MASTER THESIS

by
TAMER ÖVÜTMEN

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Master of Science

Sabanci University
Spring 2007

THE MULTINOMIAL SELECTION PROBLEM

APPROVED BY

Assist. Prof. Dr. Tonguç Ünlüyurt (Thesis Supervisor)	....
Prof. Dr. David M. Goldsman (Thesis Supervisor)	....
Assist. Prof. Dr. Murat Kaya	....
Assist. Prof. Dr. Melih Papila	....
Assist. Prof. Dr. Enes Eryarsoy	....

DATE OF APPROVAL: ....

to my beloved family

Acknowledgments

I am indebted in the preperation of this thesis to my supervisor, Prof. Dave Goldsman, whose patience and kindness, as well as academic experience, have been invaluable to me. He behaved me like a brother more than a student while we were working on the thesis, as well as searching a suitable Phd.

I am extremely grateful to my co-advisor Assist. Prof. Tonguç Ünlüyurt, Assist. Prof. Murat Kaya, Assist. Prof. Melih Papila, Assist. Prof. Enes Eryarsoy for their comments, their time spent on my thesis and serving on my thesis committee.

The informal support and encouragement of many colleagues has been indispensable, and I would like particulary to acknowledge the contribution of Aydin Tanrıverdi, Ersin Demirok, and İlkan Sarıgol. I am also grateful to Fatih Kale, Sevan Harput and Kutay Özbek for helping me get through the difficult times, and for all the emotional support, entertainment, and caring they provided.

I wish to thank my parents and my sister, the constant source of support, for their understanding, endless patience, and encouragement when it was mostly required.

THE MULTINOMIAL SELECTION PROBLEM

Abstract

In this thesis, we study indifference-zone multinomial selection procedures, that is, procedures for selecting the most probable ("best") multinomial cell. Such procedures have a number of real-world applications — for instance, which is the most popular television show in a particular time slot, or which manufacturing strategy has the highest probability of yielding the largest profit on a particular trial? The indifference-zone procedures we examine all satisfy a probability requirement that guarantees to correctly select with high probability the best multinomial category under a variety of underlying probability configurations. We show by Monte Carlo and exact calculations that certain sequential sampling procedures perform better than others. We also offer various extensions and thoughts for future research.

Keywords: Multinomial selection problems, selection procedures, ranking procedures, sequential procedures, open procedures, truncated procedures.

ÇOK TERİMLİ SEÇİM PROBLEMİ

Özet

Bu tezde biz tarafsızlık-bölgesi çok terimli seçim prosedürleri üzerine alıştık. Bu prosedürler en olası ("en iyi") çok terimli hücreyi seçmeye çalışır. Bu prosedürlerin bir çok gerçek hayat uygulaması vardr: Örneğin, belli bir zaman aralığında hangi televizyon programı en çok seviliyor, ya da hangi üretim stratejisi en yüksek kar elde etme olasılığımızı en yüksek yapıyor. Tarafsızlık- bölgesi prosedürleri gerekli olasılık değerlerini sağlayarak, en yüksek olasılıkla (ve bir çok olasılık konfigürasyonunda) en iyi çok terimli kategoriyi seçmemizi garanti eder. Bu çalışmada bazı prosedürlerin diğerlerinde daha iyi olduklarını Monte Carlo simlasyonları ve tam hesaplamalarla gösterdik. Bunun yanında birçok genişletme ve ileriki araştırmalar için fikirler öne sürüldü.

Anahtar kelimeler: Çok terimli seçim problemi, seçim prosedürleri, sıralama prosedürleri, sıralı prosedürler, açık prosedürler, kesilmiş prosedürler.

Contents

List of Figures

List of Tables

1 Introduction

One of the most important problems in statistical and industrial engineering applications is that of finding the best of a number of competing systems. For example,

- Which queueing set-up offers customers the shortest expected waiting time?
- Which simulated manufacturing layout generates the greatest expected throughput?
- Which layout has the smallest variance?
- Which drug has the highest probability of giving relief?
- Which political candidate has the highest probability of winning the election?
- Which soft drink is the favorite?

In the above examples, the experimenter or decision-maker is faced with the problem of choosing among competing *stochastic* systems, and therefore faces uncertainty when making such decisions. One could resort to classical hypothesis testing, e.g., $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$, but such hypothesis tests typically determine if any of the competing systems are simply "different" from the others — they do not necessarily determine which of the competitors is actually the best. A branch of statistics, known as ranking and selection, attempts to do more. Namely, selection procedures try to find the best among the competitors with a high probability of correct selection.

Specific procedures have been developed over the last 50 years for a number of interesting scenarios. For instance, there is a large literature on the so-called normal selection problem, where we are interested in finding the best among a number of competing normal populations, e.g., which normal population has the largest mean or the smallest variance? The normal selection problem may be appropriate if we are interested in finding that one of a number of service center configurations having the smallest expected waiting time for customers. The Bernoulli selection problem also has numerous industrial and medical applications. For example, suppose that we are interested in selecting the drug that have the highest cure rate, where each patient can be regarded as a Bernoulli trial. There is also a rich literature on this Bernoulli problem. Many other general selection problems are discussed in the literature — which is the best Poisson distribution? The best exponential? The multivariate normal distribution with the largest Mahalanobis distance?

The goal of the current thesis is to study *multinomial* selection procedures, which we regard as being almost as important as the normal and Bernoulli classes. Here, we are interested in developing and evaluating selection procedures to choose the system that has the highest probability of being the "most desirable" (which corresponds to the multinomial cell having the highest probability). For example, which television show during a particular time period is the most popular? Which political candidate is most likely to win? Which manufacturing strategy has the highest probability of yielding the largest profit on a particular trial?

Our contributions in this thesis are organized as follows. In Section 2, we start with an introduction to the multinomial distribution, along with notation that will be used in the subsequent sections. Section 3 describes and motivates a compendium of multinomial selection procedures that have been popular in the literature. We also provide comparisons of the procedures in terms of various performance criteria such as the achieved probability of correct selection and the expected sample size for certain underlying probability configurations. We show how to evaluate these performance criteria via Monte Carlo simulation methods in Section 4, and via exact methods in Sections 5 and 6. The technique described in Section 5 can be applied to special cases of the procedures under discussion; it is based on the classic gambler's ruin problem and yields explicit expressions for the performance criteria of interest. The methods given in Section 6 can be used on more-general procedures and yield exact numerical results. Section 7 proposes some procedure extensions, while Section 8 gives conclusions and describes future work.

2 Notation and Set-Up

To get things going, this section discusses notation and set-up. We begin in Section 2.1 with an elementary introduction to the multiniomial distribution, followed in Section 2.2 by a general discussion on the problem of selecting the most probable multinomial category. More specifically, Section 2.3 deals with the so-called indifference-zone methodology for selecting the most probable multinomial category. Finally, Section 2.4 provides a short literature review related to relevant procedures and other issues.

2.1 The Multinomial Distribution

Our goal for now is to find the cell of a multinomial distribution that is the most probable. We will expand the problem purview later on by showing how this problem can be interpreted as that of finding that one of k competing general systems having the highest probability of yielding the "most desirable" observation.

So for the time being, we shall consider an experiment with k possible outcomes, $E_1, E_2, \ldots, E_k$, where the E_j's form a partition of the associated sample space, i.e., the E_j's are mutually exclusive and exhaustive. Let the random variables $X_{ij} = 1$ or 0 according as E_i does or does not occur on the jth trial of the experiment, for $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots$, i.e., $X_{ij} = 1$ if event i "wins" trial j, and $X_{ij} = 0$ if event i "loses" trial j. Further, let $\boldsymbol{X}_j \equiv (X_{1j}, X_{2j}, \ldots, X_{kj})$ denote the vector-observation corresponding to the outcome of the jth trial of the experiment. In addition, let $Y_{in} \equiv \sum_{j=1}^{n} X_{ij}$ be the total number of wins for event i after n observations, where we also define the vector notation $\boldsymbol{Y}_n \equiv (Y_{1n}, Y_{2n}, \ldots, Y_{kn})$.

Example 1. We are conducting a survey on the soft drink preferences of university students. Suppose we ask person j whether she likes Coke, Pepsi, or Sprite the best. If she chooses Coke, then $\boldsymbol{X}_j = (1, 0, 0)$; a choice of Pepsi yields $\boldsymbol{X}_j = (0, 1, 0)$; and Sprite gives $\boldsymbol{X}_j = (0, 0, 1)$. After we ask 150 students, we find that 73 students preferred Coke, 36 chose Pepsi, and 41 said Sprite. So we have $Y_{1,150} = 73$, $Y_{2,150} = 41$, and $Y_{3,150} = 36$, i.e., $\boldsymbol{Y}_{150} = (73, 36, 41)$. □

Assume that the outcomes of any trial are independent and identically distributed (i.i.d.), that is, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_k$ are i.i.d. Suppose that p_i denotes the probability of the event E_i occurring, $i = 1, 2, \ldots, k$, where $0 \le p_i \le 1$ and $\sum_{i=1}^k p_i = 1$. Thus, $p_i = \mathsf{Pr}(X_{ij} = 1)$, for all i, j. The quantity p_i can be interpreted as the probability that event i will "win" a particular trial. Later on, we will expand the definition of p_i so that it is the probability that, on a particular trial, system i will yield the "most desirable" observation out of those coming from k competing systems. In any case, we henceforth use the vector notation $\boldsymbol{p} \equiv (p_1, p_2, \ldots, p_k)$. We are now in a position to define the multinomial distribution, which is of fundamental interest in this thesis.

Definition 1. If $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_k$ are i.i.d., each with underlying probability vector $\boldsymbol{p}$, then we say that the vector $\boldsymbol{Y}_n$ has the *multinomial* (or *k-nomial*) distribution with parameters n and $\boldsymbol{p}$.

The probability mass function (p.m.f.) of the multinomial distribution is given by the following expression (see, for example, any standard probability and statistics text such as Hines et al. [12]).

$$\begin{aligned} p(\boldsymbol{y}) &\equiv \mathsf{Pr}(\boldsymbol{Y}_n = \boldsymbol{y}) \\ &= \mathsf{Pr}(Y_{1n} = y_1, Y_{2n} = y_2, \ldots, Y_{kn} = y_k) \\ &= \binom{\sum_{i=1}^k y_i}{y_1, y_2 \ldots, y_k} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k} \\ &= \frac{n!}{\prod_{i=1}^k (y_i!)} \prod_{i=1}^k p_i^{y_i}, \end{aligned}$$

where $\boldsymbol{y} \equiv (y_1, y_2, \ldots, y_k)$ and $\sum_{i=1}^k y_i = n$.

Example 2. Suppose we are gambling with a dice which has 12 sides. If we throw a number divisible by 4, we lose 10 YTL; if we throw a prime number, we win 10 YTL; and in all other cases, we come out even. In this case, the probability vector associated with win, draw, and lose is $\boldsymbol{p} = (1/4, 1/3, 5/12)$. Now suppose we play this game 6 times. The probability of exactly two losses, one draw, and three wins is given by

$$\mathsf{Pr}\left(\boldsymbol{Y}_6 = (2, 1, 3)\right) = \frac{6!}{2!1!3!}(1/4)^2(1/3)^1(5/12)^3 = 0.090422. \quad \square$$

2.2 Selecting the Most Probable Multinomial Category

The components of the vector $\boldsymbol{p} \equiv (p_1, p_2, \ldots, p_k)$ are generally unknown in practice. For purposes of exposition, suppose we denote the ordered probabilities as $p_{|1|} \leq p_{|2|} \leq \cdots \leq p_{|k|}$. We assume that the experimenter has no knowledge concerning the values of the p_i's or of the $p_{[j]}$'s; we also assume that the pairings of the $p_{[j]}$'s with the E_i's $(1 \leq i, j \leq k)$ are completely unknown. The category associated with $p_{[k]}$ is the "best" (most probable) category. Our goal in this research is to select the event E_i (or, later on, the system) associated with the largest probability $p_{|k|}$. If, after sampling, we do indeed choose the category associated with $p_{|k|}$, we say that we have made a *correct selection (CS)*.

Example 3. Continuing with Example 2, suppose we do not actually know the probabilities for losing, drawing, or winning, and we want to determine which outcome has the largest probability of occurrence on a single trial. The obvious selection rule that we will adopt is to choose the event that occurs the most frequently during the six trials, using randomization to break ties if they occur. Let $\boldsymbol{Y}_6 = (Y_\ell, Y_d, Y_w)$ denote the number of occurrences of (lose, draw, win) in the six trials. The probability that we correctly select the win event is given by:

$$
\begin{aligned}
&\mathsf{Pr}\,\{\text{the win event occurs the most often in the six trials}\} \\
&\quad = \mathsf{Pr}\,\{Y_w > Y_\ell \text{ and } Y_w > Y_d\} \\
&\qquad\quad + \frac{1}{2}\,\mathsf{Pr}\,\{Y_w = Y_\ell \text{ and } Y_w > Y_d\} + \frac{1}{2}\,\mathsf{Pr}\,\{Y_w = Y_d \text{ and } Y_w > Y_\ell\} \\
&\qquad\quad + \frac{1}{3}\,\mathsf{Pr}\,\{Y_w = Y_\ell = Y_d\} \\
&\quad = \mathsf{Pr}\{\boldsymbol{Y}_6 = (0,0,6), (0,1,5), (1,0,5), (0,2,4), (2,0,4), (1,1,4), (1,2,3), (2,1,3)\} \\
&\qquad\quad + \frac{1}{2}\,\mathsf{Pr}\{\boldsymbol{Y}_6 = (3,0,3)\} + \frac{1}{2}\,\mathsf{Pr}\{\boldsymbol{Y}_6 = (0,3,3)\} + \frac{1}{3}\,\mathsf{Pr}\{\boldsymbol{Y}_6 = (2,2,2)\}.
\end{aligned}
$$

Table 1 lists the outcomes favorable to a CS of the win event, along with the associated probabilities of these outcomes, incorporating randomization when ties occur.

Hence, we see that the probability of correctly selecting the win event as the most probable outcome, based on $n = 6$ trials, is 0.48828. This probability can be increased by increasing the sample size n. In fact, Figure 1 plots the exact

Table 1: Correct Selection Probabilities for Example 3

Outcome (lose,draw,win)	Contribution to Pr{CS in six trials}
(0,0,6)	0.00523
(0,1,5)	0.02512
(1,0,5)	0.01884
(0,2,4)	0.05024
(2,0,4)	0.02826
(1,1,4)	0.07535
(1,2,3)	0.12056
(2,1,3)	0.09042
(3,0,3)	(1/2)(0.02261)
(0,3,3)	(1/2)(0.05358)
(2,2,2)	(1/3)(0.10851)
	0.48828

probability of correct selection (Pr(CS)) for this example as we increase n from 1 to 100; we find that the Pr(CS) increases from about 0.4 to almost 0.85 as we do so. □

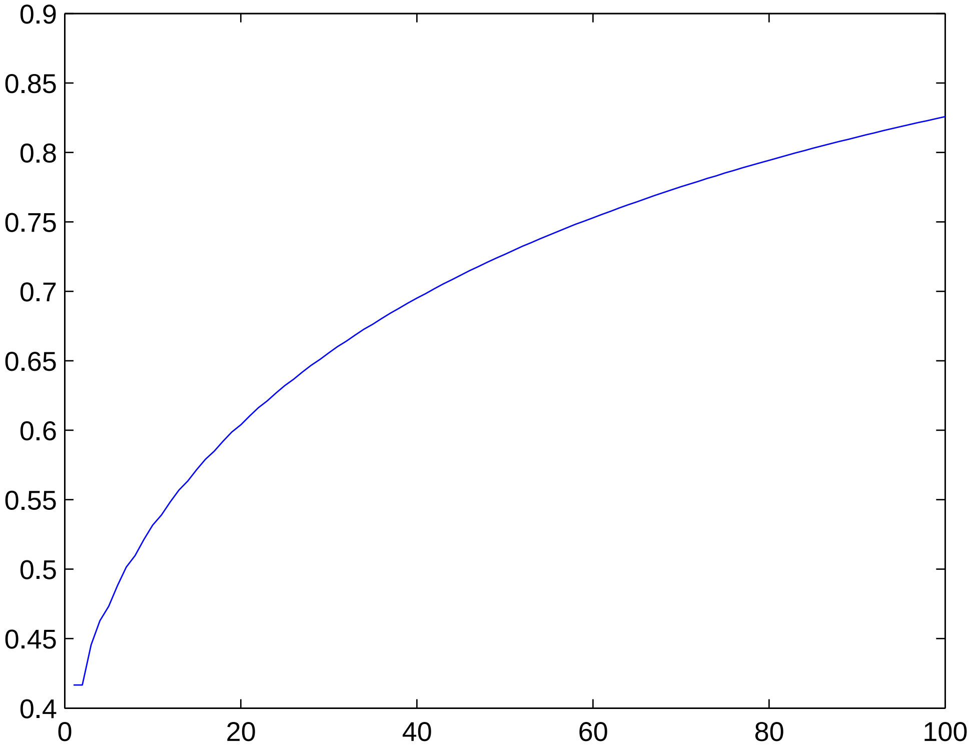

Figure 1: The Pr(CS) for Example 3 as we increase n from 1 to 100

2.3 The Indifference-Zone Approach

We will study the performance characteristics of statistical procedures that are devised to select the best category under a specified constraint on the probability of correct selection. We limit consideration to procedures that guarantee the following so-called *indifference-zone* probability requirement:

$$\text{Pr(CS)} \geq P^\star \quad \text{whenever} \quad p_{[k]} \geq \theta^\star p_{[k-1]}. \tag{1}$$

Here, $\{\theta^\star, P^\star\}$ ($\theta^\star > 1$ and $1/k < P^\star < 1$) are constants specified by the experimenter prior to the start of experimentation. The quantity $P^\star$ is obviously the experimenter's desired probability of correct selection under the indifference-zone condition $p_{[k]} \geq \theta^\star p_{[k-1]}$ where, $\theta^\star$ is the smallest ratio between the best and second-best cell probabilities. How can this indifference-zone condition be interpreted? By way of explanation, we make two fundamental definitions.

Definition 2. The *preference zone* is set of probability configurations $\Omega \equiv \{\boldsymbol{p} :$

$p_{[k]} \geq \theta^\star p_{[k-1]}\}$ for which we prefer to make a correct decision, i.e., that of selecting the category associated with $p_{[k]}$. The complement Ω^c of the preference zone is called the *indifference zone*. This is the region of $\boldsymbol{p}$-space for which we are not necessarily concerned with making a correct selection.

Some additional motivation will supply the rationale behind the indifference or preference zones.

Example 4. As a simple example, suppose that we are interested in determining which of Coke, Pepsi, and Sprite is the most popular. Clearly, the underlying probability configurations $\boldsymbol{p}_1 = (0.49, 0.48, 0.03)$ and $\boldsymbol{p}_2 = (0.50, 0.25, 0.25)$ give rise to different interpretations. One could argue that in configuration $\boldsymbol{p}_1$, Coke and Pepsi fare about the same (usually within the sampling error of most surveys); but in configuration $\boldsymbol{p}_2$, Coke obviously dominates the situation. In fact, in the case of $\boldsymbol{p}_1$, one could argue that we might be *indifferent* about declaring Coke or Pepsi to be the most popular (since they are so close); but in configuration $\boldsymbol{p}_2$, we would certainly *prefer* to correctly report that Coke is the most popular. □

Of course, in real life, we would not know the actual underlying configuration $\boldsymbol{p}$. So a good selection procedure might be designed to prefer to detect configurations such as $\boldsymbol{p}_2$ in Example 4, yet not worry about (be indifferent about detecting) configurations such as $\boldsymbol{p}_1$. Thus, in the spirit of the current discussion, the parameter $\theta^\star$ can be interpreted as the smallest ratio between the best and second-best cell probabilities, $p_{[k]}/p_{[k-1]}$, that the experimenter deems as "worth detecting." If $\theta^\star \approx 1$, then we would prefer to detect small ratios between the best and second-best cell probabilities, such as that given by configuration $\boldsymbol{p}_1$ in Example 4. On the other hand, if $\theta^\star \gg 1$, then we will only be concerned about detecting 'large" ratios. The choice of $\theta^\star$ is the responsibility of the experimenter, and may be determined by budget and other practical considerations. Further, note that specification of $\theta^\star \approx 1$ is much more demanding than specifying $\theta^\star \gg 1$, since $\theta^\star \approx 1$ requires that the procedure must be able to distinguish between category probabilities that are comparatively close to each other. Thus, if one specifies $\theta^\star \approx 1$, we would expect to take more observations, so as to guard against missing a correct selection when the two best cell probabilities are close.

Remark 1. The task of choosing the two parameters $\{\theta^\star P^\star\}$ is not onerous at all, and certainly does not mitigate against using a selection procedure instead of some kind of hypothesis test. In fact, a standard hypothesis test also requires the specification of two parameters — the level of significance α and the Type II error probability β — so the burden of specifying $\{\theta^\star, P^\star\}$ is completely reasonable. Indeed, selection procedures were originally regarded as an alternative approach to traditional hypothesis testing — instead of asking the hypothesis test question "are the cell probabilities of the multinomial distribution different?", a selection procedure asks the more-useful question "which cell probability is the largest?" Finally, whether or not one advocates one methodology over the other, there are a number of papers in the literature that combine the hypothesis testing and selection methodologies — keeping both sides happy (see, for example, the standard reference Hsu [13]). □

2.4 What to Look for in a Procedure

With n vector-observations $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ in hand, we recall the running sums $y_{in} \equiv \sum_{j=1}^{n} x_{ij}$, $i = 1, 2, \ldots, k$, $n = 1, 2, \ldots$, where the quantity y_{in} can be interpreted as the number of times category i has been sampled after n observations (stages) have been taken. We denote the ordered y_{in}-values after n observations have been taken by by $y_{[1]n} \leq y_{[2]n} \leq \cdots \leq y_{[k]n}$, $n = 1, 2, \ldots$. Typical multinomial selection procedures — described in the next section of this thesis and in the cited references — will stop sampling when the largest counter $y_{[k]n}$ is "significantly ahead" of the other y_{in}'s, or when we hit a sampling budget bound, say at $n = n_0$ observations.

In the next section, we give details on a number of indifference-zone multinomial selection procedures from the literature, including the following. Procedure $\mathcal{M}_{\text{BEM}}$ is a single-stage procedure originally discussed in Bechhofer, Elmaghraby, and Morse [2]. Bechhofer and Kulkarni [7] proposed a closed (bounded) sequential procedure $\mathcal{M}_{\text{BK}}$ that is a more-efficient implementation of procedure $\mathcal{M}_{\text{BEM}}$ in terms of the number of observations taken. Ramey and Alan [16] studied a closed sequential procedure $\mathcal{M}_{\text{RA}}$ that is usually even more parsimonious than $\mathcal{M}_{\text{BK}}$. Procedure $\mathcal{M}_{\text{BKS}}$, due to Bechhofer, Kiefer, and Sobel [6], is an open (unbounded) sequential procedure related to the classical sequential probability ratio test. Bechhofer and Goldsman [4, 5] proposed procedure $\mathcal{M}_{\rm BG}$, a truncated (bounded) version of procedure $\mathcal{M}_{\rm BKS}$, which is somewhat more efficient than the former.

How exactly would one assess the performance of a particular multinomial selection procedure, or how would we compare the performances of any of the procedures? First and foremost, any procedure must guarantee the indifference-zone probability requirement (1) — in fact, all of the procedures studied herein do so (as proven in the cited references). In addition to satisfying the probability requirement, a procedure must be frugal with observations, especially when applied to realistic configurations of the underlying unknown probability vector $\boldsymbol{p}$. Two choices of $\boldsymbol{p}$ that are of particular importance are:

1. The *slippage configuration (SC)* (often referred to as the *least-favorable configuration*),

$$p_1 \;=\; \theta^\star p,\; p_2 \;=\; p_3 \;=\; \cdots \;=\; p_k \;=\; p,$$

   where $\theta^\star > 1$, i.e.,

$$p_1 \;=\; \frac{\theta^\star}{\theta^\star + k - 1},\; p_2 \;=\; p_3 \;=\; \cdots \;=\; p_k \;=\; \frac{1}{\theta^\star + k - 1}.$$

2. The *equal-probability configuration (EP)*, $p_1 = p_2 = \cdots = p_k = 1/k$.

For all of the procedures discussed in this thesis, it can be shown (see the cited references) that $\mathsf{Pr}(\mathrm{CS}) \geq P^\star$ for $\boldsymbol{p} = \mathrm{SC}$ — which make sense since the SC is in the preference zone Ω. Furthermore, the SC can be regarded as a worst-case configuration for all $\boldsymbol{p} \in \Omega$ in that this configuration minimizes $\mathsf{Pr}(\mathrm{CS}|\boldsymbol{p})$ among all $\boldsymbol{p} \in \Omega$ (this is why the SC is also called the least favorable configuration in such cases). Not only does the SC yield the lowest $\mathsf{Pr}(\mathrm{CS})$ among all $\boldsymbol{p} \in \Omega$, it also often results in the highest expected number of observations, $\mathsf{E}(T|\boldsymbol{p})$ for $\boldsymbol{p} \in \Omega$.

When considering the EP configuration, there is no concept of a "correct selection," since all of the cells have the same probability. However, in terms of sampling requirements, the EP configuration can be regarded as a worst-case configuration for all $\boldsymbol{p}$ — not just those falling in the preference zone.

For purposes of evaluating the performance of a particular multinomial procedure (or for comparing the performances of competing multinomial procedures), the above comments suggest that we ought to report operating characteristics such as the achieved $\mathsf{Pr}(\text{CS}|\text{SC})$, $\mathsf{E}[T|\text{SC}]$, and $\mathsf{E}[T|\text{EP}]$. See, as an example, Tables 7–10 in the Appendix.

We are finally ready to discuss a number of indifferent-zone multinomial selection procedures.

3 Some Multinomial Selection Procedures

In this section, we will review several popular indifference-zone multinomial selection procedures from the literature. In each case, we will describe the procedure's setup (i.e., what needs to be specified before running the procedure), its sampling rule (i.e., how much sampling is conducted at any given stage of the procedure), its stopping rule (i.e., how we decide when to stop sampling), and its terminal decision rule (i.e., how we make our selection for the most probable cell once sampling has terminated). The terminal decision rule typically chooses as best that cell that has accumulated the most observations, using randomization in the rare case of ties.

Section 3.1 deals with the single-stage Bechhofer, Elmagrhraby, and Morse [2] procedure, while Section 3.2 concerns a more-efficient, closed, sequential version of the former, due to Bechhofer and Kulkarni [7]. Section 3.3 describes an even-more-efficient, closed, sequential procedure from Ramey and Alam [16]. Section 3.4 gives an open, sequential procedure from Bechhofer, Kiefer, and Sobel [6], while Section 3.5 discusses a closed version of the former, due to Bechhofer and Goldsman [4, 5]. Section 3.6 compares the procedures based on the criteria of $\mathsf{Pr}(\text{CS})$ and the expected value of T.

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

The first indifference-zone procedure in the literature, $\mathcal{M}_{\text{BEM}}$, was proposed by Bechhofer, Elmagrhraby, and Morse (BEM) [2]; see also the sister article, Kesten and Morse [14]. Procedure $\mathcal{M}_{\text{BEM}}$ is a single-stage procedure, that is, a procedure that takes all of its multinomial observations at the same time. The number of observations n_{BEM} is pre-determined before the experiment begins, and is chosen as the minimum number of observations that will satisfy the probability requirement (1) for the user-specified choices of $P^\star$ and $\theta^\star$.

Setup: For given k, $\theta^\star$, and $P^\star$, use Tables 7–10 to select the sample size n_{BEM}.

Sampling Rule: Take $n = n_{\text{BEM}}$ random multinomial observations $\boldsymbol{X}_j = (X_{1j}, X_{2j}, \ldots, X_{kj})$, $j = 1, 2, \ldots, n$, in single stage.

Terminal Decision Rule: For each category, calculate the sample sum

$y_{in} = \sum_{j=1}^{n} x_{ij}$, $i = 1, 2, \ldots, k$. Select the category with largest sample sum. In the case of a tie, randomize.

Example 5. Continuing our soft drink example, suppose we wish to determine which of the $k = 3$ competitors Coke, Pepsi, and Sprite is the most popular. The survey company will ask n individuals to state their preferred brand. The company will declare the favorite brand as that corresponding to largest observed proportion of positive responses. Suppose that the company wants the probability of correct selection to be at least $P^\star = 0.90$, whenever the ratio of largest to second largest true (but unknown) proportions is at least $\theta^\star = 2.0$. Referring to Table 7, we find that $n_{\text{BEM}} = 29$ individuals must be interviewed. If, after interviewing the 29 people, it turns out that Coke = 20, Pepsi = 6, and Sprite = 3, we select Coke as the most popular soda. On the other hand, if Coke = Pepsi = 13 and Sprite = 3, we flip a coin to determine the winner between Coke and Pepsi. $\square$

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

We now consider a more-efficient, sequential version of procedure $\mathcal{M}_{\text{BEM}}$. By way of motivation, we return to the previous example.

Example 6. Consider the soda survey discussed in Example 5, where we have $k = 3$ competitors, a desired $\mathsf{Pr}(\text{CS})$ of $P^\star = 0.90$, and an indifference parameter of $\theta^\star = 2.0$, so that procedure $\mathcal{M}_{\text{BEM}}$ requires that we interview $n_{\text{BEM}} = 29$ persons. But what if, after having interviewed the 25th person, the situation is that $\boldsymbol{y}_{25} = (14, 9, 2)$? This tally indicates that Coke has a substantial lead over Pepsi with only 4 observations left to be conducted — indeed, so substantial that it would not be possible for Pepsi to catch up with Coke, even were Pepsi to garner all of the remaining 4 observations. In other words, if $\boldsymbol{y}_{25} = (14, 9, 2)$, then Coke is *guanrateed* to be chosen as the favorite product in the final analysis. In such a case, we could allow procedure $\mathcal{M}_{\text{BEM}}$ to terminate sampling prematurely without affecting the procedure's ultimate selection of Coke. $\square$

With the scenario of Example 6 in mind, Bechhofer and Kulkarni (B-K) [7] devised a sequential procedure for the selecting the most probable cell that is more efficient than procedure $\mathcal{M}_{\text{BEM}}$ (which always requires a fixed sample size n_{BEM}). The B-K sequential procedure $\mathcal{M}_{\text{BK}}$ employs *curtailment* and achieves the same probability of correct selection as procedure $\mathcal{M}_{\text{BEM}}$ does, while, at the same time, potentially requiring lower number of observations over procedure $\mathcal{M}_{\text{BEM}}$. In plain English, procedure $\mathcal{M}_{\text{BK}}$ stops sampling when the category currently in the lead is guaranteed, at worst, a tie with the category currently in second place — even if all of the remaining observations were to be awarded to the category in second place. In fact, B-K show that, for any probability configuration $\boldsymbol{p}$,

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

and

$$\mathsf{E}(T \text{ using } \mathcal{M}_{\text{BK}}|\boldsymbol{p}) \;\leq\; \mathsf{E}(T \text{ using } \mathcal{M}_{\text{BEM}}|\boldsymbol{p}) \;=\; n_{\text{BEM}},$$

where T denotes the (random) number of observations taken until the point that the procedure terminates sampling.

Setup: For given k, $\theta^\star$, and P^*, use Tables 7–10 to select the (maximum possible) sample size n_{BEM}.
Sampling Rule: At the mth stage of sampling, $m = 1, 2, \ldots$, take the multinomial observation $\boldsymbol{X}_m = (X_{1m}, X_{2m} \ldots, X_{km})$.
Stopping Rule: Calculate the sample sums y_{im}, $i = 1, 2, \ldots, k$, through stage m. Stop sampling at first stage m where there exists a category i such that

$$y_{im} \;\geq\; y_{jm} + n_{\text{BEM}} - m \quad \text{for all } j \neq i.$$

Terminal Decision Rule: Let the random variable T represent the value of m at termination. If $T < n_{\text{BEM}}$, then the procedure terminated with a single category having the largest tally $y_{[k]T}$; and we select this category as the winner. If $T = n_{\text{BEM}}$, then we may have multiple categories tied for the lead; so we randomize, if necessary, to pick the winner.

Thus, curtailment of the procedure takes place when one of the categories has sufficiently more successes than all of the other categories; and even if the category in second place were to experience a "reversal of fortune" with *all* of the remaining outcomes occurring from that category, it would still be unable to defeat the current leader (at best, it could only *tie* the leader). The following examples illustrate how the procedures runs under various sampling scenarios.

Example 7. Continue under the setup of Example 5, where $k = 3$, $P^\star = 0.90$, $\theta^\star = 2.0$, and $n_{\text{BEM}} = 29$. Suppose we have the following sequence of observations.

m	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	0	1	0	12	6	3
22	0	0	1	12	6	4
23	1	0	0	13	6	4

We stop sampling at observation $T = 23$ and select category 1 as the best because

$$y_{1m} \;=\; 13 \;>\; y_{2m} + n_{\text{BEM}} - m \;=\; 6 + 29 - 23$$

and

$$y_{1m} \;=\; 13 \;>\; y_{3m} + n_{\text{BEM}} - m \;=\; 4 + 29 - 23.$$

Hence, both categories 2 and 3 have no chance to win, even if they are preferred in all of the remaining interviews. □

Example 8. This is a slight permutation of Example 7. Suppose we have the following sequence of observations.

m	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	0	1	0	12	6	3
22	0	0	1	12	7	3
23	1	0	0	13	7	3

Now, we stop sampling at observation $T = 23$ and select category 1 because

$$y_{1m} \;=\; 13 \;\geq\; y_{2m} + n_{\text{BEM}} - m \;=\; 7 + 29 - 23$$

and

$$y_{1m} \;=\; 13 \;>\; 3 + 29 - 23.$$

In this case, category 2 can at best only tie category 1, while category 3 have no chance even to tie. □

Example 9. Assume that in yet another favorite drink survey we come up with the following results.

m	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
27	0	1	0	10	11	6
28	1	0	0	11	11	6
29	0	0	1	11	11	7

At the end of the survey, since $y_{1,29} = y_{2,29} = 11$, we randomize among the two categories using probability 1/2 for each, and then select the category chosen by the random device as the winner. □

Remark 2. Note that procedure $\mathcal{M}_{\rm BK}$ employs the slightly non-intuitive termination criterion of stopping when the cell currently in first place can at worst end up in a *tie* (instead of being guaranteed a win) were sampling to continue to the maximum possible number of observations $n_{\rm BEM}$. Bechhofer and Kulkarni [7] proved that either termination criterion (stopping when at worst a tie is guaranteed or when at worst a win is guaranteed) gives precisely the same $\mathsf{Pr}(\text{CS})$ as the original single-stage procedure $\mathcal{M}_{\rm BEM}$. Since stopping when at worst a tie can be achieved is more parsimonious than stopping when a win is guaranteed, B-K adopted the former approach. □

We make some brief comments on the entries in Tables 7–10. First of all, we obtained the values of $n_{\rm BEM}$ for the procedure $\mathcal{M}_{\rm BEM}$ from Table 8.1 in Bechhofer, Santner, and Goldsman [8]. In order to generate the entries for procedure $\mathcal{M}_{\rm BK}$ in Tables 7–10, we used simulations programmed in Matlab; for each table entry, we ran 40000 independent replications of the simulation. We were also able to calculate many of the table values analytically. (More details on our exact calculations as well as the Monte Carlo implementation will be given in Section 4.) In our simulations, the required inputs are the number of competing categories (k), the indifference-zone ration of the largest to second largest probabilities ($\theta^\star$), the desired probability of correct selection ($P^\star$), and original single-stage sample size (n_{BEM}).

When we analyze the entries in Tables 7–10, we see that the attained $\mathsf{Pr}(\text{CS}|\text{SC})$ values all meet or exceed the nominal required value $P^\star$. In addition, the expected numbers of observations required in the SC, $\mathsf{E}[T|\text{SC}]$, are typically about 10% less than the single-stage procedure's corresponding truncation numbers n_{BEM}. Hence, we can conclude that the performance of procedure $\mathcal{M}_{\text{BK}}$ is superior to that of procedure $\mathcal{M}_{\text{BEM}}$. $\square$

3.3 Procedure $\mathcal{M}_{\text{RA}}$

Ramey and Alam (R-A) [16] proposed a closed, sequential procedure $\mathcal{M}_{\text{RA}}$. In this procedure, the observations are taken one-at-a-time until either the count of any category is equal to N, or the difference between the count of the leading category and that of the next largest is r. The procedure is closed since the maximum possible number of observations that can be taken is $k(N-1)+1$ — corresponding to a permutation of the sample-sum vector $\boldsymbol{y}_{k(N-1)+1} = (k, k-1, k-1, \ldots, k-1)$.

Setup: For given k, $\theta^\star$, and P^*, use Tables 11 and 12 to select the termination pair (r, N).

Sampling Rule: At the mth stage of sampling, $m = 1, 2, \ldots$, take the multinomial observation $\boldsymbol{X}_m = (X_{1m}, X_{2m} \ldots, X_{km})$.

Stopping Rule: Calculate the sample sums y_{im}, $i = 1, 2, \ldots, k$, through stage m, and then order them, $y_{[1]m} \leq y_{[2]m} \leq \cdots \leq y_{[k]m}$. Stop sampling at first stage m where there exists a category such that

$$y_{[k]m} = N$$

or

$$y_{[k]m} \; = \; y_{[k-1]m} + r.$$

Terminal Decision Rule: At the stopping point T, the category having the largest count $y_{[k]T}$ is selected as best (no randomization ever being necessary).

The values of (r, N) are dependent on k, $\theta^\star$, and $P^\star$, and are chosen in such a way as to satisfy the probability requirement (1), while at the same time minimizing $\mathsf{E}[T|\text{SC}]$. The determination of the optimal (r, N) values is typically carried out by what amounts to a complete enumeration of a reasonable set of possible (r, N) values. Ramey and Alam provide (r, N) tables for a variety of choices of $k, \theta^\star, P^\star$; but see Bechhofer and Goldsman [3] for some corrections to their tables. Our Tables 11 and 12 extend the range of applicable table values over those given in [3]. For more details on how we actually carry out the calculations, see the Monte Carlo and exact methodologies outlined in Sections 4–6 of this thesis, which can be used to determine appropriate (r, N) values.

Example 10. Consider the soda survey discussed in Example 5, where we have $k = 3$ competitors, a desired $\mathsf{Pr}(\text{CS})$ of $P^\star = 0.90$, and an indifference parameter of $\theta^\star = 2.0$, so that, by Table 11, procedure $\mathcal{M}_{\text{RA}}$ terminates at the pair $(r, N) = (4, 15)$. Thus, the procedure terminates as soon as one of the products receives 15 votes, or as soon as one of the products receives 4 more votes than the other two competitors.

m	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	1	0	0	2	8	5
16	1	0	0	3	8	5
17	0	1	0	3	9	5

Since

$$y_{2,17} \;\geq\; y_{1,17} + r \;=\; y_{1,17} + 4 \quad \text{and} \quad y_{2,17} \;\geq\; y_{3,17} + r \;=\; y_{3,17} + 4,$$

we stop at observation $T = 17$ and select cell 2 as the most probable. $\square$

Example 11. Under the same set-up as Example 10, with $(r, N) = (4, 15)$, suppose we have the following sequence of observations.

m	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
32	1	0	0	13	14	8
33	0	0	1	13	14	9
34	0	1	0	13	15	9

Since $y_{2,34} = N = 15$, we stop at observation $T = 34$ and select cell 2 as the most probable. $\square$

3.4 Procedure $\mathcal{M}_{\text{BKS}}$

Bechhofer, Kiefer, and Sobel $\mathcal{M}_{\text{BKS}}$[6] proposed an open, sequential sampling procedure $\mathcal{M}_{\text{BKS}}$ for selecting the multinomial category having the highest cell probability. Their procedure is related to a classical Wald-style sequential probability ratio test [18]. Since the procedure is open, it can continue sampling for an arbitrarily long time. In fact, the stopping rule depends only on the differences between the total numbers of wins (and not on a pre-determined truncation number). The procedure runs as follows.

Setup: Determine the k, $\theta^\star$, and $P^\star$ values.

Sampling Rule: At the mth stage of sampling, $m = 1, 2, \ldots$, take the multinomial observation $\boldsymbol{X}_m = (X_{1m}, \ldots, X_{km})$.

Stopping Rule: Calculate the sample sums y_{im}, $i = 1, 2, \ldots, k$, through stage m, and then order them, $y_{[1]m} \le y_{[2]m} \le \cdots \le y_{[k]m}$. Also calculate

$$z_m \equiv \sum_{i=1}^{k-1} (1/\theta^\star)^{y_{[k]m} - y_{[i]m}}.$$

Stop sampling at first stage m where there exists a category such that

$$z_m \le \frac{1 - P^\star}{P^\star}.$$

Terminal Decision Rule: At the stopping point T, select the event associated with $y_{[k]T}$. (Ties will not be possible under the stopping rule in play here.)

Example 12. Let us return to our soft drink example, for which we had $k = 3$ competitors, a desired Pr(CS) of $P^\star = 0.90$, and an indifference parameter of $\theta^\star = 2.0$. Consider the data realization

m	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	1	0	0	8	2	5
16	0	1	0	8	3	5
17	1	0	0	9	3	5

We stop sampling at stage $T = 17$ and select category 1 as the most probable since

$$z_{17} \;=\; (1/2)^6 + (1/2)^4 \;=\; 5/64 \;\le\; (1 - P^\star)/P^\star \;=\; 1/9. \quad \Box$$

3.5 Procedure $\mathcal{M}_{\text{BG}}$

While studying procedure $\mathcal{M}_{\text{BKS}}$, Bechhofer and Goldsman (B-G) [4, 5] found that the Pr(CS) achieved in the least favorable configuration always exceeded the probability requirement (1)'s lower bound of $P^\star$, sometimes by a substantial amount. In an effort to reduce the expected sample size, while still adhering to the probability requirement, B-G incorporate a truncation point (i.e., a limit on the total number of observations that can be taken) in their procedure $\mathcal{M}_{\text{BG}}$. The truncation point n_{BG} is chosen as the minimum limit such that the probability requirement is satisfied; thus, procedure $\mathcal{M}_{\text{BG}}$ trades some of the wasteful, extra Pr(CS) from procedure $\mathcal{M}_{\text{BKS}}$ for a reduction in the value of $\mathsf{E}[T]$.

Setup: For given k, $\theta^\star$, and $P^\star$, find the truncation number n_{BG} from Tables 13 and 14.

Sampling Rule: At the mth stage of sampling, $m = 1, 2, \ldots$, take the multinomial observation $\boldsymbol{X}_m = (X_{1m}, \ldots, X_{km})$.

Stopping Rule: Calculate the sample sums y_{im}, $i = 1, 2, \ldots, k$, through stage m, and then order them, $y_{[1]m} \le y_{[2]m} \le \cdots \le y_{[k]m}$. Also calculate

$$z_m \;\equiv\; \sum_{i=1}^{k-1} (1/\theta^\star)^{y_{[k]m} - y_{[i]m}}.$$

Stop sampling at first stage m where there exists a category such that *either*

$$z_m \leq \frac{1-P^\star}{P^\star} \tag{2}$$

or

$$m = n_{\rm BG} \tag{3}$$

or

$$y_{[k]m} \geq y_{[k-1]m} + n_{\rm BG} - m. \tag{4}$$

Terminal Decision Rule: At the stopping point T, select the event associated with $y_{[k]T}$. Break ties with randomization. (Ties will not be possible if we happen to stop at time $T < n_{\rm BG}$.)

The first stopping criterion (2) is the stopping rule from the open procedure $\mathcal{M}_{\rm BKS}$; the second criterion (3) is simply the truncation rule; and the third (4) is a curtailment rule in the spirit of procedure $\mathcal{M}_{\rm BK}$. Note that (3) is redundant in light of (4), but we retain it for ease of exposition. Some examples will serve to illustrate this procedure's multiple stopping criteria.

Example 13. Going back to our soft drink example with $k = 3$ three competitors, $P^\star = 0.9$, and $\theta^\star = 2.0$, Table 13 shows that we can use a truncation number of a survey for $n_{\rm BG} = 34$ people. Consider the data

m	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	1	0	0	8	2	5
16	0	1	0	8	3	5
17	1	0	0	9	3	5

As in Example 12, we stop sampling by the first criterion (2) and select category 1. □

Example 14. Under the same setup as in Example 13, with truncation number $n_{\rm BG} = 34$, consider the following sequence of observations.

m	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
32	0	0	1	11	9	12
33	1	0	0	12	9	12
34	1	0	0	13	9	12

We stop sampling by the second criterion (3) and select category 1 because $m = n_{\text{BG}} = 34$ observations have been taken. □

Example 15. Yet again under the setup of Example 13, with $n_{\text{BG}} = 34$, consider the following sequence.

m	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
32	1	0	0	11	12	9
33	1	0	0	12	12	9
34	0	0	1	12	12	10

We stop sampling by the second criterion (3) because $m = n_{\text{BG}} = 34$ observations have been taken. Since we have a tie between $y_{1,34}$ and $y_{2,34}$, we randomly select between categories 1 and 2. □

Example 16. Consider one last visit to the soft drink survey of Example 13, still using $n_{\text{BG}} = 34$. Suppose we observe

m	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
28	0	1	0	11	8	9
29	1	0	0	12	8	9
30	1	0	0	13	8	9

As categories 2 and 3 can do no better than tie category 1 in the $n_{\text{BG}} - m = 34 - 30 = 4$ potential remaining observations, we stop by the third criterion (4), and we select category 1. □

3.6 Comparison of Procedures

All of the procedures that we have looked at in this section are designed to satisfy the probability requirement (1). Generally speaking, the sequential procedures $\mathcal{M}_{\rm BK}$, $\mathcal{M}_{\rm RA}$, $\mathcal{M}_{\rm BKS}$, and $\mathcal{M}_{\rm BG}$ tend to be more parsimonious with observations than the single-stage procedure $\mathcal{M}_{\rm BEM}$. In fact, we have already seen that, for any $\boldsymbol{p}$-configuration, procedure $\mathcal{M}_{\rm BK}$ achieves the same Pr(CS) as procedure $\mathcal{M}_{\rm BEM}$, yet is also more efficient in terms of $\mathsf{E}[T]$ than is procedure $\mathcal{M}_{\rm BEM}$. Further, although both procedures $\mathcal{M}_{\rm BKS}$ and $\mathcal{M}_{\rm BG}$ satisfy (1), procedure $\mathcal{M}_{\rm BG}$ is — by definition — the more efficient of the two. So with procedures $\mathcal{M}_{\rm BEM}$ and $\mathcal{M}_{\rm BKS}$ out of the way, we shall only compare the sampling efficiency of procedures $\mathcal{M}_{\rm BK}$, $\mathcal{M}_{\rm RA}$, and $\mathcal{M}_{\rm BG}$ in the sequel.

In addition, a comparison of Tables 7 and 8 (for procedure $\mathcal{M}_{\rm BK}$), Tables 11 and 12 (for procedure $\mathcal{M}_{\rm RA}$), and Tables 13 and 14 (for procedure $\mathcal{M}_{\rm BG}$) shows that procedure $\mathcal{M}_{\rm BK}$ only rarely defeats procedures $\mathcal{M}_{\rm RA}$ and $\mathcal{M}_{\rm BG}$ in terms of $\mathsf{E}[T]$ — and then only for the occasional $\boldsymbol{p}$ = EP entry. So, for all intents and purposes, we only need to continue with our consideration of procedures $\mathcal{M}_{\rm RA}$ and $\mathcal{M}_{\rm BG}$.

When we compare the performances of procedures $\mathcal{M}_{\rm RA}$ and $\mathcal{M}_{\rm BG}$, we see that there is no uniform dominance of one of the procedures over the other — for some choices of $k, P^\star, \theta^\star$ and $\boldsymbol{p}$, procedure $\mathcal{M}_{\rm RA}$ gives smaller $\mathsf{E}[T|\boldsymbol{p}]$ values than does $\mathcal{M}_{\rm RA}$; in some cases, vice versa.

When we look in our simulation results we can compare $\mathcal{M}_{\rm RA}$and $\mathcal{M}_{\rm BG}$in 120 different k, $P^\star$, $\theta^\star$ combinations: (for $k = 2, 3, 4, 5$; $P^\star = 0.75, 0.90, 0.95$; and $\theta^\star = 1.2, 1.4, \ldots, 3.0$), $\mathcal{M}_{\rm BG}$performs better in 76 cases, $\mathcal{M}_{\rm RA}$performs better in 27 cases and they have the same performance within ± 0.01 values in 17 cases. The performances of the two procedures have the closest values for $k = 2$, where we have all the 17 ties. As k increases we see that $\mathcal{M}_{\rm BG}$performs better than $\mathcal{M}_{\rm RA}$, for $k = 3, 4, 5$ the $\mathcal{M}_{\rm BG}$procedure performs better in 18, 22, and 25 ($P^\star$, $\theta^\star$ combinations) cases respectively. As $\theta^\star$ increases, for the same k and $P^\star$ combination, we observe the $\mathcal{M}_{\rm BG}$procedure performs better than the $\mathcal{M}_{\rm RA}$procedure (the $\mathcal{M}_{\rm BG}$performs better for all combinations of $\theta^\star = 1.2$ and $\theta^\star = 1.4$, except for $k = 3$, $P^\star = 0.75$ combination).

4 Monte Carlo Estimation of Performance Criteria

To generate the tables at the Appendix and, for testing the procedures we have described in the previous section, we have used Monte Carlo simulations. To obtain Monte Carlo simulation results, we have used Matlab. In this section we will briefly explain how the simulation results were obtained. In all simulations the inputs are: number of competing systems (k), the ratio of largest to second larger proportions ($\theta^\star$), desired probability of correct selection($P^\star$), and truncation number (n_0).

In the initialization part, we define the required intervals for each category, for being able to match generated random variable with the corresponding category. For example, lets say we have three categories ($k = 3$) and ($\theta^\star = 2$). The required probability interval for each category are, (0,0.5], for category one, (0.5,0.75] for category two, and (0.75,1.0] for category three. The initialization procedure is the same for all procedures.

After the initialization, we began "sampling"; we generate standard uniform random number and look for the corresponding category for that number. For corresponding category i we increase the y_i value by 1. We continue this procedure till one of the stopping criteria is achieved. At the termination part we determine which category is the winner of that sampling procedure, and how many samples we had before the process terminates. In our simulations we have done these procedure 40000 times. After each replication we store which category is the winner, and how many sample we had. We count how many times our desired category won (say W_i). The ratio of $\frac{W_i}{40000}$ is the probability of correct selection value of the simulation. We also take the average of sampling numbers at the termination, to obtain expected number of observations $E(n)$.

5 Exact Results via Random Walk Methods

For the $\mathcal{M}_{\text{BKS}}$ procedure (untruncated version of $\mathcal{M}_{\text{BG}}$) which is an open sequential procedure, we can calculate the performance characteristics by using random walk arguments. The procedure is said to be open since it is not possible, before the experiment starts, to state an upper bound on the number of observations required to terminate sampling. In this procedure with $k = 2$ the observations are taken one at a time until

$$(1/\theta^\star)^{y_{[2]m}-y_{[1]m}} \;\leq\; \left(\frac{1-P^\star}{P^\star}\right)$$

is equivalent to

$$|y_{1m} - y_{2m}| \;\geq\; \ell\text{n}\left(\frac{P^\star}{1-P^\star}\right)/\ell\text{n}(\theta^\star).$$

Hence we are only interested in the difference between the total number of wins for system i after m observations. Let

$$R \;=\; \left\lceil \ell\text{n}\left(\frac{P^\star}{1-P^\star}\right)/\ell\text{n}(\theta^\star)\right\rceil,$$

where $\lceil\cdot\rceil$ is the "ceiling" (or round-up) function, so that we can model the procedure as a Gambler's Ruin problem, at which the gambler starts at R and the game ends when he hits 0 or $2R$. Hence, it is a Markov chain with transition probabilities

$P_{0,0} = P_{2R,2R} = 1$

$P_{i,i+1} = p_1 = 1 - P_{i,i-1}, \quad i = 0, 1, \ldots, 2R-1$

The Pr(CS)can also be defined as the probability of starting from i, the gambler's fortune will eventually reach $2R$ (P_i). By conditioning on the initial selection we obtain

$P_i = p_1 P_{i+1} + p_2 P_{i-1}, \qquad i = 1, 2, \ldots, 2R-1$

since $p_1 + p_2 = 1$,

$p_1 P_i + p_2 P_i = p_1 P_{i+1} + p_2 P_{i-1}$

or,

$P_{i+1} - P_i = \frac{p_2}{p_1}(P_i - P_{i-1}), \quad i = 1, 2, \ldots, 2R-1$

by using $P_0 = 0$ we obtain that,

$$P_i = \begin{cases} \frac{1-(p_2/p_1)^i}{1-(p_2/p_1)^{2R}} & \text{if } p_1 \neq p_2 \\ \frac{i}{2R} & \text{if } p_1 = p_2 \end{cases} .$$

For $p_1 > p_2$ (so that category 1 is the better),

$$\mathsf{Pr}(\text{CS}) = \frac{1-(p_2/p_1)^R}{1-(p_2/p_1)^{2R}} = [(p_1/p_2)^R + 1]^{-1}.$$

Thus, for instance, if $P_1 = 0.6$ and $P_2 = 0.4$ then the probability of correct selection is 0.9997 when $2R = 10$ In this case the expected value for the number of observations required can be calculated by:

$$E[N] = \begin{cases} R^2 & \text{if } p_1 = p_2 \\ \frac{R}{p_1-p_2} - \frac{2R}{p_1-p_2} \cdot \frac{1}{1+(p_1/p_2)^R} & \text{if } p_1 \neq p_2 \end{cases} .$$

We can also show that the probability of stopping at observation n is:

$$\left[p_1^{\frac{n-d}{2}} p_2^{\frac{n+d}{2}} + p_1^{\frac{n+d}{2}} p_2^{\frac{n-d}{2}}\right] \times \frac{2^{n-1}}{d} \sum_{k=1}^{2d-1} \cos^{n-1}(\pi k/2d) \sin(\pi k/2d) \sin(\pi k/2).$$

For different θ values we can plot the the probability of stopping at observation n. Figure 2 gives such plots for θ values 3, 2, 1.6, and 1.2.

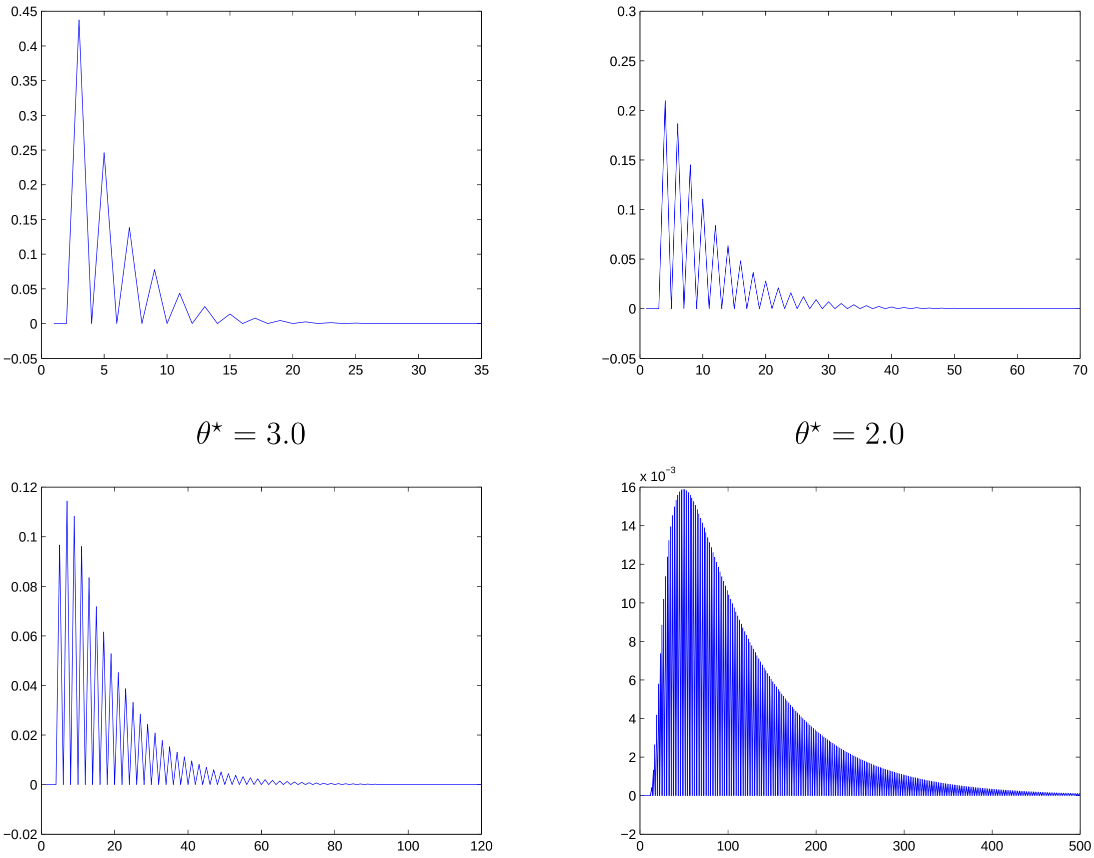

Figure 2: The Probability of Stopping at n

6 Exact Calculation Methodology

In this section, we will formulate an iterative method for calculating the performance characteristics of various sequential procedures for selecting the most probable multinomial cell. Such performance characteristics include the exact probability of obtaining a correct selection and the exact expected number of observations before procedure termination, all under arbitrary configurations of the underlying probabilities.

We now describe a methodology to calculate exactly various performance characteristics of generic multinomial selection procedures — namely, the exact probability of obtaining a correct selection and the exact expected number of observations before procedure termination, all under arbitrary configurations of the underlying probabilities.

Let us denote by T the number of vector-observations that a particular procedure $\mathcal{P}$ requires before termination. The quantity T may be fixed, as in the Bechhofer, Elmaghraby, and Morse [2] (BEM) procedure, or — more likely — a random variable, as in most other procedures of interest. We will give algorithms to calculate the exact values of $\mathsf{Pr}(\mathrm{CS}|\boldsymbol{p})$ and $\mathsf{E}(T|\boldsymbol{p})$ for a variety of procedures $\mathcal{P}$ under arbitrary underlying probability vectors $\boldsymbol{p}$.

To start things off, consider the running counts $\boldsymbol{y} = (y_{1m}, y_{2m}, \ldots, y_{km})$ after m stages of sampling. We define the notation $\#(\ell_1, \ell_2, \ldots, \ell_k)$ to be the number of distinct paths of the sampling process $\{\boldsymbol{y}_m : m = 1, 2, \ldots\}$ that lead to procedure termination exactly when $\boldsymbol{y}_m = \boldsymbol{\ell}$, where $\boldsymbol{\ell} \equiv (\ell_1, \ell_2, \ldots, \ell_k)$.

Example 17. Consider the BEM procedure with k categories, and suppose that we are directed to take $T = n$ vector observations. Then it is obvious that $\#\boldsymbol{\ell} = \binom{n}{\ell_1, \ell_2 \ldots, \ell_k}$, the usual multinomial coefficient. $\triangleleft$

Similar calculations for sequential procedures take a little more thought, though we begin with a trivial example.

Example 18. Consider the Ramey and Alam [16] (R-A) procedure with $k = 2$, $r = 2$, and $N = 3$, so that the procedure terminates sampling when either $y_{[k]m} - y_{[k-1]m} = r$ or $y_{[k]m} = N$. Then $\#(2, 0) = 1$ since only one path of the sampling process leads to termination exactly when $\boldsymbol{y}_2 = (2,0)$, namely, the path $\boldsymbol{y}_1 = (1,0) \to \boldsymbol{y}_2 = (2,0)$. $\triangleleft$

It is obvious that the number of paths such that the procedure terminates at $\boldsymbol{\ell}$ is equal to the total number of potential paths to $\boldsymbol{\ell}$ minus the number of paths to $\boldsymbol{\ell}$ that terminate en route. In other words,

$$\#\boldsymbol{\ell} = \binom{\sum_{i=1}^k \ell_i}{\ell_1, \ell_2 \ldots, \ell_k} - [\text{ number of paths to } \boldsymbol{\ell} \text{ that terminate en route }].$$

Example 19. Suppose we apply the R-A procedure to the case in which $k = 2$, $r = 2$, and $N = 3$. Further suppose that the want to calculate $\#(3,1)$. Noting that the R-A procedure terminates (en route to $\boldsymbol{y}_4 = (3,1)$) if $\boldsymbol{y}_2 = (2,0)$, we have

$$\#(3,1) = \binom{4}{1} - [\text{ number of paths from (2,0) to (3,1) }]\#(2,0) = 4 - \binom{2}{1} = 2. \quad \triangleleft$$

Generalizing the above example by giving an explicit expression for the number of ways to terminate enroute, it is easy to see that

$$\#\boldsymbol{\ell} = \binom{\sum_{i=1}^k \ell_i}{\ell_1, \ell_2 \ldots, \ell_k} - \sum_{j_1=0}^{\ell_1} \sum_{j_2=0}^{\ell_2} \cdots \sum_{j_k=0}^{\ell_k} \binom{\sum_{i=1}^k (\ell_i - j_i)}{\ell_1 - j_1, \ell_2 - j_2, \ldots, \ell_k - j_k} \#(j_1, j_2, \ldots, j_k). \tag{5}$$

Remark 3. By symmetry,

$$\#\boldsymbol{\ell} \;=\; \#(\ell_1, \ell_2, \ldots, \ell_k) \;=\; \#(\text{any permutation of } \ell_1, \ell_2, \ldots, \ell_k).$$

Hence, we need only explicitly calculate values of $\#\boldsymbol{\ell} = \#(j_1, j_2, \ldots, j_k)$ such that $j_1 \geq j_2 \geq \cdots \geq j_k$, since all other values will follow by symmetry. $\triangleleft$

Definition 3. The only nonzero $\#\boldsymbol{\ell}$'s are those for which the procedure in question terminates. In fact, for a given procedure, we introduce the *termination set* (or *stopping set*) $\mathcal{T} \equiv \{\boldsymbol{\ell} : \text{the procedure terminates}\} = \{\boldsymbol{\ell} : \#\boldsymbol{\ell} > 0\}$.

We are now in a position to present a more-interesting example.

Example 20. Consider the R-A procedure using some choice of termination parameters (r, N). In this case, we need only calculate the $\#\boldsymbol{\ell}$'s for the following configurations of $\boldsymbol{\ell}$.

$$\#(j_2 + r, j_2, j_3, \ldots, j_k), \quad 0 \le j_2 \le N - r - 1, \; j_2 \ge j_3 \ge \cdots \ge j_k \tag{6}$$

and

$$\#(N, j_2, j_3, \ldots, j_k), \quad N - r \le j_2 \le N - 1, \; j_2 \ge j_3 \ge \cdots \ge j_k. \tag{7}$$

Any $\#\boldsymbol{\ell}$ that is not a permutation of (6) or (7) must equal 0, because it is impossible for the R-A procedure to terminate at such $\#\boldsymbol{\ell}$ values. $\triangleleft$

Remark 4. It will facilitate matters if we calculate the $\#\boldsymbol{\ell}$'s in the following iterative manner.

1. Initialize all $\#\boldsymbol{\ell}$'s equal to zero.

2. Using Equation (5), calculate the next (*left-lexicographically*) $\#\boldsymbol{\ell}$ corresponding to a termination configuration. By the above Remarks, we obtain at this step (with no further calculations) all of the $\#\boldsymbol{\ell}$'s that are permutations of the current case.

3. If there are no other configurations left to check, stop. Otherwise, go to Step 2. $\triangleleft$

Remark 5. The left-lexicographic order of calculation is necessary since the computation of $\#\boldsymbol{\ell}$ involves all of the previous $\#\boldsymbol{\ell}$'s (as well as their permutations). If we store all of the values of these previous $\#\boldsymbol{\ell}$'s as they are calculated, we avoid recursive re-computation in Equation (5).

Example 21. Consider the Bechhofer and Kulkarni [7] (B-K) curtailed procedure with $k = 3$ and upper bound $B = 5$. Recall that B-K samples up to B vector-observations, but terminates if the category currently in first place can do no worse than tie. Then the algorithm proceeds as follows.

1. Initialize all $\#(\ell_1, \ell_2, \ell_3)$'s to 0.

2. Using Equation (5), set $\#(2,1,1) = \binom{4}{2,1,1} - 0 = 12$. Note that symmetry implies that $\#(1,2,1) = \#(1,1,2) = 12$.

2. Again using Equation (5), set $\#(2,2,1) = \binom{5}{2,2,1} - 1 \cdot \#(2,1,1) - 1 \cdot \#(1,2,1) = 30 - 12 - 12 = 6$. By symmetry, we have $\#(2,1,2) = \#(1,2,2) = 6$.

2. By (5), set $\#(3,0,0) = 1$. Thus, $\#(0,3,0) = \#(0,0,3) = 1$.

2. By (5), set $\#(3,1,0) = \binom{4}{3,1,0} - 1 \cdot \#(3,0,0) = 3$. Thus, $\#(0,1,3) = \#(0,3,1) = \#(1,0,3) = \#(1,3,0) = \#(3,0,1) = 3$.

2. By (5), set $\#(3,2,0) = \binom{5}{3,2,0} - 1 \cdot \#(3,0,0) - 1 \cdot \#(3,1,0) = 6$. Thus, $\#(0,2,3) = \#(0,3,2) = \#(2,0,3) = \#(2,3,0) = \#(3,0,2) = 6$.

3. End, since there are no more ways to stop. $\triangleleft$

The only (small) difficulty lies in determining which $\boldsymbol{\ell}$-configurations correspond to stopping configurations $\boldsymbol{\ell} \in \mathcal{T}$. A more-substantive example may help to explain.

Example 22. Consider the R-A procedure with arbitrary k, r, N. All terminating configurations $\boldsymbol{\ell} \in \mathcal{T}$ are of (or are permutations of) the following forms.

$$(j_2 + r, j_2, j_3, \dots, j_k), \quad 0 \le j_2 \le N - r - 1,\ j_2 \ge j_3 \ge \cdots \ge j_k \tag{8}$$

and

$$(N, j_2, j_3, \dots, j_k), \quad N - r \le j_2 \le N - 1,\ j_2 \ge j_3 \ge \cdots \ge j_k. \tag{9}$$

Thus, in the case of R-A, we would need to calculate the following quantities (as well as all of their permutations with no additional effort).

$$\left.\begin{array}{c}
\#(j+r, j, 0, \dots, 0) \\
\#(j+r, j, 1, 0, \dots, 0) \\
\#(j+r, j, 1, 1, \dots, 0) \\
\vdots \\
\#(j+r, j, 1, 1, \dots, 1) \\
\#(j+r, j, 2, \dots, 0) \\
\vdots \\
\#(j+r+1, j+1, 0, \dots, 0) \\
\vdots \\
\#(N-1, N-r-1, \dots, N-r-1)
\end{array}\right\} \#\boldsymbol{\ell}\text{'s of the form in (8)}$$

$$\left.\begin{array}{c} \#(N, N-r, 0, \ldots, 0) \\ \#(N, N-r, 1, 0, \ldots, 0) \\ \vdots \\ \#(N, N-1, N-1, \ldots, N-1) \end{array}\right\} \quad \#\boldsymbol{\ell}\text{'s of the form in (9)}$$

$\triangleleft$

Example 23. As an example within Example 22, consider the case $k = 3$, $r = 2$, $N = 4$.

$$\left.\begin{array}{rcl} \#(2{,}0{,}0) & = & 1 \\ \#(3{,}1{,}0) & = & 2 \\ \#(3{,}1{,}1) & = & 10 \end{array}\right\} \quad \#\boldsymbol{\ell}\text{'s of the form in (8)}$$

$$\left.\begin{array}{rcl} \#(4{,}2{,}0) & = & 4 \\ \#(4{,}2{,}1) & = & 28 \\ \#(4{,}2{,}2) & = & 123 \\ \#(4{,}3{,}0) & = & 8 \\ \#(4{,}3{,}1) & = & 64 \\ \#(4{,}3{,}2) & = & 320 \\ \#(4{,}3{,}3) & = & 960 \end{array}\right\} \quad \#\boldsymbol{\ell}\text{'s of the form in (9)}$$

$\triangleleft$

We are now in the position to calculate the probability that a correct selection takes place. We will assume, without loss of generality, that the most-probable category is category 1. Therefore, a CS takes place if, at sample termination time T,

1. Category 1 has the more wins than any other category (i.e., $y_{1T} = y_{[k]T} > y_{[k-1]T}$), or

2. If category 1 is tied with other categories for the most wins, we randomize among these contenders and happen to select category 1 (for example, if $y_{1T} = y_{[k]T} = y_{[k-1]T} = \cdots = y_{[k-s+1]T} > y_{[k-s]T}$, then category 1 is selected with probability $1/s$).

Henceforth, let $r(\boldsymbol{\ell})$ denote the randomization constant associated with the Pr(CS) if we were to stop sampling at state $\boldsymbol{\ell}$. In other words, if $\ell_{[1]} \leq \ell_{[2]} \leq \cdots \leq \ell_{[k]}$

denote the ordered ℓ_i's, and if we again assume without loss of generality that category 1 is the most probable, then

$$r(\boldsymbol{\ell}) \equiv \begin{cases} 0 & \text{if } \ell_1 < \ell_{[k]} \\ 1/s & \text{if } \ell_1 = \ell_{[k]} = \ell_{[k-1]} = \cdots = \ell_{[k-s+1]} > \ell_{[k-s]} \end{cases}.$$

In addition, if the true vector of underlying probabilities is $\boldsymbol{p}$, then for any procedure, we have

$$\mathsf{Pr}(\mathrm{CS}|\boldsymbol{p}) \;=\; \sum_{\ell_1=0}^{\infty}\sum_{\ell_2=0}^{\infty}\cdots\sum_{\ell_k=0}^{\infty} \#\boldsymbol{\ell}\, r(\boldsymbol{\ell}) \prod_{i=1}^{k} p_i^{\ell_i} \;=\; \sum_{\boldsymbol{\ell}\in\mathcal{T}} \#\boldsymbol{\ell}\, r(\boldsymbol{\ell}) \prod_{i=1}^{k} p_i^{\ell_i}, \tag{10}$$

where the term $\#\boldsymbol{\ell}\, r(\boldsymbol{\ell}) \prod_{i=1}^{k} p_i^{\ell_i}$ is simply the probability that the procedure will terminate at configuration $\boldsymbol{\ell}$, scaled by the randomization constant $r(\boldsymbol{\ell})$. Further, the expected number of vector-observations until procedure termination is

$$\mathsf{E}(T|\boldsymbol{p}) \;=\; \sum_{\ell_1=0}^{\infty}\sum_{\ell_2=0}^{\infty}\cdots\sum_{\ell_k=0}^{\infty} \#\boldsymbol{\ell}\; s(\boldsymbol{\ell}) \prod_{i=1}^{k} p_i^{\ell_i} \;=\; \sum_{\boldsymbol{\ell}\in\mathcal{T}} \#\boldsymbol{\ell}\; s(\boldsymbol{\ell}) \prod_{i=1}^{k} p_i^{\ell_i}, \tag{11}$$

where $s(\boldsymbol{\ell}) \equiv \sum_{i=1}^{k} \ell_i$ and where the term $\#\boldsymbol{\ell}\, s(\boldsymbol{\ell}) \prod_{i=1}^{k} p_i^{\ell_i}$ represents each $\boldsymbol{\ell}$'s contribution to the expected value. Note that in the cases of both Equations (10) and (11), we only need to sum over those values of $\boldsymbol{\ell} \in \mathcal{T}$ since $\#\boldsymbol{\ell} = 0$ for all $\boldsymbol{\ell} \notin \mathcal{T}$.

Example 24. Consider the B-K curtailed procedure with $k = 3$ and upper bound $B = 5$. The first three columns of Table 2 give all of the possible termination points $\boldsymbol{\ell}$ (i.e., all $\boldsymbol{\ell} \in \mathcal{T}$), along with their associated $\#\boldsymbol{\ell}$ and $r(\boldsymbol{\ell})$ values. Column 4 of Table 2 gives each $\boldsymbol{\ell}$'s contribution $\#\boldsymbol{\ell}\, r(\boldsymbol{\ell}) \prod_{i=1}^{k} p_i^{\ell_i}$ to the overall $\mathsf{Pr}(\mathrm{CS}|\boldsymbol{p})$ for any underlying probability configuration $\boldsymbol{p}$. Column 5 of the table gives each $\boldsymbol{\ell}$'s contribution $\#\boldsymbol{\ell}\, s(\boldsymbol{\ell}) \prod_{i=1}^{k} p_i^{\ell_i}$ to the overall $\mathsf{E}(T|\boldsymbol{p})$ for any underlying probability configuration $\boldsymbol{p}$.

At this point, we can study the behavior of a procedure for certain probability configurations $\boldsymbol{p}$ of interest.

In the case of the SC, Equations (10) and (11) simplify to

$$\mathsf{Pr}(\mathrm{CS}|\boldsymbol{p} = \mathrm{SC}) \;=\; \sum_{\boldsymbol{\ell}\in\mathcal{T}} \#\boldsymbol{\ell}\; r(\boldsymbol{\ell}) (\theta^{\star})^{\ell_1} p^{s(\ell)} \tag{12}$$

Table 2: Performance characteristics for Example 24

$\boldsymbol{\ell}$	$\#\boldsymbol{\ell}$	$r(\boldsymbol{\ell})$	contribution to $\mathsf{Pr}(\mathrm{CS}\|\boldsymbol{p})$	contribution to $\mathsf{E}(T\|\boldsymbol{p})$
$(1,1,2)$	12	0	0	$48p_1p_2p_3^2$
$(1,2,1)$	12	0	0	$48p_1p_2^2p_3$
$(2,1,1)$	12	1	$12p_1^2p_2p_3$	$48p_1^2p_2p_3$
$(1,2,2)$	6	0	0	$30p_1p_2^2p_3^2$
$(2,1,2)$	6	$1/2$	$3p_1^2p_2p_3^2$	$30p_1^2p_2p_3^2$
$(2,2,1)$	6	$1/2$	$3p_1^2p_2^2p_3$	$30p_1^2p_2^2p_3$
$(0,0,3)$	1	0	0	$3p_3^2$
$(0,3,0)$	1	0	0	$3p_2^3$
$(3,0,0)$	1	1	p_1^3	$3p_1^3$
$(0,1,3)$	3	0	0	$12p_2p_3^3$
$(0,3,1)$	3	0	0	$12p_2^3p_3$
$(1,0,3)$	3	0	0	$12p_1p_3^3$
$(1,3,0)$	3	0	0	$12p_1p_2^3$
$(3,0,1)$	3	1	$3p_1^3p_3$	$12p_1^3p_3$
$(3,1,0)$	3	1	$3p_1^3p_2$	$12p_1^3p_2$
$(0,2,3)$	6	0	0	$30p_2^2p_3^3$
$(0,3,2)$	6	0	0	$30p_2^3p_3^2$
$(2,0,3)$	6	0	0	$30p_1^2p_3^3$
$(2,3,0)$	6	0	0	$30p_1^2p_2^3$
$(3,0,2)$	6	1	$6p_1^3p_3^2$	$30p_1^3p_3^2$
$(3,2,0)$	6	1	$6p_1^3p_2^2$	$30p_1^3p_2^3$

and

$$\mathsf{E}(T|\boldsymbol{p} = \text{SC}) \;=\; \sum_{\boldsymbol{\ell}\in\mathcal{T}} \#\boldsymbol{\ell}\; s(\boldsymbol{\ell})(\theta^\star)^{\ell_1} p^{s(\ell)}. \tag{13}$$

In addition, in the case of the EP, we have

$$\mathsf{Pr}(\text{CS}|\boldsymbol{p} = \text{EP}) \;=\; 1/k \tag{14}$$

and

$$\mathsf{E}(T|\boldsymbol{p} = \text{EP}) \;=\; \sum_{\boldsymbol{\ell}\in\mathcal{T}} \#\boldsymbol{\ell}\; s(\boldsymbol{\ell})(1/k)^{s(\ell)}. \tag{15}$$

Example 25. Again consider the B-K curtailed procedure with $k = 3$ and upper bound $B = 5$ from Example 24. The first three columns of Table 3 give all of the possible termination points ℓ, along with their associated $\#\boldsymbol{\ell}$ and $r(\boldsymbol{\ell})$ values. Column 4 of Table 3 gives each ℓ's contribution $\#\boldsymbol{\ell}\, r(\boldsymbol{\ell}) \prod_{i=1}^{k} p_i^{\ell_i}$ to the overall $\mathsf{Pr}(\text{CS}|\boldsymbol{p})$ for any underlying probability configuration $\boldsymbol{p}$. Column 5 of the table gives each ℓ's contribution $\#\boldsymbol{\ell}\, s(\boldsymbol{\ell}) \prod_{i=1}^{k} p_i^{\ell_i}$ to the overall $\mathsf{E}(T|\boldsymbol{p})$ for any underlying probability configuration $\boldsymbol{p}$. Note that, if we add up all the $\mathsf{E}(T|\boldsymbol{p} = \text{EP})$ expected values, we get 37/9.

7 Extensions:Multivariate Normal

Up to this point we have used i.i.d. observations for simulating the multinomial procedures, but it is also possible that the simulator may induce positive correlation among different competing simulated systems. In some cases, the simple technique of common random numbers can be used. In other cases, more complex methods can also be implemented. The reason we use correlation is, as correlation increases among populations, it becomes easier for the experimenter to distinguish which of the populations is the "best".

In the previous selection procedures we used, it is obvious that, an increase in $\theta^\star$ results in the distinction of the best multinomial system. In the following example we illustrate how positive correlation induction yields increased $\theta^\star$.

Example 26. Suppose that $k = 2$, and X_i's are normally distributed with unknown mean μ_i and known common variance σ^2. If one observation is larger that another,

Table 3: Performance characteristics for Example 25

ℓ	$\#\boldsymbol{\ell}$	$r(\boldsymbol{\ell})$	contribution to $\mathsf{Pr}(\mathrm{CS}\|\boldsymbol{p}=\mathrm{SC})$	contribution to $\mathsf{E}(T\|\boldsymbol{p}=\mathrm{SC})$	contribution to $\mathsf{E}(T\|\boldsymbol{p}=\mathrm{EP})$
$(1,1,2)$	12	0	0	$48(\theta^\star)^1p^4$	48/81
$(1,2,1)$	12	0	0	$48(\theta^\star)^1p^4$	48/81
$(2,1,1)$	12	1	$12(\theta^\star)^2p^4$	$48(\theta^\star)^2p^4$	48/81
$(1,2,2)$	6	0	0	$30(\theta^\star)^1p^5$	30/243
$(2,1,2)$	6	1/2	$3(\theta^\star)^2p^5$	$30(\theta^\star)^2p^5$	30/243
$(2,2,1)$	6	1/2	$3(\theta^\star)^2p^5$	$30(\theta^\star)^2p^5$	30/243
$(0,0,3)$	1	0	0	$3p^3$	3/27
$(0,3,0)$	1	0	0	$3p^3$	3/27
$(3,0,0)$	1	1	$(\theta^\star)^3p^3$	$3(\theta^\star)^3p^3$	3/27
$(0,1,3)$	3	0	0	$12p^4$	12/81
$(0,3,1)$	3	0	0	$12p^4$	12/81
$(1,0,3)$	3	0	0	$12(\theta^\star)^1p^4$	12/81
$(1,3,0)$	3	0	0	$12(\theta^\star)^1p^4$	12/81
$(3,0,1)$	3	1	$3(\theta^\star)^3p^4$	$12(\theta^\star)^3p^4$	12/81
$(3,1,0)$	3	1	$3(\theta^\star)^3p^4$	$12(\theta^\star)^3p^4$	12/81
$(0,2,3)$	6	0	0	$30p^5$	30/243
$(0,3,2)$	6	0	0	$30p^5$	30/243
$(2,0,3)$	6	0	0	$30(\theta^\star)^2p^5$	30/243
$(2,3,0)$	6	0	0	$30(\theta^\star)^2p^5$	30/243
$(3,0,2)$	6	1	$6(\theta^\star)^3p^5$	$30(\theta^\star)^3p^5$	30/243
$(3,2,0)$	6	1	$6(\theta^\star)^3p^5$	$30(\theta^\star)^3p^5$	30/243

the larger one is taken to be more desirable. Define $p_1 \equiv P(X_1 > X_2)$ and $p_2 \equiv 1-p_1$. Suppose $\mu_1 > \mu_2$; so we can let $p_1 \equiv \theta p$ and $p_2 \equiv p$, where $\theta = \frac{1-p}{p} > 1$. Also let, $\rho \equiv Corr(X_1, X_2) \geq 0$.

Then,

$$\begin{aligned} p_1 &= P(X_1 > X_2) = P(X_1 - X_2 > 0) \\ &= P\{[X_1 - X_2 - (\mu_1 - \mu_2)]/\omega > -(\mu_1 - \mu_2)]/\omega\}, \\ &\quad \text{where } \omega = \sqrt{2\sigma^2(1-p)} \\ &= 1 - \Phi(-(\mu_1 - \mu_2)/\omega) = \Phi((\mu_1 - \mu_2)/\omega), \\ &\quad \text{where } \Phi(.) \text{is the N(0,1) cdf} \\ &= \theta_\rho p, \text{say}, \ = 1 - p. \end{aligned}$$

So $\theta_\rho = (1-p)/p = \Phi(\eta)/(1-\Phi(\eta))$,
where $\eta = (\mu_1 - \mu_2)/\omega$.
Hence,
$\theta_\rho/\theta_0 = [\Phi(\eta)/\Phi(\eta^{‘})] \times [(1-\Phi(\eta))/(1-\Phi(\eta^{‘}))]$,
where $\eta^{‘} = \eta\sqrt{1-p}$.
This quantity is obviously greater than 1; $\theta_\rho > \theta_0$.

Simulating Multivariable Normal

To clarify correlation induction, we have simulated different setups, In this subsection, you may find tables that illustrate how positive correlation induction can result in increased $\theta^\star$.

For simulation we have generated multivariate normal random numbers with different systems. For each multivariate normal distribution we used correlation coefficient ρ between values 0 and 1 with 0.1 increment.

In the table below, you can find results generated from 40000 replications. In the top row of the table, we have the systems we are comparing. In the first column of the table you can see the simulation results, when we compare two identical $N(0,1)$ (Normally distributed with $mean = 0$ and $variance = 1$) systems, by using different ρ (correlation coefficient) values.

In this example we used, $\theta^\star = 2$, $P^\star = 0,9$ values and corresponding $n_0 = 15$

value from table 13. For termination, we used the same conditions that we have used at $\mathcal{M}_{\text{BG}}$procedure.

Table 4: Multivariate Normal with $k = 2$, $\theta^\star = 2$, $P^\star = .90$ values

	$N(0,1)$ vs. $N(0,1)$		$N(1,1)$ vs. $N(0,1)$		$N(1,1)$ vs. $N(1,2)$		$N(1,1)$ vs. $N(1,4)$		$N(1,1)$ vs. $N(2,4)$	
ρ	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$
0.0	0.4998	10.57	0.9812	7.20	0.4992	10.59	0.5042	10.57	0.9115	8.79
0.1	0.4981	10.59	0.9870	6.97	0.4994	10.62	0.5017	10.59	0.9151	8.74
0.2	0.4994	10.57	0.9899	6.71	0.5002	10.60	0.5011	10.60	0.9191	8.67
0.3	0.4987	10.58	0.9934	6.44	0.4970	10.56	0.4986	10.57	0.9229	8.61
0.4	0.4969	10.62	0.9965	6.13	0.4986	10.56	0.4999	10.57	0.9296	8.52
0.5	0.4983	10.58	0.9980	5.82	0.5014	10.58	0.5029	10.58	0.9334	8.45
0.6	0.4974	10.59	0.9993	5.40	0.5050	10.58	0.4956	10.58	0.9411	8.34
0.7	0.5001	10.55	0.9998	4.97	0.5020	10.57	0.5021	10.59	0.9447	8.29
0.8	0.5001	10.60	1.0000	4.51	0.5020	10.57	0.5010	10.59	0.9486	8.18
0.9	0.5039	10.55	1.0000	4.11	0.5022	10.60	0.4985	10.57	0.9526	8.08
1.0	0.5000	15.00	1.0000	4.00	0.5015	10.61	0.5008	10.60	0.9592	7.97

Table 5: Multivariate Normal with $k = 3$, $\theta^\star = 2$, $P^\star = .90$ values

	$N(1,1)$ vs. $N(1,1)$ & $N(1,1)$		$N(1,2)$ vs. $N(1,1)$ & $N(1,1)$		$N(2,1)$ vs. $N(1,1)$ & $N(1,1)$		$N(2,2)$ vs. $N(1,1)$ & $N(1,1)$		$N(1,4)$ vs. $N(1,2)$ & $N(1,1)$	
ρ	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$
0.0	0.3340	23.38	0.4713	23.04	0.9923	10.83	0.9855	11.90	0.5306	22.55
0.1	0.3329	23.23	0.4853	22.84	0.9945	10.25	0.9890	11.45	0.5332	22.52
0.2	0.3302	23.28	0.4982	22.86	0.9960	9.72	0.9917	11.04	0.5451	22.43
0.3	0.3310	23.33	0.5143	22.73	0.9971	9.12	0.9929	10.53	0.5523	22.38
0.4	0.3335	23.35	0.5277	22.79	0.9986	8.49	0.9947	10.09	0.5655	22.28
0.5	0.3325	23.35	0.5553	22.58	0.9993	7.82	0.9966	9.56	0.5815	22.19
0.6	0.3294	23.31	0.5790	22.39	0.9997	7.11	0.9978	9.05	0.5868	22.01
0.7	0.3335	23.24	0.6156	22.15	1.0000	6.40	0.9984	8.50	0.6029	21.82
0.8	0.3321	23.28	0.6662	21.63	1.0000	5.70	0.9993	7.89	0.6236	21.65
0.9	0.3389	23.24	0.7422	20.65	1.0000	5.13	0.9998	7.25	0.6340	21.53
1.0	0.0000	18.07	0.7646	17.18	1.0000	5.00	0.9998	6.45	0.6544	21.12

These tables justify our claim that positive correlation induction results same as increased $\theta^\star$. When we examine the tables closely, when comparing two systems with different mean and same standard deviation we see that as the ρ value increases, the process is more in favor of the desired category, i.e. the probability of correct selection increases and the expected number of observations to be taken decreases. If we look at "N(1,1) vs. N(0,1)" column we clearly see these results. The Pr(CS)value increases from 0.9812 to 1 and $\mathsf{E}(T|\boldsymbol{p})$value decreases from 7.1954 to 4.

Another observation from Tables 4,5 is, when we compare the results of two different observations, when the mean of desired category is larger than the other categories, by increasing the variance of desired category and keeping all other values the same, it is against the favorable category. For example take "N(2,1) vs. N(1,1) & N(1,1)"and "N(2,2) vs. N(1,1) & N(1,1)" columns, As the variance increases the Pr(CS)value decreases (from 0.9923 to 0.9855) and $\mathsf{E}(T|\boldsymbol{p})$value increases (from 10.8279 to 11.9047).

Table 6: Comparision of $k = 2, 3$, and 4 for $\theta^{\star} = 2$, $\theta^{\star} = 1.4$ and $P^{\star} = 0.90$

		$N(1,2)$ vs. $N(0,4)$		$N(1,2)$ vs. $N(0,4)$ and $N(0,4)$		$N(1,2)$ vs.$N(0,4)$, $N(0,4)$ and $N(0,4)$	
$\theta^{\star}$	ρ	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$	Pr(CS)	$\mathsf{E}(T\|\boldsymbol{p})$
	0,0	0.8896	9.05	0.8644	18.23	0.8365	28.86
	0.1	0.8973	8.98	0.8733	18.08	0.8388	28.57
	0.2	0.8977	8.94	0.8714	18.04	0.8420	28.43
	0.3	0.9059	8.91	0.8798	17.83	0.8463	28.17
	0.4	0.9076	8.82	0.8854	17.68	0.8542	27.90
2.0	0.5	0.9101	8.79	0.8894	17.50	0.8579	27.88
	0.6	0.9156	8.72	0.8966	17.34	0.8660	27.59
	0.7	0.9196	8.67	0.9005	17.16	0.8712	27.31
	0.8	0.9245	8.59	0.9042	16.99	0.8762	27.01
	0.9	0.9291	8.53	0.9131	16.78	0.8792	26.71
	1.0	0.9356	8.43	0.9165	16.66	0.8894	26.44
	0	0.9892	21.53	0.9903	43.89	0.9893	69.24
	0.1	0.9897	21.17	0.9916	43.06	0.9890	68.36
	0.2	0.9915	20.96	0.9925	42.32	0.9895	67.41
	0.3	0.9918	20.59	0.9929	41.71	0.9912	66.36
	0.4	0.9919	20.32	0.9936	40.93	0.9917	65.22
1.4	0.5	0.9926	19.98	0.9937	40.24	0.9916	64.09
	0.6	0.9937	19.71	0.9948	39.56	0.9931	63.06
	0.7	0.9949	19.36	0.9952	38.69	0.9933	62.03
	0.8	0.9954	18.99	0.9960	38.09	0.9943	60.82
	0.9	0.9959	18.57	0.9967	36.99	0.9944	59.38
	1	0.9968	18.14	0.9967	36.28	0.9953	58.32

In Table 6 the desired category is $N\ (1,2)$ in both cases. In the first case, we have two categories, $N\ (1,2)$ and $N\ (0,4)$, but in the second case we have one additional rival -which is identical to the rival in the previous case- so we have three systems: $N\ (1,2)$, $N\ (0,4)$ and $N\ (0,4)$. When comparison is with two systems it is easier for desired category to dominate, on the other hand when we have three categories, the probability of selecting the desired category decreases. In the first case, the probability of selecting the desired category is $P(N\ (1,2) > N\ (0,4))$, but in the second case the probability will become, $P(N\ (1,2) > N\ (0,4)\ \text{and} N\ (0,4))$. Hence, the desired category should defeat both systems. As it can also be seen from the simulation results, it is clear that the first probability is larger than the second one.

8 Conclusions and Future Work

In this thesis we studied indifference- zone multinomial selection procedures, which are for selecting the most probable ("best") multinomial cell. We have reviewed five popular multinomial selection procedures from the literature, and tested their performances by using Monte Carlo simulations. After reviewing these processes, we have proposed an alternative approach: Random Walk, in order to show that the procedures are consistent compared with different point if views. Beside the simulations, we also discussed the exact calculation methodologies for the procedures, in a generic form. In addition to multinomial selection procedures, we proposed a multivariate normal extension, which we induce positive correlation among different competing systems.

In this thesis we work on indifference-zone multinomial selection procedures, which stop when one of the competitors is sufficiently ahead of the others, i.e. guarantees to win. As an alternative approach we can use elimination of the worst competitor. If one of the competitors is sufficiently behind the others, i.e. guarantees to lose, we can remove that category from further selection procedure. Because of cost considerations, we are looking for taking the minimum number observations, which will ensure the probability of correct being greater than $P^{\star}$. For a closed, sequential procedure we can also model it as a dynamic programming problem and solve it to find the required number of observations. In this study we proposed an extension: Multivariate Normal, which we use multivariate normal distribution instead of i.i.d. multivariate, but we have not deeply studied this idea. New procedures which will handle with the multivariate normal distributed cases can be formed.

9 Appendix

Table 7: Performance Characteristics of Procedures $\mathcal{M}_{\text{BEM}}$ and $\mathcal{M}_{\text{BK}}$ for $k = 2$ and 3. ($\mathsf{Pr}(\text{CS}|\text{SC})$ values are for both procedures; $\mathsf{E}[T|\text{SC}]$ and $\mathsf{E}[T|\text{EP}]$ values are only for procedure $\mathcal{M}_{\text{BK}}$.)

		$k=2$				$k=3$			
$P^\star$	$\theta^\star$	n_{BEM}	$\mathsf{Pr}(\text{CS}\|\text{SC})$	$\mathsf{E}[T\|\text{SC}]$	$\mathsf{E}[T\|\text{EP}]$	n_{BEM}	$\mathsf{Pr}(\text{CS}\|\text{SC})$	$\mathsf{E}[T\|\text{SC}]$	$\mathsf{E}[T\|\text{EP}]$
	3.0	1	0.7456	1.00	1.00	5	0.7696	3.95	4.11
	2.8	3	0.8257	2.39	2.50	6	0.7834	4.49	4.92
	2.6	3	0.8094	2.40	2.50	6	0.7519	4.55	4.92
	2.4	3	0.7919	2.42	2.50	7	0.7457	5.56	5.79
0.75	2.2	3	0.7703	2.43	2.50	9	0.7523	7.29	7.68
	2.0	5	0.7911	3.96	4.12	12	0.7568	9.91	10.43
	1.8	5	0.7505	4.00	4.12	17	0.7594	14.42	15.08
	1.6	9	0.7642	7.31	7.53	26	0.7484	22.74	23.59
	1.4	17	0.7631	14.23	14.68	52	0.7498	47.21	48.52
	1.2	55	0.7550	49.12	50.04	181	0.7499	171.78	174.45
	3.0	7	0.9286	5.16	5.81	11	0.9027	8.46	9.49
	2.8	7	0.9165	5.23	5.81	13	0.9050	10.11	11.32
	2.6	7	0.8984	5.28	5.82	15	0.9047	11.83	13.23
	2.4	9	0.9087	6.82	7.54	18	0.9022	14.43	16.02
0.90	2.2	11	0.9068	8.43	9.29	22	0.9044	17.98	19.79
	2.0	15	0.9120	11.70	12.85	29	0.9044	24.23	26.44
	1.8	19	0.9020	15.16	16.48	40	0.9019	34.29	37.00
	1.6	31	0.9064	25.50	27.54	64	0.9010	56.56	60.16
	1.4	59	0.9027	50.71	53.83	126	0.9003	115.20	120.59
	1.2	199	0.8984	181.95	188.70	427	0.8974	406.81	416.97
	3.0	9	0.9503	6.54	7.55	17	0.9559	12.96	15.07
	2.8	11	0.9550	8.03	9.31	19	0.9519	14.69	16.94
	2.6	13	0.9576	9.55	11.06	22	0.9535	17.26	19.78
	2.4	15	0.9552	11.21	12.84	26	0.9519	20.77	23.59
0.95	2.2	19	0.9573	14.41	16.45	32	0.9519	26.10	29.32
	2.0	23	0.9535	17.79	20.15	42	0.9498	35.02	38.91
	1.8	33	0.9550	26.22	29.38	59	0.9514	50.49	55.31
	1.6	49	0.9503	40.34	44.38	94	0.9505	82.89	89.32
	1.4	97	0.9522	83.60	90.08	186	0.9510	169.83	179.39
	1.2	327	0.9502	299.90	313.66	645	0.9504	614.03	632.56
	3.0	19	0.9903	13.30	16.50	29	0.9900	21.81	26.45
	2.8	21	0.9905	14.91	18.31	33	0.9902	25.20	30.29
	2.6	25	0.9913	17.98	21.96	39	0.9904	30.25	36.04
	2.4	29	0.9903	21.21	25.66	46	0.9900	36.38	42.77
0.99	2.2	37	0.9912	27.62	33.13	58	0.9916	46.84	54.33
	2.0	47	0.9909	35.97	42.50	75	0.9909	62.03	70.85
	1.8	65	0.9910	51.27	59.52	106	0.9914	90.08	101.06
	1.6	101	0.9914	82.80	93.97	167	0.9908	146.43	160.72
	1.4	193	0.9907	166.20	182.88	330	0.9908	300.12	321.17
	1.2	653	0.9900	599.29	633.52	1148	0.9896	1090.57	1131.48

Table 8: Performance Characteristics of Procedures $\mathcal{M}_{\text{BEM}}$ and $\mathcal{M}_{\text{BK}}$ for $k = 4$ and 5. (Pr(CS|SC) values are for both procedures; $\mathsf{E}[T|\text{SC}]$ and $\mathsf{E}[T|\text{EP}]$ values are only for procedure $\mathcal{M}_{\text{BK}}$.)

		$k=4$				$k=5$			
$P^\star$	$\theta^\star$	n_{BEM}	Pr(CS|SC)	E[T|SC]	E[T|EP]	n_{BEM}	Pr(CS|SC)	E[T|SC]	E[T|EP]
0.75	3.0	8	0.7685	6.43	6.91	11	0.7687	9.28	9.92
	2.8	9	0.7670	7.39	7.88	12	0.7509	10.22	10.89
	2.6	10	0.7575	8.28	8.82	14	0.7477	12.08	12.80
	2.4	12	0.7543	10.10	10.68	17	0.7529	14.87	15.69
	2.2	15	0.7523	12.86	13.56	22	0.7541	19.55	20.50
	2.0	20	0.7524	17.49	18.31	29	0.7543	26.17	27.29
	1.8	29	0.7558	25.90	27.00	41	0.7523	37.64	38.98
	1.6	46	0.7532	42.07	43.48	68	0.7551	63.63	65.43
	1.4	92	0.7495	86.46	88.41	137	0.7515	130.87	133.36
	1.2	326	0.7496	315.43	319.29	486	0.7559	474.52	479.26
0.90	3.0	16	0.9030	12.97	14.49	21	0.9029	17.64	19.54
	2.8	19	0.9072	15.58	17.38	24	0.9013	20.41	22.44
	2.6	22	0.9046	18.31	20.25	29	0.9048	24.97	27.29
	2.4	26	0.9011	21.98	24.09	35	0.9028	30.55	33.13
	2.2	33	0.9039	28.36	30.84	44	0.9039	38.99	41.89
	2.0	43	0.8985	37.69	40.56	58	0.9027	52.19	55.61
	1.8	61	0.9046	54.55	58.10	83	0.9031	76.06	80.16
	1.6	98	0.9031	89.77	94.33	134	0.9037	125.15	130.42
	1.4	196	0.9043	184.15	190.78	271	0.9041	258.44	265.97
	1.2	692	0.9006	669.53	682.31	964	0.9004	940.36	954.55
0.95	3.0	23	0.9523	18.53	21.22	29	0.9511	24.19	27.29
	2.8	26	0.9497	21.23	24.09	34	0.9543	28.70	32.15
	2.6	31	0.9516	25.64	28.93	40	0.9525	34.24	38.01
	2.4	37	0.9514	31.14	34.73	48	0.9520	41.69	45.84
	2.2	46	0.9506	39.38	43.46	61	0.9539	53.74	58.55
	2.0	61	0.9528	53.20	58.09	81	0.9531	72.56	78.18
	1.8	87	0.9530	77.50	83.52	115	0.9523	104.99	111.66
	1.6	139	0.9518	126.89	134.64	185	0.9522	172.27	180.80
	1.4	278	0.9522	260.65	271.84	374	0.9531	355.81	368.08
	1.2	979	0.9519	946.18	967.47	1331	0.9533	1296.80	1320.00
0.99	3.0	39	0.9898	30.96	36.66	48	0.9910	39.49	45.82
	2.8	45	0.9900	36.26	42.50	56	0.9907	46.66	53.65
	2.6	53	0.9907	43.29	50.27	66	0.9918	55.78	63.43
	2.4	63	0.9908	52.39	60.06	80	0.9915	68.68	77.20
	2.2	79	0.9911	66.90	75.68	100	0.9914	87.25	96.87
	2.0	104	0.9921	89.89	100.23	133	0.9919	118.20	129.45
	1.8	147	0.9910	130.02	142.52	189	0.9913	171.22	184.78
	1.6	235	0.9900	213.19	229.33	305	0.9909	282.32	299.68
	1.4	471	0.9909	439.63	462.99	616	0.9908	583.61	608.42
	1.2	1660	0.9894	1600.54	1644.99	2191	0.9904	2130.06	2176.92

Table 9: Performance Characteristics of Procedures $\mathcal{M}_{\text{BEM}}$ and $\mathcal{M}_{\text{BK}}$ for $k = 6$ and 7. (Pr(CS|SC) values are for both procedures; $\mathsf{E}[T|\text{SC}]$ and $\mathsf{E}[T|\text{EP}]$ values are only for procedure $\mathcal{M}_{\text{BK}}$.)

		$k=6$				$k=7$			
$P^\star$	$\theta^\star$	n_{BEM}	Pr(CS|SC)	E[T|SC]	E[T|EP]	n_{BEM}	Pr(CS|SC)	E[T|SC]	E[T|EP]
	3.0	14	0.7624	12.13	12.94	17	0.7618	15.01	15.92
	2.8	16	0.7607	14.01	14.85	20	0.7604	17.85	18.85
	2.6	19	0.7598	16.83	17.76	23	0.7573	20.74	21.78
	2.4	23	0.7600	20.62	21.63	28	0.7514	25.51	26.66
0.75	2.2	29	0.7547	26.34	27.49	36	0.7537	33.18	34.50
	2.0	38	0.7525	34.95	36.30	48	0.7510	44.79	46.28
	1.8	56	0.7599	52.29	53.94	70	0.7519	66.16	67.95
	1.6	90	0.7567	85.38	87.44	114	0.7543	109.17	111.43
	1.4	184	0.7555	177.46	180.40	234	0.7542	227.10	230.38
	1.2	658	0.7532	645.89	651.29	840	0.7520	827.40	833.29
	3.0	26	0.9022	22.41	24.57	31	0.9016	27.22	29.60
	2.8	30	0.9005	26.14	28.48	36	0.8997	31.94	34.49
	2.6	36	0.9032	31.72	34.32	43	0.9022	38.53	41.36
	2.4	44	0.9059	39.28	42.16	53	0.9060	48.01	51.19
0.90	2.2	56	0.9071	50.62	53.95	68	0.9059	62.36	65.98
	2.0	74	0.9060	67.83	71.65	90	0.9058	83.57	87.70
	1.8	106	0.9044	98.64	103.21	130	0.9064	122.33	127.26
	1.6	172	0.9078	162.70	168.49	211	0.9063	201.32	207.56
	1.4	349	0.9032	335.82	344.03	430	0.9029	416.52	425.16
	1.2	1249	0.9003	1224.51	1239.83	1545	0.9012	1519.92	1535.94
	3.0	36	0.9526	30.79	34.33	42	0.9496	36.65	40.36
	2.8	41	0.9517	35.48	39.24	49	0.9528	43.20	47.27
	2.6	49	0.9507	42.93	47.05	59	0.9542	52.54	57.10
	2.4	60	0.9532	53.22	57.87	72	0.9533	64.85	69.91
0.95	2.2	76	0.9555	68.35	73.61	91	0.9517	83.07	88.68
	2.0	101	0.9538	92.20	98.28	121	0.9539	111.87	118.36
	1.8	144	0.9528	133.51	140.76	174	0.9538	163.10	170.86
	1.6	233	0.9491	219.75	228.97	283	0.9524	269.27	279.03
	1.4	475	0.9522	456.18	469.26	578	0.9510	558.69	572.43
	1.2	1697	0.9504	1661.96	1686.42	2075	0.9520	2039.16	2064.58
	3.0	58	0.9912	49.00	55.91	68	0.9907	58.58	65.98
	2.8	68	0.9909	58.13	65.75	79	0.9912	68.84	76.83
	2.6	80	0.9911	69.25	77.58	94	0.9919	82.85	91.65
	2.4	97	0.9906	85.19	94.33	114	0.9908	101.76	111.44
0.99	2.2	122	0.9913	108.69	119.01	145	0.9917	131.15	142.12
	2.0	163	0.9914	147.55	159.59	193	0.9910	177.11	189.69
	1.8	233	0.9911	214.55	228.95	277	0.9914	258.17	273.08
	1.6	377	0.9916	353.70	371.84	450	0.9906	426.17	445.06
	1.4	766	0.9912	732.83	758.76	918	0.9905	884.18	911.03
	1.2	2737	0.9904	2674.89	2723.51	3297	0.9911	3234.11	3283.89

Table 10: Performance Characteristics of Procedures $\mathcal{M}_{\text{BEM}}$ and $\mathcal{M}_{\text{BK}}$ for $k = 8$ and 10. (Pr(CS|SC) values are for both procedures; E[T|SC] and E[T|EP] values are only for procedure $\mathcal{M}_{\text{BK}}$.)

		$k = 8$				$k = 10$			
$P^\star$	$\theta^\star$	n_{BEM}	Pr(CS|SC)	E[T|SC]	E[T|EP]	n_{BEM}	Pr(CS|SC)	E[T|SC]	E[T|EP]
	3.0	20	0.7545	17.95	18.93	27	0.7578	24.76	25.93
	2.8	23	0.7496	20.83	21.87	31	0.7532	28.64	29.87
	2.6	28	0.7576	25.57	26.77	37	0.7529	34.44	35.77
	2.4	34	0.7564	31.36	32.65	46	0.7532	43.17	44.66
0.75	2.2	43	0.7530	40.08	41.51	60	0.7631	56.78	58.48
	2.0	59	0.7623	55.58	57.27	81	0.7614	77.32	79.27
	1.8	86	0.7600	81.91	83.95	118	0.7632	113.63	115.95
	1.6	140	0.7600	134.87	137.41	192	0.7557	186.65	189.43
	1.4	286	0.7559	278.87	282.40	396	0.7576	388.51	392.43
	1.2	1030	0.7537	1016.96	1023.34	1433	0.7550	1419.48	1426.30
	3.0	37	0.9052	32.97	35.60	47	0.8997	42.80	45.64
	2.8	43	0.9067	38.65	41.49	57	0.9086	52.28	55.52
	2.6	51	0.9047	46.30	49.38	67	0.9046	62.00	65.40
	2.4	63	0.9073	57.78	61.22	83	0.9112	77.42	81.26
0.90	2.2	80	0.9081	74.18	78.01	105	0.9087	98.81	103.05
	2.0	107	0.9074	100.28	104.72	141	0.9065	133.96	138.78
	1.8	154	0.9080	146.03	151.33	204	0.9092	195.68	201.37
	1.6	251	0.9039	241.02	247.60	335	0.9049	324.59	331.67
	1.4	514	0.9026	500.04	509.18	688	0.9041	673.51	683.32
	1.2	1851	0.9039	1825.35	1842.21	2489	0.9024	2462.66	2480.29
	3.0	49	0.9536	43.39	47.41	64	0.9576	57.88	62.45
	2.8	58	0.9549	51.82	56.29	74	0.9571	67.48	72.35
	2.6	69	0.9559	62.25	67.15	89	0.9557	81.91	87.19
	2.4	84	0.9526	76.68	81.98	109	0.9565	101.24	107.03
0.95	2.2	106	0.9557	97.81	103.73	138	0.9545	129.40	135.79
	2.0	142	0.9556	132.59	139.39	186	0.9548	176.10	183.46
	1.8	205	0.9549	193.76	201.90	269	0.9544	257.29	266.00
	1.6	334	0.9539	319.90	330.15	440	0.9521	425.37	436.21
	1.4	684	0.9519	664.20	678.51	903	0.9532	882.61	897.68
	1.2	2464	0.9517	2427.59	2453.83	3269	0.9515	3231.58	3259.05
	3.0	78	0.9911	68.26	76.02	99	0.9920	88.68	97.10
	2.8	91	0.9919	80.50	88.90	115	0.9918	103.94	112.97
	2.6	108	0.9915	96.57	105.72	137	0.9917	125.03	134.79
	2.4	132	0.9915	119.36	129.49	168	0.9918	154.81	165.59
0.99	2.2	167	0.9914	152.91	164.19	214	0.9912	199.33	211.32
	2.0	224	0.9907	207.65	220.79	287	0.9912	270.16	283.89
	1.8	322	0.9910	302.78	318.17	415	0.9899	395.08	411.34
	1.6	525	0.9904	500.61	520.17	679	0.9913	653.97	674.36
	1.4	1074	0.9900	1039.85	1067.18	1394	0.9910	1358.90	1387.42
	1.2	3869	0.9903	3805.54	3856.22	5043	0.9908	4978.42	5030.75

Table 11: Performance Characteristics of Procedure $\mathcal{M}_{\text{RA}}$ for $k = 2$ and 3

		$k=2$				$k=3$			
$P^\star$	$\theta^\star$	(r,N)	Pr(CS|SC)	E[T|SC]	E[T|EP]	(r,N)	Pr(CS|SC)	E[T|SC]	E[T|EP]
0.75	3.0	(1,1)	0.7500	1.00	1.00	(2,3)	0.7962	3.68	4.25
	2.8	(2,2)	0.8274	2.39	2.50	(2,3)	0.7733	3.74	4.25
	2.6	(2,2)	0.8113	2.40	2.50	(2,3)	0.7510	3.81	4.25
	2.4	(2,2)	0.7914	2.42	2.50	(2,5)	0.7601	4.70	5.54
	2.2	(2,2)	0.7661	2.43	2.50	(3,5)	0.7555	6.39	7.05
	2.0	(2,3)	0.7737	3.09	3.25	(4,5)	0.7556	8.81	9.63
	1.8	(2,4)	0.7555	3.44	3.61	(4,7)	0.7570	12.39	13.74
	1.6	(3,5)	0.7559	5.96	6.26	(4,12)	0.7572	18.24	20.93
	1.4	(5,9)	0.7553	12.68	13.27	(6,15)	0.7075	30.28	32.67
	1.2	(11,29)	0.7552	46.89	48.87	(14,48)	0.7038	117.16	123.16
0.90	3.0	(2,12)	0.9000	3.20	4.00	(3,5)	0.9004	6.76	8.74
	2.8	(3,4)	0.9156	4.63	5.34	(3,6)	0.9004	7.58	10.13
	2.6	(3,5)	0.9154	5.23	6.24	(3,8)	0.9030	8.62	12.12
	2.4	(3,6)	0.9113	5.72	6.94	(4,8)	0.9104	11.79	15.51
	2.2	(3,8)	0.9024	6.32	7.83	(4,10)	0.9026	13.67	18.47
	2.0	(4,8)	0.9033	8.90	10.59	(4,15)	0.9015	16.51	23.60
	1.8	(4,14)	0.9011	11.00	13.92	(5,19)	0.9040	24.43	34.21
	1.6	(5,21)	0.9006	17.00	21.48	(6,30)	0.9021	37.82	53.07
	1.4	(9,32)	0.9023	39.71	47.80	(9,52)	0.9051	79.36	106.76
	1.2	(20,104)	0.9062	152.02	177.59	(27,157)	0.9018	361.32	415.56
0.95	3.0	(3,6)	0.9522	5.25	6.94	(4,7)	0.9505	9.77	13.73
	2.8	(4,6)	0.9542	6.88	8.57	(4,8)	0.9504	10.75	15.52
	2.6	(4,7)	0.9545	7.55	9.65	(4,10)	0.9510	12.18	18.48
	2.4	(4,9)	0.9548	8.47	11.38	(4,14)	0.9509	13.84	22.82
	2.2	(4,12)	0.9506	9.39	13.12	(5,14)	0.9511	18.58	27.88
	2.0	(5,14)	0.9537	13.09	17.90	(5,22)	0.9505	22.40	37.03
	1.8	(6,18)	0.9511	18.04	24.33	(6,31)	0.9548	32.31	53.77
	1.6	(8,27)	0.9520	29.44	39.18	(8,40)	0.9520	53.56	82.35
	1.4	(11,54)	0.9518	56.80	77.67	(12,72)	0.9532	112.10	164.23
	1.2	(27,172)	0.9532	238.31	301.29	(36,228)	0.9517	532.21	636.84

Table 12: Performance Characteristics of Procedure $\mathcal{M}_{\text{RA}}$ for $k = 4$ and 5

		$k=4$				$k=5$			
$P^\star$	$\theta^\star$	(r,N)	Pr(CS|SC)	E[T|SC]	E[T|EP]	(r,N)	Pr(CS|SC)	E[T|SC]	E[T|EP]
0.75	3.0	(2,4)	0.7718	5.15	6.39	(2,5)	0.7544	6.66	8.76
	2.8	(2,5)	0.7621	5.68	7.17	(3,5)	0.8046	10.25	12.90
	2.6	(3,5)	0.8045	8.92	10.88	(3,5)	0.7707	10.59	12.89
	2.4	(3,5)	0.7731	9.22	10.90	(3,6)	0.7683	12.65	15.58
	2.2	(3,6)	0.7626	10.88	12.96	(3,8)	0.7683	15.70	19.94
	2.0	(3,8)	0.7525	13.37	16.17	(3,11)	0.7504	18.85	24.29
	1.8	(4,10)	0.7579	20.72	24.30	(4,12)	0.7518	28.75	34.88
	1.6	(4,20)	0.7514	30.05	37.29	(5,19)	0.7533	48.60	58.56
	1.4	(6,33)	0.7530	65.27	78.23	(7,35)	0.7516	99.71	118.62
	1.2	(16,89)	0.7532	273.40	302.04	(21,106)	0.7532	431.10	470.55
0.90	3.0	(3,7)	0.9091	9.87	14.71	(3,8)	0.9046	12.58	19.95
	2.8	(3,9)	0.9069	11.02	17.42	(3,11)	0.9038	14.19	24.26
	2.6	(4,8)	0.9050	14.39	19.81	(4,9)	0.9044	18.74	26.96
	2.4	(4,10)	0.9107	16.98	24.35	(4,11)	0.9046	21.87	32.46
	2.2	(4,13)	0.9050	19.92	29.76	(5,13)	0.9081	29.83	42.48
	2.0	(5,15)	0.9060	27.93	39.22	(5,18)	0.9093	37.76	56.94
	1.8	(5,25)	0.9022	36.07	55.15	(6,22)	0.9004	53.05	74.41
	1.6	(7,33)	0.9098	63.33	89.82	(7,36)	0.9016	84.33	122.00
	1.4	(12,56)	0.9021	142.59	181.62	(10,68)	0.9014	176.25	245.95
	1.2	(38,108)	0.9006	617.24	672.24	(36,210)	0.9031	864.02	964.39
0.95	3.0	(4,9)	0.9577	13.75	22.19	(4,10)	0.9573	17.37	29.80
	2.8	(4,10)	0.9520	14.96	24.36	(4,11)	0.9505	18.97	32.39
	2.6	(4,12)	0.9504	16.70	28.11	(4,15)	0.9506	21.72	40.99
	2.4	(5,13)	0.9531	22.30	34.65	(5,15)	0.9577	29.05	48.16
	2.2	(5,17)	0.9518	26.28	43.35	(5,19)	0.9506	34.15	58.96
	2.0	(6,21)	0.9534	36.26	57.56	(6,23)	0.9518	47.27	78.04
	1.8	(7,28)	0.9507	50.97	79.66	(8,29)	0.9507	72.32	108.05
	1.6	(9,43)	0.9521	84.53	128.40	(10,45)	0.9511	118.88	175.37
	1.4	(16,76)	0.9517	196.68	226.94	(14,87)	0.9516	245.24	355.48
	1.2	(54,258)	0.9506	886.33	966.43	(44,283)	0.9509	1147,24	1315.94

Table 13: Performance Characteristics of Procedure $\mathcal{M}_{\text{BG}}$ for $k = 2$ and 3

		$k = 2$				$k = 3$			
$P^\star$	$\theta^\star$	n_{BG}	Pr(CS|SC)	E[T|SC]	E[T|EP]	n_{BG}	Pr(CS|SC)	E[T|SC]	E[T|EP]
0.75	3.0	1	0.7456	1.00	1.00	5	0.7570	3.24	3.84
	2.8	3	0.8257	2.39	2.50	6	0.7638	3.70	4.27
	2.6	3	0.8094	2.40	2.50	7	0.7567	3.93	4.60
	2.4	3	0.7919	2.42	2.50	8	0.7588	5.41	6.21
	2.2	3	0.7703	2.43	2.50	10	0.7535	6.00	7.03
	2.0	5	0.7763	3.07	3.25	13	0.7471	7.99	9.25
	1.8	7	0.7504	3.44	3.62	18	0.7506	11.32	13.04
	1.6	9	0.7521	5.97	6.25	32	0.7529	17.62	20.47
	1.4	19	0.7588	11.28	12.05	71	0.7507	33.87	40.11
	1.2	67	0.7522	36.74	39.42	285	0.7526	117.96	139.75
0.90	3.0	6	0.8969	3.77	5.06	12	0.9013	6.99	9.26
	2.8	7	0.9147	4.64	5.35	15	0.9037	7.76	10.62
	2.6	9	0.9179	5.23	6.28	16	0.9001	9.15	12.21
	2.4	11	0.9113	5.73	6.95	22	0.9015	10.42	14.54
	2.2	15	0.9025	6.32	7.84	25	0.9012	13.31	18.13
	2.0	15	0.9024	8.89	10.61	34	0.9026	17.18	23.57
	1.8	27	0.9032	11.00	13.98	50	0.9007	23.65	33.22
	1.6	41	0.9005	17.07	21.48	83	0.8989	37.46	53.15
	1.4	79	0.8989	32.93	42.15	170	0.8994	73.70	104.73
	1.2	267	0.8991	111.83	142.51	670	0.8978	253.04	368.39
0.95	3.0	11	0.9516	5.26	6.96	20	0.9508	8.87	13.73
	2.8	15	0.9516	5.62	7.82	22	0.9523	10.47	16.04
	2.6	13	0.9535	7.57	9.68	25	0.9519	12.30	18.68
	2.4	17	0.9563	8.43	11.39	31	0.9512	14.52	22.38
	2.2	23	0.9506	9.40	13.17	41	0.9521	17.62	28.01
	2.0	27	0.9537	13.08	17.88	52	0.9484	23.07	36.27
	1.8	35	0.9491	18.00	24.35	71	0.9525	32.57	50.72
	1.6	59	0.9502	26.54	37.19	125	0.9505	50.39	81.79
	1.4	151	0.9475	48.33	72.12	266	0.9499	99.50	164.76
	1.2	455	0.9506	166.12	242.39	960	0.9523	344.97	575.82
0.99	3.0	21	0.9916	9.57	15.39	33	0.9903	14.19	25.87
	2.8	23	0.9902	10.13	16.36	38	0.9901	15.81	29.49
	2.6	31	0.9909	10.91	19.92	46	0.9902	18.19	35.07
	2.4	35	0.9900	14.11	24.17	54	0.9900	22.33	42.31
	2.2	48	0.9903	15.58	28.27	69	0.9907	26.82	52.66
	2.0	59	0.9936	20.53	37.17	92	0.9901	34.60	68.69
	1.8	89	0.9901	27.26	51.71	137	0.9900	48.46	99.70
	1.6	158	0.9904	42.46	84.53	255	0.9907	77.28	171.65
	1.4	273	0.9907	82.12	156.78	505	0.9904	152.27	338.44
	1.2	935	0.9909	278.93	543.95	2000	0.9905	529.42	1239.04

Table 14: Performance Characteristics of Procedure $\mathcal{M}_{\text{BG}}$ for $k = 4$ and 5

		$k = 4$				$k = 5$			
$P^\star$	$\theta^\star$	n_{BG}	Pr(CS|SC)	E[T|SC]	E[T|EP]	n_{BG}	Pr(CS|SC)	E[T|SC]	E[T|EP]
	3.0	9	0.7666	5.79	6.81	12	0.7595	7.45	9.16
	2.8	9	0.7508	6.02	7.09	13	0.7518	8.37	10.23
	2.6	11	0.7539	7.03	8.48	17	0.7595	9.80	12.26
	2.4	15	0.7556	8.30	10.12	20	0.7521	11.97	14.78
0.75	2.2	17	0.7505	10.50	12.52	25	0.7519	14.99	18.47
	2.0	24	0.7559	13.79	16.65	34	0.7503	19.78	24.54
	1.8	35	0.7513	19.45	23.53	50	0.7477	28.29	35.30
	1.6	57	0.7480	31.31	37.86	86	0.7511	45.60	57.72
	1.4	124	0.7512	62.48	76.18	184	0.7504	92.81	117.08
	1.2	495	0.7492	220.84	273.27	730	0.7505	333.13	422.21
	3.0	19	0.9065	9.91	14.33	24	0.9039	13.12	19.13
	2.8	22	0.9042	11.22	16.33	28	0.9024	15.02	22.06
	2.6	26	0.9000	13.18	19.18	34	0.9016	17.33	26.03
	2.4	31	0.9039	15.91	23.03	42	0.9013	21.22	31.97
0.90	2.2	39	0.9033	19.80	28.72	52	0.9011	26.58	39.85
	2.0	53	0.9011	25.59	37.49	71	0.9020	35.28	53.17
	1.8	75	0.8998	36.85	53.92	104	0.9030	50.36	76.44
	1.6	126	0.8980	58.86	86.85	172	0.9001	81.10	123.58
	1.4	274	0.9035	117.14	177.37	374	0.9012	164.91	253.87
	1.2	1050	0.9000	417.16	636.50	1460	0.9010	590.56	923.03
	3.0	26	0.9526	12.98	20.69	34	0.9509	16.56	27.58
	2.8	30	0.9478	14.73	23.69	39	0.9509	19.23	31.85
	2.6	36	0.9510	17.13	27.81	46	0.9510	22.63	37.43
	2.4	44	0.9504	20.67	33.81	58	0.9518	27.19	45.88
0.95	2.2	56	0.9514	25.88	42.53	74	0.9518	33.92	57.96
	2.0	74	0.9508	33.81	55.73	98	0.9512	45.17	76.58
	1.8	106	0.9506	48.26	79.50	142	0.9508	64.78	109.31
	1.6	180	0.9496	76.46	129.08	240	0.9489	103.83	179.88
	1.4	380	0.9497	153.00	263.76	510	0.9520	209.40	370.56
	1.2	1500	0.9516	544.74	957.39	2000	0.9502	750.17	1352.08
	3.0	46	0.9900	18.85	37.64	57	0.9906	23.65	48.49
	2.8	52	0.9901	21.17	42.46	65	0.9900	27.02	55.26
	2.6	61	0.9902	25.02	50.21	78	0.9904	31.99	66.24
	2.4	73	0.9903	30.30	60.36	92	0.9903	38.81	78.53
0.99	2.2	95	0.9902	37.35	76.97	117	0.9905	48.48	98.44
	2.0	130	0.9900	48.93	103.71	158	0.9901	63.94	132.60
	1.8	192	0.9906	70.10	151.46	225	0.9900	91.45	189.40
	1.6	340	0.9905	111.41	262.97	390	0.9904	148.21	319.71
	1.4	650	0.9902	224.62	495.42	730	0.9901	297.02	608.78
	1.2	2500	0.9903	792.39	1811.22	2950	0.9904	1068.97	2307.34

References

[1] J. Banks, J. Carson, B. Nelson, and D. Nicol (2005). *Discrete-Event System Simulation*, 4th ed. Prentice Hall, Upper Saddle River, New Jersey.

[2] R. E. Bechhofer, S. Elmaghraby, and N. Morse (1959). "A single-sample multiple decision procedure for selecting the multinomial event which has the largest probability." *Annals of Mathematical Statistics*, 30, 102–119.

[3] R. E. Bechhofer and D. M. Goldsman (1985) "On the Ramey-Alam sequential procedure for selecting the multinomial event which has the largest probabilty." *Communications in Statistics–Simulation and Computation*, 14, 263–282.

[4] R. E. Bechhofer and D. M. Goldsman (1985). "Truncation of the Bechhofer-Kiefer-Sobel sequential procedure for selecting the multinomial event which has the largest probability." *Communications in Statistics–Simulation and Computation*, B14, 283–315.

[5] R. E. Bechhofer and D. M. Goldsman (1986). "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,* B15, 829–851.

[6] R. E. Bechhofer, J. Kiefer, and M. Sobel *Sequential Identification and Ranking Procedures (with Special Reference to Koopman-Darmois Populations)*. University of Chicago Press, Chicago.

[7] R. E. Bechhofer and R. V. Kulkarni (1984). "Closed sequential procedures for selecting the multinomial events which have the largest probabilities." *Communications in Statistics–Theory and Methods*, A13, 2997–3031.

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

[9] T. Cacoullos and M. Sobel (1966). "An inverse-sampling procedure for selecting the most probable event in a multinomial distribution." In *Multivariate Analysis* (ed. P. R. Krishhnaiah), 423–455. Academic Press, New York.

[10] P. Chen (1992). "Truncated selection procedures for the most probable event and the least probable event." *Annals of Mathematical Statistics*, B44, 613–622.

[11] D. M. Goldsman and B. L. Nelson (2001). "Statistical selection of the best system." *Proceedings of the 2001 Winter Simulation Conference* (eds. J. Joines, R. R. Barton, P. Fishwick, and K. Kang), 139–146. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

[12] W. W. Hines, D. C. Montgomery, D. M. Goldsman, and C. M. Borror (2003). *Probability and Statistics in Engineering*, 4th ed. John Wiley & Sons, New York.

[13] J. C. Hsu (1984). "Constrained simultaneous confidence intervals for multiple comparisons with the best." *Annals of Statistics*, 12, 1136–1144.

[14] H. Kesten and N. Morse (1959). "A property of the multinomial distribution." *Annals Math. Statist.*, 30, 120–127.

[15] A. M. Law (2006). *Simulation Modeling and Analysis*, 4th ed. McGraw-Hill, New York.

[16] J. T. Ramey and K. Alam (1979). "A sequential procedure for selecting the most probable multinomial event." *Biometrika*, 66, 171–173.

[17] S. M. Ross (2003). *Introduction to Probability Models*, 8th ed. Academic Press, San Diego.

[18] A. Wald (1947). *Sequential Analysis*, Wiley, New York. (Reprinted by Dover Publications, 1973.)