmatrix} + \begin{vmatrix} \boldsymbol{v}_{1,2} & \boldsymbol{v}_{1,3} & 1 \\ \boldsymbol{v}_{2,2} & \boldsymbol{v}_{2,3} & 1 \\ \boldsymbol{v}_{3,2} & \boldsymbol{v}_{3,3} & 1 \end{vmatrix} + \begin{vmatrix} \boldsymbol{v}_{1,3} & \boldsymbol{v}_{1,1} & 1 \\ \boldsymbol{v}_{2,3} & \boldsymbol{v}_{2,1} & 1 \\ \boldsymbol{v}_{3,3} & \boldsymbol{v}_{3,1} & 1 \end{vmatrix}}, \tag{D.2}$$

where $\boldsymbol{v}_{1,i}$ is the ith component of vertex $\boldsymbol{v}_1$ and $|\boldsymbol{A}|$ is the determinant of matrix $\boldsymbol{A}$. Application of (D.2) to Ω'_{IZ3} gives us

$$V_{\text{IZ3}'} = \frac{1}{2}\sqrt{\begin{vmatrix} \frac{1}{3} & \frac{1}{3} & 1 \\ \frac{\theta^\star}{2\theta^\star+1} & \frac{\theta^\star}{2\theta^\star+1} & 1 \\ \frac{\theta^\star}{\theta^\star+2} & \frac{1}{\theta^\star+2} & 1 \end{vmatrix} + \begin{vmatrix} \frac{1}{3} & \frac{1}{3} & 1 \\ \frac{\theta^\star}{2\theta^\star+1} & \frac{1}{2\theta^\star+1} & 1 \\ \frac{1}{\theta^\star+2} & \frac{1}{\theta^\star+2} & 1 \end{vmatrix} + \begin{vmatrix} \frac{1}{3} & \frac{1}{3} & 1 \\ \frac{1}{2\theta^\star+1} & \frac{\theta^\star}{2\theta^\star+1} & 1 \\ \frac{1}{\theta^\star+2} & \frac{\theta^\star}{\theta^\star+2} & 1 \end{vmatrix}},$$

which reduces to

$$V_{\text{IZ3}'} = \frac{(\theta^\star-1)^2}{2\sqrt{3}(2\theta^\star+1)(\theta^\star+2)}. \tag{D.3}$$

In order to calculate the required probability under our assumptions, we need the volume of the permutation of Ω in which $p_1 \geq p_2 \geq p_3$. In that permutation of the space, the Ω' polytope is defined as follows:

$$\Omega' \equiv \left\{(p_1, p_2, p_3) \in \mathbb{R}^3 : \; p_1 \geq p_2, \; p_2 \geq p_3, \; p_3 \geq 0, \; \textstyle\sum_{i=1}^3 p_i = 1\right\}, \tag{D.4}$$

which has the following three vertices:

$$\begin{aligned} \boldsymbol{v}_1 &= (1\,,\,0\,,\,0), \\ \boldsymbol{v}_2 &= (1/2\,,\,1/2\,,\,0), \\ \boldsymbol{v}_3 &= (1/3\,,\,1/3\,,\,1/3). \end{aligned}$$

We use (D.2) to solve for the volume of Ω' (but omit the algebra for brevity):

$$V_{\Omega'} = \frac{\sqrt{3}}{12}. \tag{D.5}$$

We now have the volumes necessary to calculate the desire probability under our uniformity assumption:

$$\begin{aligned} \mathsf{P}_{\text{u}}(\boldsymbol{p} \in \Omega_{\text{IZ3}}) &= \frac{V_{\text{IZ3}'}}{V_{\Omega'}} \\ &= \frac{2(\theta^\star-1)^2}{(2\theta^\star+1)(\theta^\star+2)}. \end{aligned} \tag{D.6}$$

D.2 $\mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{IZ2}})$ *for* $k = 3$

For this probability, we could simply use the information we already have and the law of total probability:

$$\begin{aligned}
\mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{IZ2}}) &= 1 - \mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{IZ3}}) - \mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{PZ}}) \\
&= 1 - \frac{2(\theta^\star - 1)^2}{(2\theta^\star + 1)(\theta^\star + 2)} - 3! \, \frac{1}{(\theta^\star + 1)(\theta^\star + 2)} \\
&= \frac{3(3\theta^\star + 2)(\theta^\star - 1)}{(\theta^\star + 1)(\theta^\star + 2)(2\theta^\star + 1)}. \qquad \text{(D.7)}
\end{aligned}$$

However, we derive (D.7) using the same procedure as we did in §D.1, in order to show a unique aspect of the volumes involved in this space. We are interested in the following permutation of Ω_{IZ2}:

$$\Omega'_{\mathrm{IZ2}} \equiv \left\{ (p_1, p_2, p_3) \in \mathbb{R}^3 : \begin{array}{l} p_1 \geq \theta^\star \, p_3, \; p_1 \leq \theta^\star \, p_2, \; p_1 \geq p_2, \\ p_3 \geq 0, \; \sum_{i=1}^3 p_i = 1 \end{array} \right\}. \qquad \text{(D.8)}$$

Since we have one equality, we must choose among the four inequalities to determine the two remaining equations that will define each feasible solution. Following that procedure for the $\binom{4}{2}$ possible sets of equations, we identify four vertices:

$$\begin{aligned}
\boldsymbol{v}_1 &= \left(\frac{1}{2}, \frac{1}{2}, 0 \right), \\
\boldsymbol{v}_2 &= \left(\frac{\theta^\star}{\theta^\star + 1}, \frac{1}{\theta^\star + 1}, 0 \right), \\
\boldsymbol{v}_3 &= \left(\frac{\theta^\star}{\theta^\star + 2}, \frac{1}{\theta^\star + 2}, \frac{1}{\theta^\star + 2} \right), \\
\boldsymbol{v}_4 &= \left(\frac{\theta^\star}{2\theta^\star + 1}, \frac{\theta^\star}{2\theta^\star + 1}, \frac{1}{2\theta^\star + 1} \right).
\end{aligned}$$

Unlike our volume for Ω_{IZ3}, we have four vertices, so we must divide the region into two simplices to calculate the volume. We use visual methods to determine appropriate simplices.

The first simplex is defined by vertices $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$. Using (D.2), we get

$$V^1_{\mathrm{IZ2}'} = \frac{\sqrt{3}(\theta^\star - 1)}{4(\theta^\star + 1)(\theta^\star + 2)}. \qquad \text{(D.9)}$$

The second simplex is defined by vertices $\boldsymbol{v}_1$, $\boldsymbol{v}_3$, and $\boldsymbol{v}_4$. Using (D.2), we get

$$V^2_{\mathrm{IZ2}'} = \frac{\sqrt{3}(\theta^\star - 1)}{4(2\theta^\star + 1)(\theta^\star + 2)}. \qquad \text{(D.10)}$$

We combine the volumes (D.9) and (D.10) of the two simplices to get the total volume:

$$\begin{aligned} V_{\text{IZ2}'} &= V^1_{\text{IZ2}'} + V^2_{\text{IZ2}'} \\ &= \frac{\sqrt{3}(\theta^\star - 1)}{4(\theta^\star + 1)(\theta^\star + 2)} + \frac{\sqrt{3}(\theta^\star - 1)}{4(2\theta^\star + 1)(\theta^\star + 2)} \\ &= \frac{\sqrt{3}(3\theta^\star + 2)(\theta^\star - 1)}{4(\theta^\star + 1)(\theta^\star + 2)(2\theta^\star + 1)}. \end{aligned}$$

We now have the volumes necessary to calculate the desire probability under our uniformity assumption:

$$\begin{aligned} \mathsf{P}_{\text{u}}(\boldsymbol{p} \in \Omega_{\text{IZ2}}) &= \frac{V_{\text{IZ2}'}}{V_{\Omega'}} \\ &= \frac{3(3\theta^\star + 2)(\theta^\star - 1)}{(\theta^\star + 1)(\theta^\star + 2)(2\theta^\star + 1)}, \end{aligned} \tag{D.11}$$

which is equivalent to what we calculated in (D.7).

D.3 $\mathsf{P}_{\text{u}}(\boldsymbol{p} \in \Omega_{\text{IZ}m})$ *for* $\boldsymbol{k = 4}$

The derivations for $k = 4$ follow the same methodology as that of §D.1 and §D.2. In this section, we will show the results of each major step for the volumes, but leave out a majority of the algebra and relationships that we have already shown previously.

First, (D.2) is not appropriate for the three-dimensional volumes in $\mathbb{R}^4$ that we must work with here. That equation, adapted for four-dimensional vertices, is:

$$V = \frac{1}{6}\sqrt{\begin{aligned} &\begin{vmatrix} \boldsymbol{v}_{1,1} & \boldsymbol{v}_{1,2} & \boldsymbol{v}_{1,3} & 1 \\ \boldsymbol{v}_{2,1} & \boldsymbol{v}_{2,2} & \boldsymbol{v}_{2,3} & 1 \\ \boldsymbol{v}_{3,1} & \boldsymbol{v}_{3,2} & \boldsymbol{v}_{3,3} & 1 \\ \boldsymbol{v}_{4,1} & \boldsymbol{v}_{4,2} & \boldsymbol{v}_{4,3} & 1 \end{vmatrix} + \begin{vmatrix} \boldsymbol{v}_{1,2} & \boldsymbol{v}_{1,3} & \boldsymbol{v}_{1,4} & 1 \\ \boldsymbol{v}_{2,2} & \boldsymbol{v}_{2,3} & \boldsymbol{v}_{2,4} & 1 \\ \boldsymbol{v}_{3,2} & \boldsymbol{v}_{3,3} & \boldsymbol{v}_{3,4} & 1 \\ \boldsymbol{v}_{4,2} & \boldsymbol{v}_{4,3} & \boldsymbol{v}_{4,4} & 1 \end{vmatrix} \\ &+ \begin{vmatrix} \boldsymbol{v}_{1,3} & \boldsymbol{v}_{1,4} & \boldsymbol{v}_{1,1} & 1 \\ \boldsymbol{v}_{2,3} & \boldsymbol{v}_{2,4} & \boldsymbol{v}_{2,1} & 1 \\ \boldsymbol{v}_{3,3} & \boldsymbol{v}_{3,4} & \boldsymbol{v}_{3,1} & 1 \\ \boldsymbol{v}_{4,3} & \boldsymbol{v}_{4,4} & \boldsymbol{v}_{4,1} & 1 \end{vmatrix} + \begin{vmatrix} \boldsymbol{v}_{1,4} & \boldsymbol{v}_{1,1} & \boldsymbol{v}_{1,2} & 1 \\ \boldsymbol{v}_{2,4} & \boldsymbol{v}_{2,1} & \boldsymbol{v}_{2,2} & 1 \\ \boldsymbol{v}_{3,4} & \boldsymbol{v}_{3,1} & \boldsymbol{v}_{3,2} & 1 \\ \boldsymbol{v}_{4,4} & \boldsymbol{v}_{4,1} & \boldsymbol{v}_{4,2} & 1 \end{vmatrix} \end{aligned}}, \tag{D.12}$$

D.3.1 Volume of Ω' for $k = 4$

This space Ω' in $\mathbb{R}^4$ is defined by:

$$\Omega' \equiv \left\{ (p_1, p_2, p_3, p_4) \in \mathbb{R}^4 : \begin{array}{l} p_1 \geq p_2,\ p_2 \geq p_3,\ p_3 \geq p_4, \\ p_4 \geq 0,\ \sum_{i=1}^4 p_i = 1 \end{array} \right\},$$

with the following vertices:

$$\begin{aligned} \boldsymbol{v}_1 &= (1\,,\,0\,,\,0\,,\,0)\,, \\ \boldsymbol{v}_1 &= (1/2\,,\,1/2\,,\,0\,,\,0)\,, \\ \boldsymbol{v}_1 &= (1/3\,,\,1/3\,,\,1/3\,,\,0)\,, \\ \boldsymbol{v}_1 &= (1/4\,,\,1/4\,,\,1/4\,,\,1/4)\,. \end{aligned}$$

The volume of the simplex defined by the vertices is

$$V_{\Omega'} = \frac{1}{72}. \tag{D.13}$$

D.3.2 $\mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{IZ4}})$ for $k = 4$

This space is defined by:

$$\Omega'_{\mathrm{IZ4}} \equiv \left\{ (p_1, p_2, p_3, p_4) \in \mathbb{R}^4 : \begin{array}{l} p_1 \leq \theta^\star\, p_4,\ p_1 \geq p_2,\ p_2 \geq p_3, \\ p_3 \geq p_4,\ p_4 \geq 0,\ \sum_{i=1}^4 p_i = 1 \end{array} \right\},$$

with the following vertices:

$$\begin{aligned} \boldsymbol{v}_1 &= \left(\frac{1}{4}\,,\,\frac{1}{4}\,,\,\frac{1}{4}\,,\,\frac{1}{4}\right), \\ \boldsymbol{v}_2 &= \left(\frac{\theta^\star}{\theta^\star+3}\,,\,\frac{1}{\theta^\star+3}\,,\,\frac{1}{\theta^\star+3}\,,\,\frac{1}{\theta^\star+3}\right), \\ \boldsymbol{v}_3 &= \left(\frac{\theta^\star}{2\theta^\star+2}\,,\,\frac{\theta^\star}{2\theta^\star+2}\,,\,\frac{1}{2\theta^\star+2}\,,\,\frac{1}{2\theta^\star+2}\right), \\ \boldsymbol{v}_4 &= \left(\frac{\theta^\star}{3\theta^\star+1}\,,\,\frac{\theta^\star}{3\theta^\star+1}\,,\,\frac{\theta^\star}{3\theta^\star+1}\,,\,\frac{1}{3\theta^\star+1}\right). \end{aligned}$$

The volume of the simplex defined by the vertices is

$$V_{\mathrm{IZ4}'} = \frac{(\theta^\star - 1)^3}{12(3\theta^\star+1)(2\theta^\star+2)(\theta^\star+3)}. \tag{D.14}$$

We now have the volumes necessary to calculate the desire probability under our uniformity assumption:

$$\begin{aligned} \mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{IZ4}}) &= \frac{V_{\mathrm{IZ4}'}}{V_{\Omega'}} \\ &= \frac{6(\theta^\star - 1)^3}{(3\theta^\star + 1)(2\theta^\star + 2)(\theta^\star + 3)}. \end{aligned} \tag{D.15}$$

D.3.3 $\mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{IZ3}})$ for $k = 4$

This space is defined by:

$$\Omega'_{\mathrm{IZ3}} \equiv \left\{ (p_1, p_2, p_3, p_4) \in \mathbb{R}^4 : \begin{array}{l} p_1 \geq \theta^\star\, p_4,\ p_1 \leq\ \theta^\star p_3,\ p_1 \geq\ p_2, \\ p_2 \geq p_3,\ p_4 \geq 0,\ \sum_{i=1}^4 p_i = 1 \end{array} \right\},$$

with the following vertices:

$$\begin{aligned} \boldsymbol{v}_1 &= \left(\frac{\theta^\star}{3\theta^\star + 1}, \frac{\theta^\star}{3\theta^\star + 1}, \frac{\theta^\star}{3\theta^\star + 1}, \frac{1}{3\theta^\star + 1} \right), \\ \boldsymbol{v}_2 &= \left(\frac{\theta^\star}{2\theta^\star + 2}, \frac{\theta^\star}{2\theta^\star + 2}, \frac{1}{2\theta^\star + 2}, \frac{1}{2\theta^\star + 2} \right), \\ \boldsymbol{v}_3 &= \left(\frac{\theta^\star}{2\theta^\star + 1}, \frac{\theta^\star}{2\theta^\star + 1}, \frac{1}{2\theta^\star + 1}, 0 \right), \\ \boldsymbol{v}_4 &= \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0 \right), \\ \boldsymbol{v}_5 &= \left(\frac{\theta^\star}{\theta^\star + 3}, \frac{1}{\theta^\star + 3}, \frac{1}{\theta^\star + 3}, \frac{1}{\theta^\star + 3} \right), \\ \boldsymbol{v}_6 &= \left(\frac{\theta^\star}{\theta^\star + 2}, \frac{1}{\theta^\star + 2}, \frac{1}{\theta^\star + 2}, 0 \right). \end{aligned}$$

The polygon defined by the vertices is not a simplex, but can be divided into three three-dimensional simplices. The first is defined by vertices $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_4$, and $\boldsymbol{v}_5$. Its volume is

$$V^1_{\mathrm{IZ3}'} = \frac{(\theta^\star - 1)^2}{9(3\theta^\star + 1)(2\theta^\star + 2)(\theta^\star + 3)}. \tag{D.16}$$

The next is defined by vertices $\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$, and $\boldsymbol{v}_5$. Its volume is

$$V^2_{\mathrm{IZ3}'} = \frac{(\theta^\star - 1)^2}{9(2\theta^\star + 2)(2\theta^\star + 1)(\theta^\star + 3)}. \tag{D.17}$$

The last is defined by vertices $\boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5$, and $\boldsymbol{v}_6$. Its volume is

$$V^3_{\mathrm{IZ3}'} = \frac{(\theta^\star - 1)^2}{9(2\theta^\star + 1)(\theta^\star + 3)(\theta^\star + 2)}. \tag{D.18}$$

The total volume is the sum of volumes (D.16), (D.17), and (D.18):

$$\begin{aligned} V_{\text{IZ3}'} &= V^1_{\text{IZ3}'} + V^2_{\text{IZ3}'} + V^3_{\text{IZ3}'} \\ &= \frac{(11\theta^{\star 2} + 20\theta^\star + 6)(\theta^\star - 1)^2}{9(3\theta^\star + 1)(2\theta^\star + 2)(2\theta^\star + 1)(\theta^\star + 3)(\theta^\star + 2)}. \end{aligned}$$

We now have the volumes necessary to calculate the desire probability under our uniformity assumption:

$$\begin{aligned} \mathsf{P}_{\text{u}}(\boldsymbol{p} \in \Omega_{\text{IZ3}}) &= \frac{V_{\text{IZ3}'}}{V_{\Omega'}} \\ &= \frac{8(11\theta^{\star 2} + 20\theta^\star + 6)(\theta^\star - 1)^2}{(3\theta^\star + 1)(2\theta^\star + 2)(2\theta^\star + 1)(\theta^\star + 3)(\theta^\star + 2)}. \end{aligned} \tag{D.19}$$

D.3.4 $\mathsf{P}_{\text{u}}(\boldsymbol{p} \in \Omega_{\text{IZ2}})$ for $k = 4$

This space is defined by:

$$\Omega'_{\text{IZ2}} \equiv \left\{ (p_1, p_2, p_3, p_4) \in \mathbb{R}^4 : \begin{array}{l} p_1 \geq \theta^\star\, p_3,\ p_1 \leq\ \theta^\star p_2,\ p_1 \geq\ p_2, \\ p_3 \geq p_4,\ p_4 \geq 0,\ \sum_{i=1}^4 p_i = 1 \end{array} \right\},$$

with the following vertices:

$$\begin{aligned} \boldsymbol{v}_1 &= \left(\frac{\theta^\star}{2\theta^\star + 2}, \frac{\theta^\star}{2\theta^\star + 2}, \frac{1}{2\theta^\star + 2}, \frac{1}{2\theta^\star + 2} \right), \\ \boldsymbol{v}_2 &= \left(\frac{\theta^\star}{2\theta^\star + 1}, \frac{\theta^\star}{2\theta^\star + 1}, \frac{1}{2\theta^\star + 1}, 0 \right), \\ \boldsymbol{v}_3 &= \left(\frac{1}{2}, \frac{1}{2}, 0, 0 \right), \\ \boldsymbol{v}_4 &= \left(\frac{\theta^\star}{\theta^\star + 3}, \frac{1}{\theta^\star + 3}, \frac{1}{\theta^\star + 3}, \frac{1}{\theta^\star + 3} \right), \\ \boldsymbol{v}_5 &= \left(\frac{\theta^\star}{\theta^\star + 2}, \frac{1}{\theta^\star + 2}, \frac{1}{\theta^\star + 2}, 0 \right), \\ \boldsymbol{v}_6 &= \left(\frac{\theta^\star}{\theta^\star + 1}, \frac{1}{\theta^\star + 1}, 0, 0 \right). \end{aligned}$$

The polygon defined by the vertices is not a simplex, but can be divided into three three-dimensional simplices. The first is defined by vertices $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$, and $\boldsymbol{v}_4$. Its volume is

$$V^1_{\text{IZ2}'} = \frac{(\theta^\star - 1)}{6(2\theta^\star + 2)(2\theta^\star + 1)(\theta^\star + 3)}. \tag{D.20}$$

The next is defined by vertices $\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$, and $\boldsymbol{v}_5$. Its volume is

$$V^2_{\text{IZ2}'} = \frac{(\theta^\star - 1)}{6(2\theta^\star + 1)(\theta^\star + 3)(\theta^\star + 2)}. \tag{D.21}$$

The last is defined by vertices $\boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5$, and $\boldsymbol{v}_6$. Its volume is

$$V^3_{\mathrm{IZ}2'} = \frac{(\theta^\star - 1)}{6(\theta^\star + 3)(\theta^\star + 2)(\theta^\star + 1)}. \tag{D.22}$$

The total volume is the sum of volumes (D.20), (D.21), and (D.22):

$$\begin{aligned} V_{\mathrm{IZ}2'} &= V^1_{\mathrm{IZ}2'} + V^2_{\mathrm{IZ}2'} + V^3_{\mathrm{IZ}2'} \\ &= \frac{(7\theta^{\star 2} + 13\theta^\star + 6)(\theta^\star - 1)}{6(2\theta^\star + 2)(2\theta^\star + 1)(\theta^\star + 3)(\theta^\star + 2)(\theta^\star + 1)}. \end{aligned}$$

We now have the volumes necessary to calculate the desire probability under our uniformity assumption:

$$\begin{aligned} \mathsf{P}_{\mathrm{u}}(\boldsymbol{p} \in \Omega_{\mathrm{IZ}2}) &= \frac{V_{\mathrm{IZ}2'}}{V_{\Omega'}} \\ &= \frac{12(7\theta^{\star 2} + 13\theta^\star + 6)(\theta^\star - 1)}{(2\theta^\star + 2)(2\theta^\star + 1)(\theta^\star + 3)(\theta^\star + 2)(\theta^\star + 1)}. \end{aligned} \tag{D.23}$$

REFERENCES

[1] Alam, K., "On selecting the most probable category," *Technometrics*, vol. 13, no. 4, pp. 843–850, 1971.

[2] Alam, K. and Thompson, J. R., "On selecting the least probable multinomial event," *Annals of Mathematical Statistics*, vol. 43, no. 6, pp. 1981–1990, 1972.

[3] Aoshima, M. and Chen, P. Y., "A two-stage procedure for selecting the largest multinomial cell probability when nuisance cell is present," *Sequential Analysis*, vol. 18, no. 2, pp. 143–155, 1999.

[4] Aoshima, M., Chen, P. Y., and Panchapakesan, S., "Sequential procedures for selecting the most probable multinomial cell when a nuisance cell is present," *Communications in Statistics — Theory and Methods*, vol. A32, no. 4, pp. 893–906, 2003.

[5] Bechhofer, R. E., "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *Annals of Mathematical Statistics*, vol. 25, no. 1, pp. 16–39, 1954.

[6] Bechhofer, R. E. and Chen, P. Y., "A note on a curtailed sequential procedure for subset-selection of multinomial cells," *American Journal of Mathematical and Management Sciences*, vol. 11, no. 3/4, pp. 309–324, 1991.

[7] Bechhofer, R. E., Dunnett, C. W., and Sobel, M., "A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance," *Biometrika*, vol. 41, no. 1/2, pp. 170–176, 1954.

[8] Bechhofer, R. E., Elmaghraby, S., and Morse, N., "A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability," *Annals of Mathematical Statistics*, vol. 30, no. 1, pp. 102–119, 1959.

[9] Bechhofer, R. E. and Goldsman, D. M., "On the Ramey-Alam sequential procedure for selecting the multinomial event which has the largest probability," *Communications in Statistics — Simulation and Computation*, vol. B14, no. 2, pp. 263–282, 1985a.

[10] Bechhofer, R. E. and Goldsman, D. M., "Truncation of the Bechhofer-Kiefer-Sobel sequential procedure for selecting the multinomial event which has the largest probability," *Communications in Statistics — Simulation and Computation*, vol. B14, no. 2, pp. 283–315, 1985b.

[11] Bechhofer, R. E. and Goldsman, D. M., "Truncation of the Bechhofer-Kiefer-Sobel sequential procedure for selecting the multinomial event which has the largest probability (II): Extended tables and an improved procedure," *Communications in Statistics — Simulation and Computation*, vol. B15, no. 3, pp. 829–851, 1986.

[12] Bechhofer, R. E., Goldsman, D. M., and Jennison, C., "A single-stage selection procedure for multi-factor multinomial experiments with multiplicativity," *Communications in Statistics — Simulation and Computation*, vol. B18, no. 1, pp. 31–61, 1989.

[13] Bechhofer, R. E., Kiefer, J., and Sobel, M., *Sequential Identification and Ranking Procedures, with Special Reference to Koopman-Darmois Populations*. Statistical Research Monographs, Chicago: University of Chicago Press, 1968.

[14] Bechhofer, R. E. and Kulkarni, R. V., "Closed sequential procedures for selecting the multinomial events which have the largest probabilities," *Communications in Statistics — Theory and Methods*, vol. A13, no. 24, pp. 2997–3031, 1984.

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

[16] Bechhofer, R. E. and Sobel, M., "A single-sample multiple decision procedure for ranking variances of normal populations," *Annals of Mathematical Statistics*, vol. 25, no. 2, pp. 273–289, 1954.

[17] Bertsimas, D. and Tsitsiklis, J. N., *Introduction to Linear Optimization*. Belmont, MA: Athena Scientific, 1997.

[18] Bhandari, S. K. and Ali, M. M., "An asymptotically minimax procedure for selecting the t-best multinomial cells," *Journal of Statistical Planning and Inference*, vol. 38, no. 1, pp. 65–74, 1994.

[19] Bhandari, S. K. and Bose, A., "On selecting the most likely event," *Journal of Statistical Planning and Inference*, vol. 17, no. 2, pp. 227–240, 1987.

[20] Bhandari, S. K. and Bose, A., "Selecting the t-best cells in a multinomial distribution," *Communications in Statistics — Theory and Methods*, vol. A18, no. 9, pp. 3313–3326, 1989.

[21] Branke, J., Chick, S. E., and Schmidt, C., "Selecting a selection procedure," *Management Science*, vol. 53, no. 12, pp. 1916–1932, 2007.

[22] Büeler, B., Enge, A., and Eukuda, K., "Exact volume computation for polytopes: A practical study," in *Polytopes — Combinatorics and Computation* (Kalai, G. and Ziegler, G. M., eds.), Basel, Switzerland: Birkhäuser–Verlag, 2000.

[23] Cacoullos, T. and Sobel, M., "An inverse sampling procedure for selecting the most probable event in a multinomial distribution," in *Multivariate Analysis* (Krishnaiah, P. R., ed.), (New York), pp. 423–455, Academic Press, 1966.

[24] Chen, C. H., "An effective approach to smartly allocate computing budget for discrete event simulation," in *Proceedings of the 34th IEEE Conference on Decision and Control*, vol. 3, pp. 2598–2603, IEEE, 1995.

[25] Chen, P. Y., "Closed inverse sampling procedure for selecting the largest multinomial cell probability," *Communications in Statistics — Simulation and Computation*, vol. B17, no. 3, pp. 969–994, 1988a.

[26] Chen, P. Y., "An integrated formulation for selecting the most probable multinomial cell," *Annals of the Institute of Statistical Mathematics*, vol. 40, no. 3, pp. 615–625, 1988b.

[27] Chen, P. Y., "Truncated selection procedures for the most probable event and the least probable event," *Annals of the Institute of Statistical Mathematics*, vol. 44, no. 4, pp. 613–622, 1992.

[28] Chen, P. Y. and Hsu, L. F., "A composite stopping rule for multinomial subset-selection," *British Journal of Mathematical & Statistical Psychology*, vol. 44, no. 2, pp. 403–411, 1991a.

[29] Chen, P. Y. and Hsu, L. F., "On a sequential subset selection procedure for the least probable multinomial cell," *Communications in Statistics — Theory and Methods*, vol. A20, no. 9, pp. 2845–2862, 1991b.

[30] Chen, P. Y. and Sobel, M., "A new formulation for the multinomial selection problem," *Communications in Statistics — Theory and Methods*, vol. A16, no. 1, pp. 147–180, 1987.

[31] Chen, R. W. and Hwang, F. K., "Some theorems, counterexamples, and conjectures in multinomial selection theory," *Communications in Statistics — Theory and Methods*, vol. A13, no. 10, pp. 1289–1298, 1984.

[32] Chick, S. E. and Gans, N., "Economic analysis of simulation selection problems," *Management Science*, vol. 55, no. 3, pp. 421–437, 2009.

[33] Chick, S. E. and Inoue, K., "New two-stage and sequential procedures for selecting the best simulated system," *Operations Research*, vol. 49, no. 5, pp. 732–743, 2001.

[34] Gibbons, J. D., Olkin, I., and Sobel, M., *Selecting and Ordering Populations: A New Statistical Methodology*. Wiley Series in Probability and Mathematical Statistics, New York: Wiley, 1977.

[35] Gupta, S. S. and Miescke, K. J., "Bayesian look ahead one-stage sampling allocations for selection of the best population," *Journal of Statistical Planning and Inference*, vol. 54, no. 2, pp. 229–244, 1996.

[36] Gupta, S. S. and Nagel, K., "On selection and ranking procedures and order statistics from the multinomial distribution," *Sankhyā: The Indian Journal of Statistics*, vol. B29, no. 1/2, pp. 1–34, 1967.

[37] He, D. H., Chick, S. E., and Chen, C. H., "Opportunity cost and OCBA selection procedures in ordinal optimization for a fixed number of alternative systems," *IEEE Transactions on Systems, Man, and Cybernetics — Applications and Reviews*, vol. C37, no. 5, pp. 951–961, 2007.

[38] Hines, W. W., Montgomery, D. C., Goldsman, D. M., and Borror, C. M., *Probability and Statistics in Engineering*. Hoboken, NJ: Wiley, 2003.

[39] Hong, L. J. and Nelson, B. L., "The tradeoff between sampling and switching: New sequential procedures for indifference-zone selection," *IIE Transactions*, vol. 37, no. 7, pp. 623–634, 2005.

[40] Hsu, J. C., *Multiple Comparisons: Theory and Methods*. London: Chapman and Hall, 1996.

[41] Kesten, H. and Morse, N., "A property of the multinomial distribution," *Annals of Mathematical Statistics*, vol. 30, no. 1, pp. 120–127, 1959.

[42] Kim, S. H. and Nelson, B. L., "A fully sequential procedure for indifference-zone selection in simulation," *ACM Transactions on Modeling and Computer Simulation*, vol. 11, no. 3, pp. 251–273, 2001.

[43] Kolojna, B., Kalasky, D. R., and Mutmansky, J. M., "Optimization of dispatching criteria for open-pit truck haulage system design using multiple comparisons with the best and common random numbers," in *1993 Winter Simulation Conference* (Evans, G. W., Mollaghasemi, M., Russell, E. C., and Biles, W. E., eds.), pp. 393–401, 1993.

[44] Levin, B., "On a sequential selection procedure of Bechhofer, Kiefer, and Sobel," *Statistics & Probability Letters*, vol. 2, no. 2, pp. 91–94, 1984.

[45] Marshall, A. W. and Olkin, I., *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.

[46] Matejcik, F. J. and Nelson, B. L., "Two-stage multiple comparisons with the best for computer simulation," *Operations Research*, vol. 43, no. 4, pp. 633–640, 1995.

[47] Miller, J. O., Nelson, B. L., and Reilly, C. H., "Efficient multinomial selection in simulation," *Naval Research Logistics*, vol. 45, no. 5, pp. 459–482, 1998.

[48] Miller, R., *Simultaneous Statistical Inference*. New York: Springer–Verlag, 2nd ed., 1981.

[49] Mosteller, F., "A k-sample slippage test for an extreme population," *Annals of Mathematical Statistics*, vol. 19, no. 1, pp. 58–65, 1948.

[50] Panchapakesan, S., "On a subset selection procedure for the most probable event in a multinomial distribution," in *Statistical Decision Theory and Related Topics* (Gupta, S. S. and Yackel, J., eds.), pp. 275–298, New York, NY: Academic Press, 1971.

[51] Paulson, E., "A multiple decision procedure for certain problems in the analysis of variance," *Annals of Mathematical Statistics*, vol. 20, no. 1, pp. 95–98, 1949.

[52] Paulson, E., "On the comparison of several experimental categories with a control," *Annals of Mathematical Statistics*, vol. 23, no. 2, pp. 239–246, 1952a.

[53] Paulson, E., "An optimum solution to the k-sample slippage problem for the normal distribution," *Annals of Mathematical Statistics*, vol. 23, no. 4, pp. 610–616, 1952b.

[54] Paulson, E., "A sequential procedure for selecting the population with largest mean from k normal populations," *Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 174–180, 1964.

[55] Ramey, J. T. and Alam, K., "A sequential procedure for selecting the most probable multinomial event," *Biometrika*, vol. 66, no. 1, pp. 171–173, 1979.

[56] Ramey, J. T. and Alam, K., "A Bayes sequential procedure for selecting the most probable multinomial event," *Communications in Statistics — Theory and Methods*, vol. A9, no. 3, pp. 265–276, 1980.

[57] Rinott, Y., "On two-stage selection procedures and related probability-inequalities," *Communications in Statistics —Theory and Methods*, vol. A7, no. 8, pp. 799–811, 1978.

[58] Ross, S., *Introduction to Probability Models*. Boston: Academic Press, 7th ed., 2000.

[59] Sobel, M. and Frankowski, K., "Extensions of gambler's ruin with even odds: Multinomial and other models," *Journal of Statistical Planning and Inference*, vol. 101, no. 1/2, pp. 291–310, 2002.

[60] Sobel, M. and Huyett, M. J., "Selecting the best one of several binomial populations," *Bell System Technical Journal*, vol. 36, pp. 537–576, 1957.

[61] Strang, G., *Linear Algebra and Its Applications*. Fort Worth, TX: Harcourt Brace and Company, 1988.

[62] Sullivan, D. W. and Wilson, J. R., "Restricted subset selection procedures for simulation," *Operations Research*, vol. 37, no. 1, pp. 52–71, 1989.

[63] Swisher, J. R. and Jacobson, S. H., "A survey of ranking, selection, and multiple comparison procedures for discrete-event simulation," in *1999 Winter Simulation Conference* (Farrington, P. A., Nembhard, H. B., Sturrock, D. T., and Evans, G. W., eds.), pp. 492–501, 1999.

[64] Vieira, H., Kienitz, K. H., and Belderrain, M. C. N., "Multinomial selection problem: A study of BEM and AVC algorithms," *Communications in Statistics — Simulation and Computation*, vol. B39, no. 5, pp. 971–980, 2010.

[65] Vieira, H. J., Sanchez, S. M., Sanchez, P., Kienitz, K. H., and Belderrain, M. C. N., "A restricted multinomial subset selection procedure." Submitted paper, 2012.

VITA

Eric S. Tollefson was born in Spokane, WA, in 1972, but raised in Lancaster, PA, for most of his childhood. In 1994, he received a B.S. in Engineering Physics from the United States Military Academy (USMA), and was commissioned as an infantry officer in the U.S. Army. After serving at Fort Bragg, NC, and Fort Benning, GA, in positions as platoon leader, assistant operations officer, and company commander, he attended the Georgia Institute of Technology in Atlanta, GA, obtaining an M.S. in Operations Research in 2002. He subsequently taught as an instructor and assistant professor in the Department of Systems Engineering at USMA, as well as serving in the Operations Research Center of Excellence there. From 2005–2009, he served as an operations research analyst supporting Army and Department of Defense studies in Monterey, CA, and as an analyst on corps staff in Iraq. In 2009, he returned to the Georgia Institute of Technology for a doctoral program in Industrial Engineering with a concentration in System Informatics and Control. His minor is Human Decision Making. He also obtained an M.S. in Statistics as part of his studies in 2011.

He married the former Kara Benson in late 2001 and has three children: Nathan, Emma, and Claire.