Summer 2013

New Statistical Methods for Simulation Output Analysis

Huan Yu
University of Iowa

NEW STATISTICAL METHODS FOR SIMULATION OUTPUT ANALYSIS

by

Huan Yu

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Industrial Engineering
in the Graduate College of
The University of Iowa

August 2013

Thesis Supervisor: Associate Professor Yong Chen

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Huan Yu

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Industrial Engineering at the August 2013 graduation.

Thesis Committee: _____
Yong Chen, Thesis Supervisor

_____
Pavlo A. Krokhmal

_____
Ibrahim Tarik Ozbolat

_____
Russell V. Lenth

_____
Kate Cowles

To my parents.

ACKNOWLEDGMENTS

First and foremost I would like to thank my advisor Prof. Yong Chen. He has taught me a lot, both in research and course work. Not only instructed me about our research contents, he also spent efforts training me with good research habits. I really enjoy and appreciate my PhD years with him.

I would like to thank my parents for being so supportive to me about everything since I was born. Also, I appreciate the encouragement and love from all my friends.

ABSTRACT

In this thesis, there are generally three contributions to the Ranking and Selection problem in discrete-event simulation area. Ranking and selection is an important problem when people want to select single or multiple best designs from alternative pool.

There are two different types in discrete-event simulation: terminating simulation and steady-state simulation. For steady-state simulation, there is an initial trend before the data output enters into the steady-state, if we cannot start the simulation from steady state. We need to remove the initial trend before we use the data to estimate the steady-state mean. Our first contribution regards the application to eliminate the initial trend/initialization bias. In this thesis, we present a novel solution to remove the initial trend motivated by offline change detection method. The method is designed to monitor the cumulative absolute bias from the estimated steady-state mean. Experiments are conducted to compare our procedure with other existing methods. Our method is shown to be at least no worse than those methods and in some cases much better. After removing the initialization bias, we can apply a ranking and selection procedure for the data outputs from steady-state simulation.

There are two main approaches to ranking and selection problem. One is subset selection and the other one is indifference zone selection. Also by employing directed graph, some single-best ranking and selection methods can be extended to solve multi-best selection problem. Our method is designed to solve multi-best ranking and selection. And in Chapter 3, one procedure for ranking and selection in terminating simulation is extended based full sequential idea. It means we compare the sample means among all systems in contention at each stage. Also, we add a technique to do pre-selection of the superior systems at the same time of eliminating inferior systems. This can accelerate the speed of obtaining the number of best systems we want. Experiments are conducted to demonstrate the pre-selection technique can save observation significantly compared with

the procedure without it. Also compared with existing methods, our procedure can save significant number of observations. We also explore the effect of common random number. By using it in the simulation process, more observations can be saved.

The third contribution of this thesis is to extend the procedure in Chapter 3 for steady-state simulation. Asymptotic variance is employed in this case. We justify our procedure in asymptotic point of view. And by doing extensive experiments, we demonstrate that our procedure can work in most cases when sample size is finite.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Discrete event simulation is a classical tool to solve models with arbitrarily large state spaces, as long as the state of the system changes at a discrete set of points in time. In discrete event simulation, large amount of input data can be generated easily using random number generators. The output from the simulation model can be used to do further analysis and design of the system. For example, we can estimate the performance of a given system, compare the performance of multiple alternative designs based on the simulation outputs from each system, or find the optimal design of a system with the best estimated performance. Discrete event simulations have been applied in a large variety of areas such as manufacturing simulation, supply chain simulation, computer network simulation, capital investment decision evaluation, and hospital operation room occupation simulation.

One important usage of simulation outputs is to select one or more best systems from a finite number of alternatives, which is called the *ranking and selection* (R&S) problem. Here, the "best" system is typically defined by the system with the maximum/minimum mean performance. In other words, a R&S method is to determine which system(s) have the largest/smallest mean performance. R&S procedures should be designed differently for the two different categories of discrete event simulation models, namely, the *terminating simulation* and the *steady-state simulation*.

The first category of discrete event simulation is *terminating/replicated simulation*, which runs for some time interval of T_E, where E is a specified event that stops the simulation (Banks, Carson et al. 2004). For example, when we simulate the daily operations of a bank, each replication of the simulation can be used to simulate one-day operation of the bank. The terminating event E of the simulation is the closing of the bank at the end of the day. For every replication, the simulation starts from the same

initial system state corresponding to the opening of the bank at the beginning of a day when no customer is in the bank. The performance output from each replication such as the number of customers served and the waiting times of customers can be averaged across multiple replications to estimate the mean performance of the bank.

The second category of discrete event simulation is *steady-state simulation*. The analysis based on steady-state simulation focuses on the long-run performance of the system, in other words, the steady-state performance. For example, an emergency room operates continuously for 24 hours each day. There is no event *E* in this case. And we are interested in the performance of the emergency room in long terms: the daily average patient or the average time that it is occupied by a patient. It would be more appropriate to use steady-state simulation in this case. If the simulation starts in the empty state (no patient), it might take a while for the system to enter the steady state, since "empty" is obviously not the steady-state of an emergency room. System will enter into steady-state after a *transient period*, if it does not start in the steady state. Using the data from the transient period to estimate steady-state performance will cause bias in the estimation, which is called the *initialization bias*. We will introduce the initialization bias problem in more details in Section 1.1 and the R&S problems for terminating and steady-state simulation in Section 1.2 and Section 1.3, respectively.

1.1 Initialization Bias Elimination

When we are performing a steady-state simulation, the long-run performance which is independent of the initial conditions, is of our interest. However, the length of a simulation run is finite, so residual effects of the initial values will always exist in the outputs. These residual effects, also called *initialization bias* (IB), could cause trouble when we estimate parameters of interest in the steady state. For example, a queuing problem which starts empty will take a while to reach its steady state. If we want to estimate the average number of customers or waiting times at the steady state, a

truncation point must be determined and then the initialization bias could be eliminated by removing the data from the beginning to the truncation point. Then the data points left are considered as from the steady state and will be used to estimate the steady-state parameters. The problem of determining a truncation point of an output data series from a steady-state simulation in order to accurately estimate the steady-state parameters is called the *IB elimination* problem. To illustrate the importance of the IB elimination problem, Figure 1.1 shows an AR(1) data series starting from non-steady state, which causes initialization bias.

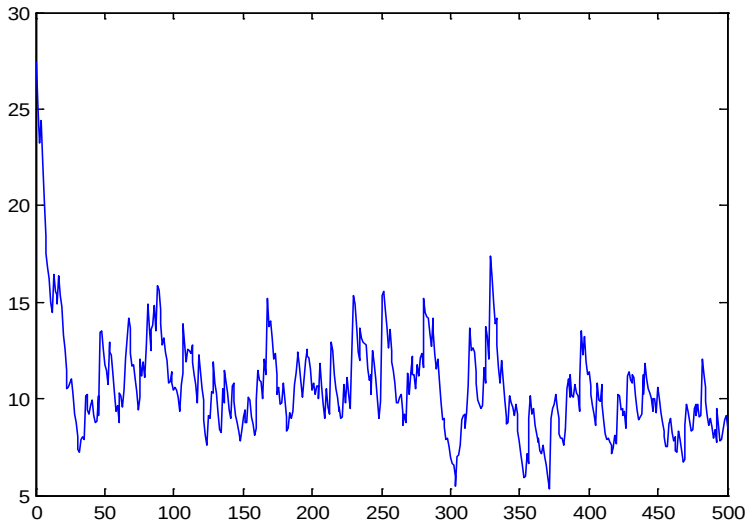

Figure 1.1. AR(1) data series with initialization bias

There is a long history of study of the IB elimination problem. Robinson (2002; 2004; 2007) and Hoad, Robinson et al.(2008) point out that there are six categories of methods. Graphical methods identify initialization bias based on visual inspection of the simulation outputs or statistics of simulation outputs. Graphical methods are simple to use but depends too much on individual's subjective judgment. The second category is heuristic methods which design rules to eliminate IB. The Marginal Standard Error Rule

(MSER) (White 1995; White 1997; Spratt 1998) belongs to this category. Heuristic methods are simple to implement and not depending on subjective judgment. The third category is statistical methods which use statistical principles to determine the initialization bias. One example is the regression method proposed by Law and Kelton (2000). Statistical methods designed so far are much more complicated than the methods belonging to the first two categories and require more computational time. The fourth category of methods are initialization bias tests (Schruben 1982; Schruben 1983; Vassilacopoulos 1989; Cash 1992; Cash, Dippold et al. 1992; Goldsman, Schruben et al. 1994), which are used to solve a hypothesis testing problem to test whether initialization bias exists. These tests focus on testing the existence of IB rather than accurate estimation of truncation point and steady-state parameters. The last one is hybrid methods which combine the methods from different categories, but are often complicated and time-consuming.

Robinson (2007) employs the principles of statistical process control (SPC) as a new category of methods to eliminate IB. In this thesis, we call this method as the SPC method. SPC method is a special case of the change detection methods, which are procedures used to detect the location of a change point within a data series after which the distribution of the data is changed. SPC is an on-line change-detection method which does not utilize information after the detected change point. If the data after change point are available, which is the case in IB elimination problem, it will be more efficient to determine the change point utilizing the information from all the data. In this case the change detection method is considered as offline. Based on our knowledge, no IB elimination method has been developed based on the idea of off-line change detection.

1.2 Ranking and Selection Procedures for Terminating Simulation

Ranking and selection methods are techniques for comparing "populations" with the goal of finding the single best or multi best among them, where "best" is defined by the maximum or minimum population mean. The selection is often assured by an assigned probability of correct selection.

As stated in Kim and Nelson (2005), ranking and selection methods can trace back to two papers: Bechhofer (1954a) proposes the indifference-zone procedure (IZP) and Gupta (1965) establishes the subset selection procedure (SSP). Both focused on single best selection. Assume we want to select the single best system which holds the largest mean among *K* alternatives and the true means of all systems satisfy $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_K$. In indifference-zone procedure, the goal is to guarantee selecting the best system, *K*, with probability at least $1-\alpha$, if $\mu_K - \mu_{K-1} \geq \delta$, where δ is called the *indifference zone parameter*, which is the practically significant difference worth to detect and determined by the experimenter. To be more specific, the procedure should guarantee $\Pr\{\text{select } K | \mu_K - \mu_{K-1} \geq \delta\} \geq 1-\alpha$. While a subset selection procedure is to obtain a subset $I \subset \{1, 2, \ldots . K\}$ such that $\Pr\{K \in I\} \geq 1-\alpha$ (Gupta 1965; Kim and Nelson 2005).

Based on IZP, Dudewicz and Dalal (1975) propose a two-stage procedure to solve the single best selection problem when variances of alternatives are unknown. The first stage is used to estimate the variance. Dudewicz and Dalal's procedure specifies N_i, the total number of observations required from alternative system *i*, if all systems are independent and normally distributed, as

$$N_i = \max\{n_0 + 1, \left\lceil (\frac{hS_i}{\delta})^2 \right\rceil\},$$

where n_0 is the number of observations in the first stage; *h* is a constant calculated from standard deviation obtained based the first stage observations; and δ is the

indifference zone parameter. The procedure by Dudewicz and Dalal (1975) is based on *weighted* sample averages of the outputs from each system, which may not be convenient to use. Rinott (1978) designed another two-stage procedure that is based on (un-weighted) sample means.

When the number of alternatives is large, it may not be efficient to generate the same number of observations for all the systems. If some systems are obviously inferior to other systems, fewer observations from the inferior systems are enough to eliminate them from contention. Based on this idea, Nelson, Swann et al. (2001) design a procedure (Combined procedure) which combines an initial screening stage to remove the obviously inferior systems based on a subset selection procedure, and a second stage of selecting the best system among the remaining systems based on an indifference zone selection procedure.

The combined procedure has only one screening stage to eliminate the inferior systems, while the fully sequential procedure by Kim and Nelson (2001) evaluates systems and eliminate the inferior ones in multiple sequential stages. After generating n_0 initial observations, in each stage, the fully sequential procedure sample one observation (or one batch of observations) for all the alternatives in contention and eliminates the obviously inferior systems immediately. This way more saving in the total number of observations is achieved. Swisher (2003; 2004) provides a comprehensive survey of the single best R&S methods.

Sometimes, selecting one single best system from the alternatives cannot satisfy experimenters' need. For example, there might be some hidden or unmodeled factors in a system, such as political or environmental reasons (Kiekhaefer 2011). It is possible that the single best system/design selected cannot be adopted. It may be more desirable to select multi-best systems (best k systems, $k>1$) to provide more options. In such a case, a ranking and selection procedure that selects multi-best systems will be useful. An example is that a consulting company may want to offer clients more alternative choices.

In medicine, multi-best R&S can answer questions like ‘what are top three diet habits most likely cause obesity?’ In agriculture, it can answer questions like ‘what are the top five crop plans having the largest yields?’

Although the majority of R&S literature focuses on the single-best problems, there are methods proposed for multi-best selection. Most of them are further developed based on the existing single best R&S methods. Extensions of SSP to solve multi-best R&S were introduced by Carroll, Gupta et al. (1975) and Bofinger and Mangersen (1986). IZP is also employed in multi-best R&S area. Bechhofer (1954b) extends his method (Bechhofer 1954a) to select more than one best systems. Dudewicz and Dalal’s (1975) two-stage procedure can solve both single-best and multi-best problems. Kiekhaefer (2011) develops multi-best R&S methods by extending Rinott’s (1978) two-stage procedure and Nelson, Swann et al.’s (2001) combined procedure. For one of the most efficient single-best R&S methods, the fully sequential procedure (Kim and Nelson 2001), there is no work existing trying to extend it into the multi-best R&S area. One of the aims of this thesis is to fill this gap.

1.3 Ranking and Selection Procedures for Steady-state Simulation

Most existing R&S methods are constructed under the assumption that the output data from each system are independent and identically distributed. And most terminating simulations can satisfy this assumption. However, the outputs from a single replication of steady-state simulation are mostly dependent. Typically, there are two possible methods to deal with this situation. For the first method, we generate *m* replications of outputs from each system. Each replication should be long enough to reach the steady state. Then we need to remove the data from the transient period to avoid initialization bias for each replication of each system. After removing initialization bias, we use the sample mean of the remaining data in each replication as one basic observation. In this way the

observations of each system are independent of each other and approximately normally distributed and the R&S methods for terminating systems can be applied in this scenario. The drawback of this method is that IB elimination for each replication causes substantial waste of data. The other method is to generate one single long replication. We only need to remove the initialization bias once. However, after removing the initialization bias, the remaining data are still dependent of each other. Although the classical batching method could normalize the data and make them more independent, the size of the batch might need to be very large to achieve this goal and it's hard to control. With large batch sizes, the interval of decision on eliminating inferior systems will be long, which causes waste of observations, especially for the fully sequential procedures. Therefore, procedures that can use basic observations from single-replication of steady-state simulation are more desirable.

Based on this idea, Nakayama (1997) and Damerdji and Nakayama (1999) extended some ranking and selection procedures for terminating systems into steady-state area. Goldsman et al. (2002) present a fully-sequential procedure FS+ to solve the single-best ranking and selection problem for steady-state simulation. FS+ is based on one single replication of simulation output and the concept of *asymptotic variance* is employed. Experiments under different settings are also conducted and the results show that FS+ is efficient in most finite-sample cases. And Kim and Nelson (2006) prove theoretically that the procedure FS+ can assure certain probability of correct selection for selecting single best system for steady-state simulation by appropriate asymptotic analysis. Both of the papers focus on the single-best R&S problem. It's very likely that the multi-best ranking and selection problem for steady-state simulation can be solved by extending the FS+ procedure.

1.4 Thesis Objectives and Outline

The objective of the thesis is to solve the following related problems in simulation output analysis:

1. Develop an efficient and robust IB elimination method for steady-state simulation based on the idea of offline change detection methods: In this thesis, we develop a new procedure, called the *cusum rule*, to remove IB bias. The derivation of the cusum rule is motived by modifying and applying an offline change detection method to solve the IB elimination problem. The cusum rule is a heuristic rule and does not rely on statistical assumptions. It is a robust and efficient method to solve initialization bias elimination problem. We will conduct extensive experiments to compare it with some other methods in the literature, particularly the MSER method which is considered as a benchmark method in the literature. This work is presented in Chapter 2 of this thesis.
2. Develop a fully sequential R&S procedure that selects k ($k \geq 1$) best alternatives with a guaranteed probability of correct selection for terminating simulation: We develop a fully sequential procedure, called FS-k procedure, which is an extension of the procedure from Kim and Nelson (2001), to solve the multi-best R&S problem for terminating simulation. We also provide theoretical justification of the new procedure by proving the ensured probability of correct selection. Extensive experiments are conducted to study the performance of this procedure. This work is presented in Chapter 3 of this thesis.
3. Adapt FS-k to steady-state simulation: In this thesis, we adapt the FS-k procedure so that it can be used for selection of multi-best systems based on basic observations from single replications of steady-state simulation. The concept of asymptotic variance is introduced. The procedure is theoretically proved to be effective in asymptotic sense. And different asymptotic variance estimators are

considered and compared. Experiments are conducted to test its performance for finite samples. More details of this work are presented in Chapter 4 of this thesis.

CHAPTER 2

CUSUM-RULE FOR INITIALIZATION BIAS ELIMINATION IN STEADY-STATE SIMULATION

2.1 Review of Related Work

2.1.1 Problem Settings and MSER Method

For an output series $\{Y_i, i = 1, \ldots n\}$ from a simulation, initialization bias elimination is to find a truncation point *d* so that the steady state mean of the output, μ, is estimated by $\hat{\mu} = \frac{1}{n-d}\sum_{i=d+1}^{n} Y_i$.

One of the most popular methods for initialization bias elimination is the Marginal Standard Error Rules (MSER). It is a method originally designed by White (1995; 1997) to determine the truncation point by minimizing the width of marginal confidence interval of the sample mean after truncation. More specifically, the truncation point *d* is found by:

$$d = \text{argmin}_{0\le d<n} \frac{1}{(n-d)^2}\sum_{i=d+1}^{n}\left(Y_i - \bar{Y}_{n,d}\right)^2, \text{ where } \bar{Y}_{n,d} = \frac{1}{n-d}\sum_{i=d+1}^{n} Y_i$$

Spratt (1998) revised the method by applying the MSER on batch means of data instead of the original data. If the batch size is *m*, then the updated method is called MSER-*m*. It is claimed that *m*=5 performs particularly well (Spratt 1998; White and Spratt 2000; Pasupathy and Schmeiser 2010).

One shortcoming of MSER-*m* is that it can mistakenly determine a truncation point very close to the end of the data series when the last a few data are quite close in values (Hoad, Robinson et al. 2010). To avoid this situation, a common practice is to require that no more than half of the data can be deleted (White and Spratt 2000; Pasupathy and Schmeiser 2010). That is, the truncation point *d* is found by:

$$d = \text{argmin}_{0\le d\le \frac{n}{2}} \frac{1}{(n-d)^2}\sum_{i=d+1}^{n}\left(Y_i - \bar{Y}_{n,d}\right)^2.$$

Many authors have conducted extensive tests and found that MSER-5 performs very well, comparing with the other methods tested. For examples, White and Spratt (2000) compare five heuristic rules from the literature and find that MSER is the most effective rule in eliminating bias and outperforms all other rules for most data types considered in the chapter. Also, Hoad, Robinson et al.(2010) test 3048 cases using data having initial bias of various shapes, severities, lengths, and degrees of autocorrelation and show that MSER-5 performs robustly and effectively for most of data sets tested. Moreover, MSER-5 is easy to understand and implement. Thus, it is generally considered as a benchmark method for initialization bias elimination.

2.1.2 Change Detection Problems

In this chapter, we will propose a new method for initialization bias elimination that is originally introduced in the area of change detection. In this subsection, we will first introduce the change detection problems and two basic change detection methods that can potentially be applied for initialization bias elimination.

For a data series $\{X_t, t = 1, \dots n\}$, first assume that X_t are independently distributed with $X_i \sim p_\theta$, where $\theta = \theta_0$, when $1 \le t \le \tau - 1$, and $\theta = \theta_1 \ne \theta_0$, when $t \ge \tau$. In a change detection problem, we want to estimate the value of change time τ based on the observed data. There are two types of change detection methods: online and offline. For *online* change detection methods, the data are observed sequentially and the decision on whether a change in θ has occurred should be made as soon as possible at each time t, $t = 1, 2, \dots$ using only the observed data up to time t. In other words, a decision should be made at time t only based on $\{X_1, \dots, X_t\}$, the data observed up to time *t*. For *offline* change detection methods, the change time is estimated after all the data $\{X_i, i = 1, \dots n\}$ are observed. That is, we can use all the data to decide whether and when a change occurs.

The most important application of online change detection is statistical process control (SPC). In SPC, X_i represents measurements on a quality characteristic of a product. The process is called *in-control* if it is operating with only the chance causes of variation, which are inherent or natural variability of the process that always exists. The process is called *out-of-control* if it is operating in the presence of variability that are not coming from chance cause pattern, such as improperly adjusted or controlled machines or operator's mistakes (Montgomery 2004). Suppose $X_i \sim p_{\theta_0}$ when the process is in-control and $X_i \sim p_{\theta_1}$ when it is out-of-control. The goal of a SPC method is to detect the out-of-control status as soon as possible by sequentially observing the quality measurement data $\{X_1, X_2, X_3, \ldots\}$. Therefore SPC is a collection of online change detection methods. The most popular and basic method of SPC for online change detection is the Shewhart control chart (Page 1955; Montgomery 2004). To construct a control chart, historical data thought to be in-control are first collected to estimate in-control process parameters such as mean and variance, and then determine the control limits (CL) of the control chart. After the control chart is constructed, the data points to be monitored will be plotted on the control chart. If a data point falls inside the control limits, the process is considered as in-control. Otherwise, a signal of a change to the out-of-control state is detected. Please refer to Montgomery (2004) for more details on construction and usage of control charts in SPC.

Online detection only uses the data before *t*. If data after the change point are available, offline detection algorithm will be more efficient. For *offline* change detection, we will introduce a method based on the cumulative sums (cusum). If the change occurs at time *t*, the likelihood function for $\{X_i, i = 1, \ldots n\}$ can be written as:

$$L_t = \prod_{i=1}^{t-1} p_{\theta_0}(X_i) \prod_{i=t}^{n} p_{\theta_1}(X_i)$$

By using the maximum likelihood principle, we can estimate τ by

$$\hat{\tau} = \underset{0 \le t \le n}{\operatorname{argmax}} L_t \tag{2.1}$$

Then (2.1) can be written as

$$\hat{\tau} = \underset{0 \le t \le n}{\operatorname{argmax}} \prod_{i=1}^{t-1} p_{\theta_0}(X_i) \prod_{i=t}^{n} p_{\theta_1}(X_i) = \underset{0 \le t \le n}{\operatorname{argmax}} \sum_{i=1}^{t-1} \ln\left(p_{\theta_0}(X_i)\right) + \sum_{i=t}^{n} \ln\left(p_{\theta_1}(X_i)\right)$$

$$= \underset{0 \le t \le n}{\operatorname{argmax}} \sum_{i=t}^{n} \ln\left(p_{\theta_1}(X_i)\right) - \sum_{i=t}^{n} \ln\left(p_{\theta_0}(X_i)\right)$$

The last equality is obtained by substracting $\sum_{i=1}^{n} \ln\left(p_{\theta_0}(X_i)\right)$ from $\sum_{i=1}^{t-1} \ln\left(p_{\theta_0}(X_i)\right) + \sum_{i=t}^{n} \ln\left(p_{\theta_1}(X_i)\right)$.

If we further assume that $X_i \sim N(\mu, \sigma^2)$, where $\mu = \mu_0$ when X_i is in-control, and $\mu = \mu_1 > \mu_0$when X_i is out of control, it can be seen that

$$\hat{\tau} = \underset{0 \le t \le n}{\operatorname{argmax}} \sum_{i=t}^{n} \ln\left(p_{\mu_1}(X_i)\right) - \sum_{i=t}^{n} \ln\left(p_{\mu_0}(X_i)\right) = \underset{0 \le t \le n}{\operatorname{argmax}} \sum_{i=t}^{n} (X_i - \mu_0 - \frac{\mu_1 - \mu_0}{2})$$

Note that $\sum_{i=t}^{n}(X_i - \mu_0 - \frac{\mu_1 - \mu_0}{2})$ is the cumulative sum (cusum) of the shifts of X_i from $\mu_0 + \frac{\mu_1 - \mu_0}{2}$. The idea of cusum derived above based on the parametric assumptions can be generalized to nonparametric situations where the change-point τ is estimated by:

$$\hat{\tau} = \operatorname{argmax}_{0 \le t \le n} \sum_{i=t}^{n}(X_i - \mu_0 - K), \tag{2.2}$$

where the parameter *K* is called the reference value. For normal distribution with known μ_1, K is set as $\frac{\mu_1 - \mu_0}{2}$. When cusum is used as a nonparametric method, *K* is a parameter of the method that needs to be decided. Intuitively, $X_i - \mu_0 - K$ can be considered as a "significant" evidence of a shift from μ_0. And the cusum value can be considered as an accumulation of these evidences of shift for all the observations after *t*. The change-point is estimated by maximizing the cumulative evidence of the significant shifts. Because of this intuitive interpretation of the cusum method, it can work as a nonparametric method with μ_1 unknown, although it is originally motivated under the assumption of normal distribution with known μ_1.

2.1.3 SPC Method for Initialization Bias Elimination

From the last two subsections, we can see that there are a lot of similarities between IB problem and the change detection problem. The steady state data from simulation can be considered as data without changes in distribution (or "in-control"). The data with initial transient can be considered as data after a change has occurred (or "out-of-control"). Then detection of the period of initialization bias becomes the problem of detecting when the distribution of the data starts to change. Therefore the change detection algorithms can be borrowed to solve the initialization bias problem. This idea has been used by Robinson (2007) in terms of applying the Shewhart control chart to determine the initialization bias period in discrete-event simulation. A general outline for this method is given in Procedure 1.

Procedure 1 (SPC method)

1. Collect *r* replications of simulation outputs $\{\{Y_{i1}, \dots, Y_{in}\}, i = 1, \dots r\}$. The mean series are defined as: $\bar{Y} = (\frac{\sum_{i=1}^{r} Y_{i1}}{r}, \dots, \frac{\sum_{i=1}^{r} Y_{in}}{r})$.
2. Determine the smallest batch size m so that the batch means can pass the test of no-autocorrelation using Fishman's procedure (Fishman 1978; Fishman 1996) and the Anderson-Darling test of normality (Anderson and Darling 1954; Law and Kelton 2000). Batched data series are obtained as: $b = \left\lfloor \frac{n}{m} \right\rfloor$
   $\bar{\bar{Y}}_j = \frac{\sum_{i=(j-1)m+1}^{jm} \bar{Y}_i}{m}, j = 1, \dots, b$
3. Assume the second half of data series is in control (or in steady-state). A control chart is constructed by using the second half of the data as the 'historical data'. The control limits are obtained by:
   $\hat{\mu} = \frac{1}{\lfloor \frac{b}{2} \rfloor} \sum_{i=b-\lfloor \frac{b}{2} \rfloor+1}^{b} \bar{\bar{Y}}_i,$

   $\hat{\sigma} = \sqrt{\frac{1}{\lfloor \frac{b}{2} \rfloor} \sum_{i=b-\lfloor \frac{b}{2} \rfloor+1}^{b} s_i^2}$ where s_i^2 is the variance for each batch mean.

   $CL = \hat{\mu} \pm z \frac{\hat{\sigma}}{\sqrt{5}}, for\ z = 1, 2, 3.$

4. Identify the truncation point and remove the initialization bias: Plot $\bar{\bar{Y}}_j$, j=1, …, b on the control chart. The truncation point is identified as the largest j, j=1, …, $\left\lfloor \frac{b}{2} \right\rfloor$ such that $\bar{\bar{Y}}_j$ is outside the control limits. To determine out-of-control signal, Robinson also applied the Western Electric rules (Montgomery 1994) plus a trace-back rule considering "the initial points all located in one side of the center line are considered as out of control". If the truncation point is located on the second half of data series, this estimation is considered as ineffective and should be dropped.

2.2 Proposed Method and Evaluation Criteria

2.2.1 Our Proposed Method

In this section, we propose a heuristic rule for initialization bias elimination by adapting the idea of the cusum change detection algorithm described in Section 2.1. We refer to this new heuristic rule as the *cusum-rule*. The cusum calculated in (2.2) requires the assumption of normal distribution, thus we use the batch-size method which is a classical method to normalize data at the beginning. For raw data $\{Y_1^{raw}, Y_2^{raw}, Y_3^{raw} \dots, Y_n^{raw}\}$ and batch size equal to m, batched data are:

$$\left\{Y_i, i = 1, 2, \dots, b | Y_i = \frac{1}{m}\sum_{j=1}^{m} Y_{(i-1)*m+j}^{raw}, b = \left\lfloor \frac{n}{m} \right\rfloor \right\}.$$

The cusum calculated in (2.2) also requires the in-control mean μ_0. We propose to estimate μ_0 by

$$\hat{\mu}_0 = \frac{1}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0} Y_t,$$

where $\mathcal{T}_0 \subset \{1, \dots, b\}$ denotes the set of indices for the data that are considered as from the steady-state. Usually $\mathcal{T}_0$ is set as

$$\left\{\left\lfloor \frac{b}{2} \right\rfloor + 1, \left\lfloor \frac{b}{2} \right\rfloor + 2, \dots, b\right\},$$

which is the second-half of the data series. For the reference value *K* in (2.2), we set $K = k\hat{\sigma}$, where $\hat{\sigma}$ is the estimated standard deviation of the data based on $\{Y_i, i \in \mathcal{T}_0\}$ and *k* is treated as a parameter of our proposed procedure.

Another assumption the cusum method in Section 2.1 requires is that we know the directions of the mean shift, that is, either $\mu_0 < \mu_1$ or $\mu_0 > \mu_1$. In initialization bias elimination, the direction of the mean shifts during the initial transient may not be known. More importantly, the mean of the transient data may oscillate above and below the steady-state mean in some simulation applications. To make the cusum method applicable to all these situations, we propose to first transform Y_i by

$$\tilde{Y}_i = |Y_i - \hat{\mu}_0|, \text{ i=1, 2, …, } b.$$

We will then apply the cusum method to detect changes in the transformed data series $\{\tilde{Y}_i, i = 1, 2, \ldots b\}$ which are the absolute deviations of the original data from the estimated steady-state mean. When X_i is from the steady-state, we expect $E[\tilde{Y}_i]$ to be unchanged and small. When Y_i is from the transient-state, we expect $E[\tilde{Y}_i]$ to be increased. Therefore, the cusum in Section 2.1 can be used to detect the mean-shift change in $\tilde{Y}_i$ from the steady-state to the transient state. The idea of using absolute deviations has been used in SPC literature (Montgomery 2004) in order to detection both mean and variance changes of a process.

To eliminate initialization bias of data series $\{Y_1, \ldots, Y_b\}$, the procedure to design the cusum-rule is:

Procedure 2 (cusum-rule):

1. Estimate in-control mean for raw data based on preset $\mathcal{T}_0$

$$\hat{\mu}_{00} = \frac{1}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0} Y_t.$$

$|\cdot|$ calculates the number of elements of a set. $\mathcal{T}_0$ is set as $\left\{\left\lfloor\frac{b}{2}\right\rfloor + 1, \left\lfloor\frac{b}{2}\right\rfloor + 2, \ldots, b\right\}$.

2. Transform batched data to the absolution deviation from $\hat{\mu}_{00}$: $\tilde{Y} = |\bar{Y} - \hat{\mu}_{00}|$

3. Determine the new steady-state parameter for updated data series W.Sample mean, $\hat{\mu}_0$, is calculated based on $\{\tilde{Y}_t, t \in \mathcal{T}_0\}$,

$$\hat{\mu}_0 = \frac{1}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0} \tilde{Y}_t.$$

To estimate the value of variance $\hat{\sigma}^2$, a batch-mean estimator V_{BM,b_1} (Goldsman, Schruben et al. 1994) is applied:

a. Divide data series $\{\tilde{Y}_t, t \in \mathcal{T}_0\}$ into b_1 non-overlapping batches. Here, we assume $|\mathcal{T}_0|$ can be divided by b_1. (Otherwise, throw away the first few in-control data (those closest to out-of-control state) to make the number of remaining data divisible by b_1) Thus, batch size $m_1 = |\mathcal{T}_0|/b_1$.

b. Batch mean series for $\{\tilde{Y}_t, t \in \mathcal{T}_0\}$ is calculated as:

$\bar{Y}_{i,m_1} = \frac{1}{m_1}\sum_{l=1}^{m_1} \tilde{Y}_{b-\left\lfloor\frac{b}{2}\right\rfloor+(i-1)m_1+l}, i = 1, 2, \dots b_1.$

c. Calculate batch-mean estimator of σ,

$\hat{\sigma}^2 = V_{BM,b_1} = \frac{m_1\sum_{i=1}^{b_1}(\bar{Y}_{i,m_1}-\bar{\bar{Y}}_{m_1})}{b_1-1}$, where $\bar{\bar{Y}}_{m_1} = \frac{1}{b_1}\sum_{i=1}^{b_1}\bar{Y}_{i,m_1}$.

4. Determine the truncation point:
Let $S_t = \sum_{i=1}^{t}(\tilde{Y}_i - \hat{\mu}_0 - K), t = 1, \dots, b$, $K = k * \hat{\sigma}$. If $\max_{1\le t\le b} S_t > 0$

$$\hat{\tau} = \underset{1\le t\le b}{\operatorname{argmax}} S_t \qquad (2.3)$$

If $\max_{1\le t\le b} S_t \le 0,\ \hat{\tau} = 0$

The selection of the parameter k will be introduced later. The cusum-rule is computationally simple, with complexity of $O(n)$.

Remark: Another initialization bias elimination method is cusum plot (Barton and Schruben 1989) which also employs "cumulative sum" to remove initialization bias. Different from cusum method in Section 2.1 and our method, cusum plot is monitoring the sum of deviations of observations from cumulative sample mean. Given $S_0 = 0$, and $S_i = \sum_{t=1}^{i}(\bar{Y} - Y_t), i = 1, \dots, n$, where $\bar{Y} = \sum_{t=1}^{n} Y_t$, cusum plot is a graph of $S_i$vs i. If there is initialization bias, plot of S_i will all tend to stay on one side of zero; otherwise, plot of S_i will tend to cross zero several times. There is also similar discussion in the

Markov Chain Monte Carlo (MCMC) literature about using the cusum plot to monitor Markov sampler and assess the convergence performance of the Markov chain (Yu 1996; Yu and Mykland 1998). Although including the same idea "cusum", cusum plot is quite different from our cusum rule. First, cusum plot sets the reference value as the global average, while our method estimate it based on the in-control data; secondly, cusum plot identifies the change point by observing S_i crossing zero, while our method finds the change point by locating the maximum of "cusum" in (2.3). Thirdly, our method employs the absolution modification which is helpful for detecting oscillating initialization bias, while cusum plot is only sensitive for one-side bias; finally, the detection rule of cusum plot is based on subjective judgment of users. Although Brooks (1998) tried to define some diagnostic parameters, an assumption that data series are symmetrically distributed about the mean value is employed. We will show the comparison results between cusum plot and our method in Section 2.3.

2.2.2 Performance Criteria

To assess and compare the effectiveness of different initialization bias elimination methods, we will use two performance criteria in this chapter. The first one is mean square error (MSE) of the estimated steady-state mean, which is defined as:

$$mse(\hat{\mu}, \mu) = E[(\hat{\mu} - \mu)^2] = bias^2(\hat{\mu}, \mu) + var(\hat{\mu}),$$

where μ represents the true steady-state mean and $\hat{\mu}$ is estimated steady-state mean using an IB elimination procedure. It is pointed out that initialization bias elimination problem is a straight-forward point estimation of μ (Fishman 1972; Snell and Schruben 1985; Pasupathy and Schmeiser 2010). MSE, considering the bias and standard error at the same time, is a statistical traditional measurement of point-estimator quality. Thus, we choose MSE as the main criterion to evaluate performance of methods in this chapter.

Another criterion we will use is *percentage bias removed by truncation* (PERC) (Hoad, Robinson et al. 2010), which is defined as:

$$PERC = \begin{cases} \frac{\sum_{i=1}^{\hat{L}} b_i}{\sum_{i=1}^{L} b_i}, \text{for } \hat{L} \leq L \\ 1, \quad \text{for } \hat{L} > L \end{cases} \tag{2.4}$$

PERC can be applied when we design the data series by plus steady-state data series with bias series which is obtained from bias function. In (2.4), b_i is the absolute value of bias function at the i^{th} observation, L is the true truncation point and $\hat{L}$ is the estimated truncation point. PERC gives a clear and quantitative idea that how much initialization bias a method can remove.

2.2.3 Selection of Parameter *k*

In order to know how the parameter *k* should be selected, first consider the case that there is no transient period and all the data are from the steady-state. A data series without transient period is referred to as the *steady-state data* in this thesis. And the MSE of a method applied to the steady-state data is referred to as *the steady-state MSE (SSMSE)*. For steady-state data, the optimal truncation point that minimizes MSE will be at zero, that is, all the data are retained to estimate the steady-state mean. Let $SSMSE_0$ denote the SSMSE when all the data are retained. Typically, the steady-state MSE of an IB elimination procedure is greater than $SSMSE_0$.

Now we look at how the parameter *k* in the cusum-rule affects its SSMSE. For any procedure, it is easy to see that its SSMSE is simply the variance of the mean estimator $\hat{\mu}$. When *k* is very large (approaching to infinity), it can be seen that $S_t \leq 0, \forall t \geq 1$ in Procedure 2. So we always have $\hat{\tau} = 0$ and the SSMSE of the cusum-rule will be exactly MSE_0. On the other hand, when *k* is very small (close to 0), S_t in Procedure 2 approximately forms a random walk (for independent data) and $\hat{\tau}$ will be randomly distributed between 0 and *b*. In this case, many steady-state data may be truncated and the variance of $\hat{\mu}$ is large. Consequently, the SSMSE when *k* is very small will be much larger than $SSMSE_0$. Although a very large *k* is optimal for the steady-state

data, it does not work well when IB exists. It can be seen that, when *k* is very large, we still have $\hat{\tau} = 0$ when initialization bias exists, which will cause large bias and MSE because no data with IB is truncated. From the above analysis, we can see that there is obviously a trade-off between the variance caused by truncating steady-data and the bias caused by using data with IB. If *k* is too large, the cusum-rule is not sensitive to the mean shifts in the IB period, which causes large bias and MSE. If *k* is too small, the cusum -rule tends to truncate many steady-state data, which causes large variance and MSE.

This tradeoff is similar to the tradeoff between the type I error (also called false-alarm rate) and detectability of mean shifts in selecting the control limit parameter of a control chart: narrow control limits cause large type I error while wide control limits reduce sensitivity to the mean shifts. In SPC literature, the control limit and other control chart parameters are selected by setting the type-I error to be within a specified value α. And two control charts are compared by first setting their parameters so that the type-I errors are close to each other. Then the detectability of the control charts for various types of mean shifts will be compared. In this chapter, we propose using the similar idea to select parameter *k* for the cusum-rule. More specifically, we will select *k* so that the SSMSE of the cusum-rule is less than or equal to $(1 + \alpha)\text{MSE}_0$, where $\alpha > 0$ and $1 + \alpha$ is a user-specified upper limit on the ratio of the MSE comparing with MSE_0. Examples of choices of α are 0.05 or 0.10. When we compare the cusum-rule with another method, *k* will be selected so that the steady-state MSE of the two methods are close to each other. The steady-state MSE of an IB procedure depends on the distribution and the autocorrelation structure of the data series. We propose using the simplest steady-state case, where the data are i.i.d. normal random variables.

In this chapter, we will focus our comparison between the cusum-rule and MSER. Since MSER does not have any parameter that can be used to adjust the SSMSE, we will select *k* so that the SSMSE of the cusum-rule is close to that of MSER. The procedure to estimate the SSMSE of an IB procedure is outlined in Procedure 3.

Procedure 3

1. Let *N* be the number of replications and *r* be replication index. Set *r*=1.
2. Generate *b* independent random observations $Y_r = \{Y_{1r}, Y_{2r}, \dots, Y_{br}\}$ with $Y_{ir} \sim N(0,1), i = 1,2, \dots, b$.
3. Apply IB procedure to Y_r. Save the estimated steady state mean $\hat{\mu}_r$. And let *r=r+1* and repeat 2 and 3 until *r=N*.
4. Estimate MSE based on $\hat{\mu}_r, r = 1, \dots, N$.

In the above procedure, we generate observations Y_{ir} from the standard normal distribution $N(0,1)$ because it is easy to see that the SSMSE of both the cusum-rule and MSER are independent of the mean and variance of Y_{ir}. Typically the SSMSE of an IB procedure varies with different values of *b*, the length of data, which is cusum-rule is the number of batches. In Table 2.1, we list the SSMSE of MSER and $SSMSE_0$ for different values *b*.

Table 2.1 SSMSE($\times 10^{-3}$) of MSER and $SSMSE_0$ and the ratio between them for different data length *N*

b	$SSMSE_m$	$SSMSE_0$	$SSMSE_m/SSMSE_0$
250	4.31	3.99	1.08
500	2.06	2	1.03
750	1.36	1.34	1.02
1000	1.02	1	1.01
1500	0.674	0.668	1.01
2000	0.504	0.501	1.01
3000	0.334	0.333	1
4000	0.251	0.250	1
5000	0.2	0.2	1
10000	0.1	0.1	1

In order to select *k* in the cusum-rule so that its SSMSE is close to that of MSER, we use the following procedure:

Procedure 4

1. k=0
2. Using Procedure 3 with N replications, estimate SSMSE of the cusum-rule, denoted by $SSMSE_c$, and the SSMSE of MSER, denoted by $SSMSE_m$. If $(SSMSE_c - SSMSE_m)/SSMSE_m \leq 0.01$, stop. Otherwise, set $k = k + \delta$, where δ is the step size used to adjust k. Repeat 2.

Using N=250,000 and δ=0.5, the k values for different values of b for the cusum-rule are given in Table 2.2.

Table 2.2 Selection of k value for different length b and corresponding SSMSE($\times 10^{-3}$)

b	k	$SSMSE_c$	$SSMSE_m$	$\frac{SSMSE_c - SSMSE_m}{SSMSE_m}$
250	0.95	4.34	4.31	0.00696
500	0.85	2.07	2.06	0.00485
750	0.75	1.37	1.36	0.00735
1000	0.70	1.03	1.02	0.0098
1500	0.60	0.680	0.674	0.0089
2000	0.55	0.508	0.504	0.00794
3000	0.45	0.337	0.334	0.009
4000	0.40	0.253	0.251	0.00797
5000	0.35	0.202	0.2	0.01
10000	0.25	0.101	0.1	0.01

Although the k value for the cusum-rule are selected based on i.i.d normally distributed data, in Section 2.3, we will conduct extensive experiments for data with various types of autocorrelation structure and distributions to compare both the steady-state and IB elimination performance of the cusum-rule and MSER.

2.3.Experiments

In this section, we will conduct extensive tests to evaluate the performance of the cusum-rule and compare it with other methods, particularly with MSER-5, the well-accepted benchmark method for IB elimination. We first use a large variety of artificial data sets similar to those used by Hoad, Robinson et al. (2010) to test and compare the methods. Then we consider ARMA time series data with IB caused by initial values. Finally, we apply the cusum-rule to the simulation outputs from M/M/1 model. In all the experiments, we set the k value of the cusum-rule according to Table 2.2 and use batch size m=5 to be consistent with the batch size of MSER-5. We use N=10,000 replications to estimate MSE and PERC for all experiments.

2.3.1 Experiments Using Artificial Data Sets

To test the performance of MSER-5, Hoad, Robinson et al. (2010) uses artificial data sets that are generated using various bias functions with different lengths, severity, shape, and orientation, as well as steady-state data with different distributions and auto-correlation structures. The selection of bias functions and steady-state data is based on their study of over 50 ‘real’ simulation models and outputs (Hoad, Robinson et al. 2010). In this chapter, we use a representative subset of the data set used by Hoad, Robinson et al. (2010). While all the data sets used by Hoad, Robinson et al. (2010) has data length n=1000, we also conduct experiments with shorter (n=250) and longer (n=4000) data lengths.

2.3.1.1 Generation of Artificial Data Sets

By testing methods with artificial data sets, we can completely control the data length n, the true truncation point L, the true steady-state mean μ, the severity of initialization bias C, the distribution of noise, and the shape of the bias function. This is helpful to make tests more adequate and complete. In this chapter, we generate steady-state data and initial bias separately. The steady-state series is generated according to the

Table 2.3 Model and parameter setting for steady-state series

Auto-correlation type	Model	Parameter setting
No auto correlation	$X_t = \varepsilon_t$	
AR(1)	$X_t = \phi_1 X_{t-1} + \varepsilon_t$	$\phi_1 = 0.9$
AR(2)	$X_t = \phi_2 X_{t-1} + \phi_3 X_{t-2} + \varepsilon_t$	$\phi_2 = -0.25, \phi_3 = 0.5$
ARMA(5)	$X_t = 1 + \varepsilon_t + \sum_{i=1}^{5} [\frac{1}{2^i}(X_{t-i} + \varepsilon_{t-i})]$	

Table 2.4 Initialization bias functions

1. Mean Shift (MS)	$\mathrm{a(t)} = \begin{cases} \pm Ct & t = 1, \dots, L \\ 0 & t = L+1, \dots, n \end{cases}$
2. Linear trend (LI)	$a(t) = \begin{cases} \pm C(1 + \frac{1}{L} - \frac{t}{L}) & t = 1, \dots, L \\ 0 & t = L+1, \dots, n \end{cases}$
3. Quadratic trend (QU)	$a(t) = \begin{cases} \pm \frac{C}{L^2}(L+1-t)^2 & t = 1, \dots, L \\ 0 & t = L+1, \dots, n \end{cases}$
4. Exponential trend (EX)	$a(t) = \begin{cases} \pm C * \exp(\frac{\log\left(\frac{C}{0.005}\right)}{L-1}(1-t)) & t = 1, \dots, L \\ 0 & t = L+1, \dots, n \end{cases}$
5. Oscillating trend (OS)	$a(t) = \begin{cases} \pm C * \exp\left(\frac{\log(0.005)}{L}\left(t - \frac{L}{20}\right)\right) * \sin(\frac{10\pi}{L}t) & t = 1, \dots, L \\ 0 & t = L+1, \dots, n \end{cases}$

models in Table 2.3. Different extents of auto-correlations are considered. Random noise ε_t is set to be distributed as either N(1, 1) or Exp(1). For AR(1), AR(2) and ARMA(5) model, the initial values of $\{X_0, X_{-1}, X_{-2}, \dots\}$ are set as the true steady-state mean of each model. In order to ensure that the data are sampled from the steady-state distributions, we treat the first 1000 data points as the warm-up period, which will be truncated, and select the next *n* data points as the steady-state series we really use.

The initialization bias functions are generated based on the bias severity parameter *C*(=10,5,2,1), the data length *n*, and the true truncation point *L*(=0.1*n*, 0.2*n*, 0.3*n*, 0.4*n*) as shown in Table 2.4. We add negative bias when the data length *n* = 250 and 4000, and positive bias when *n*=1000. After generating initialization bias series $a(t)$ and steady-state series X_t, the data series with IB is generated by $\{Y_t = X(t) + a(t), t = 1, \dots, n\}$.

2.3.1.2 Preliminary Testing Results

We first consider simple data sets with no autocorrelation, normal errors, data length *n*=1000, and with a mean shift bias or linear trend bias. These data sets will be used to quickly test the four methods mentioned in this chapter: MSER, cusum-rule, SPC, and cusum-plot. We use *C*=10 so that the bias is severe and obvious. Thus a method failing to give a good estimation of steady-state mean for these simple data sets should be rejected. The testing results are showed in Table 2.5. Also, to make the comparison be consistent with the setting in Robinson (2007), each data sample used in each replication is actually the average of 5 replications of raw data series. In the first row of Table 2.5, we show the SSMSE (*L*=0) for all four methods. MSER has the smallest SSMSE and cusum-rule has a very similar one, while SPC and cusum-plot have much larger SSMSE than MSER and the cusum-rule. The other rows of Table 2.5 compare the MSEs of the four methods for data sets with different initial bias lengths and trends. The MSE of MSER-5 and cusum-rule are very close to each other in all these cases. Comparing with

MSER and the cusum-rule, SPC method has much larger MSE in many cases, especially when L=100 or 200. Another issue worth to mention is that SPC method has less than 5% effective records because of its severe false alarm. The idea of effective record is proposed by Robinson (2007) that the estimated truncation point using the SPC method is within the first half of raw data, while an estimated truncation point within the second half is considered as ineffective and additional samples should be collected to estimate the truncation point again. The low percentage of effective records using the SPC method causes substantial waste of data and computational cost.

For the cusum-plot, we use the data point where the plot first crosses zero as the truncation point. From Table 2.5, it is obvious that the cusum-plot performs poorly in detecting the truncation point. For most cases, the cusum-plot never crosses zero (therefore the percentage of effective records is zero). The problem of the cusum-plot method can be clearly seen from Figure 2.1 which shows the cusum plot of S_i vs. i for data with linear trends and true truncation point L equal to 100 and 400. We can find that in both types of data, the cusum plot is entirely on one side of zero and it reaches a value close to zero only at the beginning and the end of the data series, which is far from the true location of the truncation point. Another useful observation of Figure 2.1 is that the maximum of the cusum plot seems to be much more related to the location of the truncation point. And the cusum-rule proposed in this chapter is exactly based on the idea of using maximum of the cusum to estimate the truncation point. As discussed in Section 2.2.1, another important difference between the cusum-rule and the cusum-plot is that the reference values are calculated differently. To see that the reference value used by the cusum-plot is not effective, we revise the cusum-plot method by using the location of the maximum of S_j value as the truncation point, which is similar to the detection rule used in the cusum-rule. We refer to this revised (and improved) cusum-plot method as cusum-plot-rev and compare its performance with other methods in Table 2.5. From these results, we can see that, although cusum-plot-rev is much better than the original cusum-

plot, its performance is quite unstable, with very large MSE for the last three cases of linear trend data. This is because cusum-plot-rev, as in the original cusum-plot, still uses the global mean of the data as the reference value when calculating the cumulative sum S_j, which causes the shape of the cusum-plot to be highly sensitive to the shape, length, and magnitude of the initialization bias. Since MSER-5 and Cusum-rule are the only methods under comparison that perform well for these simple cases, we will focus on only the comparison of these two methods in the following sections.

Table 2.5 MSE($\times 10^{-3}$) of five methods for Mean Shift and Linear trend data with *C*=10 and $\varepsilon_t \sim \mathrm{N}(1,1)$

		MSER-5	Cusum-rule	SPC	Cusum-plot	Cusum-plot-rev
No	*L*=0	0.2	0.2	0.31(4.42%)	0.256(98.9%)	2.41
MS	*L*=100	0.23	0.23	0.32(4.55%)	NaN (0%)	0.22
	L=200	0.25	0.26	0.32(4.66%)	NaN(0%)	0.25
	L=300	0.29	0.29	0.33(4.40%)	NaN(0%)	0.29
	L=400	0.34	0.34	0.36(4.37%)	NaN(0%)	0.33
LI	*L*=100	0.22	0.22	0.31(4.51%)	322(1.12%)	0.22
	L=200	0.26	0.26	0.34(4.37%)	NaN(0%)	0.41
	L=300	0.29	0.29	0.30(4.34%)	NaN(0%)	2.4
	L=400	0.34	0.35	0.33(4.42%)	NaN(0%)	14

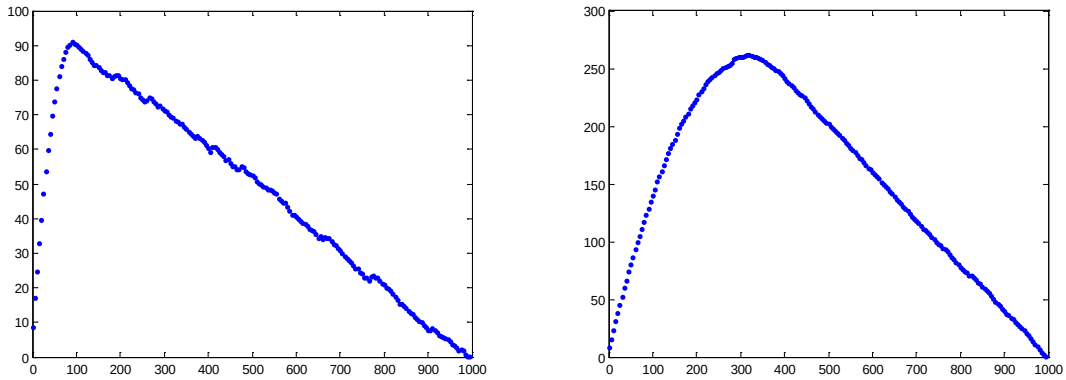

Figure 2.1 The cusum-plots for data series with linear trend, n=1000 and L=100 (left) or 400 (right)

2.3.1.3 Comparison Between the Cusum-rule and MSER-5 for artificial data sets

We conduct a total of 960 experiments to compare the performance of the cusum-rule and MSER-5 for data series with a full combination of different parameter settings as shown in Table 2.6.

Table 2.6 Parameters used to generate the artificial data sets

Parameter	Notation	Values
Noise distribution	ε_t	N(1, 1); Exp(1)
Bias severity value	C	10; 5; 2; 1
Date length	n	250; 1000; 4000
True truncation point	L	0.1n; 0.4n
Autocorrelation Type	AR	0-no autocorrelation; 1-AR(1); 2-AR(2); 5-
Shape of the bias	Shape	MS; LI; QU; EX; OS

Table 2.7 Cases with maximum $\frac{MSE_c}{MSE_m}$

n	maxCM	Error	*C*	*L*	AR	Shape	MSE_m	MSE_c	$PERC_m$	$PERC_k$
250	1.07	N(1, 1)	1	25	1	EX	0.494	0.527	42%	41%
1000	1.04	N(1, 1)	5	100	1	MS	0.119	0.123	96%	93%
4000	1.06	N(1, 1)	10	400	5	MS	1.31	1.39	79%	69%

Table 2.8 Cases with maximum $\frac{MSE_m}{MSE_c}$

n	maxMC	Error	*C*	*L*	AR	shape	MSE_m	MSE_c	$PERC_m$	$PERC_c$
250	2.2	Exp(1)	1	100	2	MS	0.0358	0.0163	79%	93%
1000	8.13	Exp(1)	1	400	2	MS	0.0373	0.00458	77%	98%
4000	7.09	N(1, 1)	1	1600	2	MS	0.00562	0.000793	97%	100%

Table 2.9 All records with MSER and cusum-rule having more than 10% difference in MSE when *n*=250

Error	*C*	*n*	*L*	AR	shape	MSE_m	MSE_c	$PERC_m$	$PERC_c$
N(1,1)	1	250	100	0	MS	0.0178	0.00885	91%	97%
N(1,1)	1	250	100	0	LI	0.0143	0.0121	60%	66%
N(1,1)	5	250	100	1	MS	1.04	0.876	86%	91%
N(1,1)	1	250	100	2	MS	0.046	0.0219	73%	89%
N(1,1)	1	250	100	2	LI	0.0218	0.0193	48%	55%
Exp(1)	1	250	100	0	MS	0.0123	0.00747	94%	98%
Exp(1)	1	250	100	0	LI	0.0124	0.0101	64%	71%
Exp(1)	5	250	100	1	MS	0.86	0.771	90%	92%
Exp(1)	1	250	100	2	MS	0.0358	0.0163	78%	93%
Exp(1)	1	250	100	2	LI	0.0202	0.017	53%	61%

In Table 2.7 and Table 2.8, we first show the cases where the MSE of the two methods are the most different. Table 2.7 lists the data series that gives the largest ratio of MSE of the cusum-rule (MSE_c) over that of the MSER-5 (MSE_m) for each *n* value, where

maxCM=$max\frac{MSE_c}{MSE_m}$. Table 2.8 lists the data series that gives the largest ratio of MSE_m over MSE_c, where maxMC= $max\frac{MSE_m}{MSE_c}$.

From Table 2.7, we can find cusum-rule is never more than 7% worse than MSER in terms of MSE based on all the experiments we have performed. However, MSER could be more than eight times worse than cusum-rule in some cases according to Table 2.8. To explore the difference of behaviors between MSER and cusum-rule, we conduct further study on all the records with more than 10% difference in MSE, including 10 (3.1%) records when data length *n*=250, 28(8.8%) records when *n*=1000 and 90 (28%) records when *n*=4000. When the data length is longer, there tends to be bigger difference in performance between the two methods. MSER has a bigger MSE and eliminates smaller percentage of initialization bias than the cusum-rule in all 128 cases. This indicates that both performance criteria support the claim that the cusum-rule is significantly better than the MSER for these records.

We show all the records with more than 10% difference in MSE between the two methods when *n*=250 in Table 2.9 and when *n*=1000 in Table 2.10 and Table 2.11. To save space and show the results more clearly, the 90 records with more than 10% difference in MSE when *n*=4000 are showed in Table 2.9, Table 2.10 and Table 2.11 grouped by the values of AR, *L*, Error distribution, and the shape of initialization bias, respectively.

In all three tables, the first six columns show the parameter setting of each record: the noise distribution of each record, the bias severity value *C*, the data length *n*, the true truncation point *L*, the autocorrelation variable *AR*, and the shape of initialization bias. Then in the last four columns, the MSE value and PERC value of two methods are shown. From these three tables, we can find *L*=0.4*n* in all records which indicates that cusum-rule tends to behave substantially better than MSER when the percentage of initialization bias is bigger; severity value *C* is 1, 2 or 5 which indicates when the severity of bias is smaller, cusum-rule has more chance to have significantly better performance

than MSER-5; autocorrelation AR is 0, 1 or 2 which indicates we cannot find significant difference between the two methods in ARMA(5, 5) data; for the shape of initialization bias, only mean shift and linear trend are involved when *n*=250, and then exponential and quadratic trends show up in the table when *n*=1000. No oscillating data shows significant difference between the two methods.

Table 2.10 All records with MSER and cusum-rule having more than 10% difference in MSE with *n*=1000 and N(1,1)

Error	*C*	*n*	*L*	AR	shape	MSE_m	MSE_c	$PERC_m$	$PERC_c$
N(1,1)	5	1000	400	0	QU	0.00253	0.00226	96%	97%
N(1,1)	5	1000	400	0	EX	0.00228	0.00206	90%	91%
N(1,1)	2	1000	400	0	LI	0.00368	0.00292	92%	94%
N(1,1)	2	1000	400	0	QU	0.00355	0.00284	85%	88%
N(1,1)	2	1000	400	0	EX	0.00247	0.00219	74%	77%
N(1,1)	1	1000	400	0	MS	0.0025	0.0018	99%	100%
N(1,1)	1	1000	400	0	LI	0.0096	0.0051	61%	76%
N(1,1)	1	1000	400	0	QU	0.0050	0.0036	57%	66%
N(1,1)	1	1000	400	0	EX	0.0027	0.0024	47%	53%
N(1,1)	5	1000	400	1	MS	0.2933	0.2089	96%	98%
N(1,1)	5	1000	400	1	LI	0.3258	0.2884	61%	66%
N(1,1)	2	1000	400	1	MS	0.6082	0.4999	14%	29%
N(1,1)	5	1000	400	2	QU	0.0043	0.0038	95%	96%
N(1,1)	2	1000	400	2	LI	0.0070	0.0052	88%	91%
N(1,1)	2	1000	400	2	QU	0.0060	0.0048	80%	84%
N(1,1)	1	1000	400	2	MS	0.0276	0.0036	83%	99%

When data length *n*=4000, the percentage of data records with at least 10% difference in MSE grows to 28% of the total number of experiments. The cusum-rule is much more robust than MSER-5 when *n*=4000, which is shown in Figure 2.2, Figure 2.3 and Figure 2.4 shows the boxplots of the ratios $R_{m/c} = \frac{MSE_m}{MSE_c}$ for difference autocorrelation settings. The median lines of $R_{m/c}$ are all either higher than or close to 1.

These figures clearly show that for all cases the performance of the cusum-rule is similar to or better than that of MSER-5. However, for many cases, with the only exceptions of AR=5 and Shape=OS, the MSER-5 is much worse than the cusum-rule for many data sets, which indicates again that the cusum-rule is more robust than MSER-5, which is consistent with what we find when *n*=250 and 1000.

Table 2.11 All records with MSER and cusum-rule having more than 10% difference in MSE with *n*=1000 and Exp(1)

Error	*C*	*n*	*L*	AR	shape	MSE_m	MSE_c	$PERC_m$	$PERC_c$
Exp(1)	2	1000	400	0	LI	0.00402	0.00333	92%	93%
Exp(1)	2	1000	400	0	QU	0.00371	0.00323	85%	87%
Exp(1)	1	1000	400	0	MS	0.00705	0.0018	96%	100%
Exp(1)	1	1000	400	0	LI	0.0101	0.00627	61%	72%
Exp(1)	1	1000	400	0	QU	0.00535	0.00436	57%	62%
Exp(1)	5	1000	400	1	MS	0.403	0.241	93%	97%
Exp(1)	2	1000	400	1	MS	0.581	0.519	19%	27%
Exp(1)	2	1000	400	2	LI	0.00771	0.00601	87%	90%
Exp(1)	2	1000	400	2	QU	0.00626	0.00544	80%	83%
Exp(1)	1	1000	400	2	MS	0.0373	0.00458	77%	98%
Exp(1)	1	1000	400	2	LI	0.0164	0.0108	47%	60%
Exp(1)	1	1000	400	2	QU	0.00798	0.00687	45%	52%

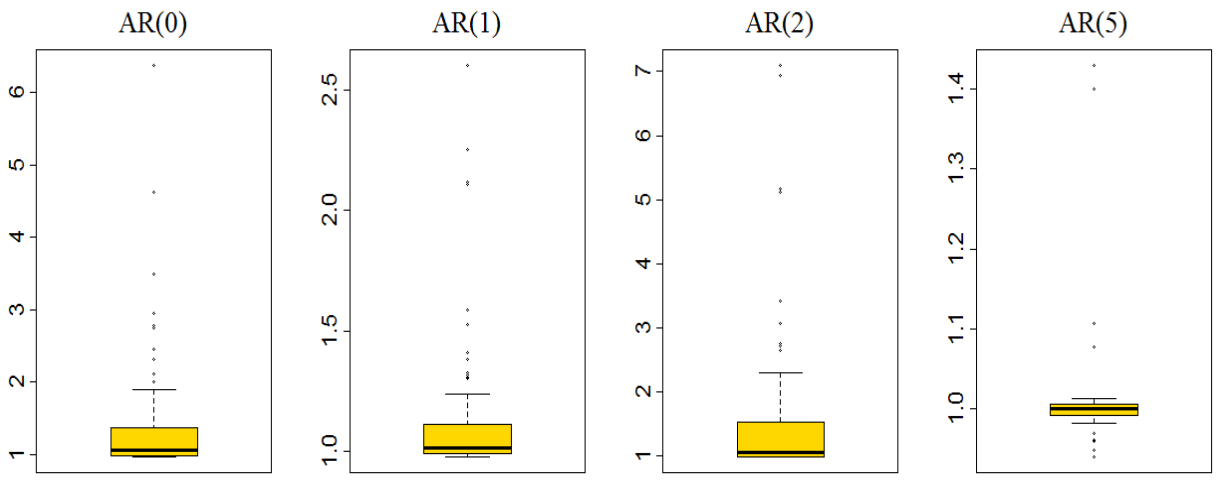

Figure 2.2 Boxplots of MSE ratio between MSER and cusum-rule grouped by AR when n=4000

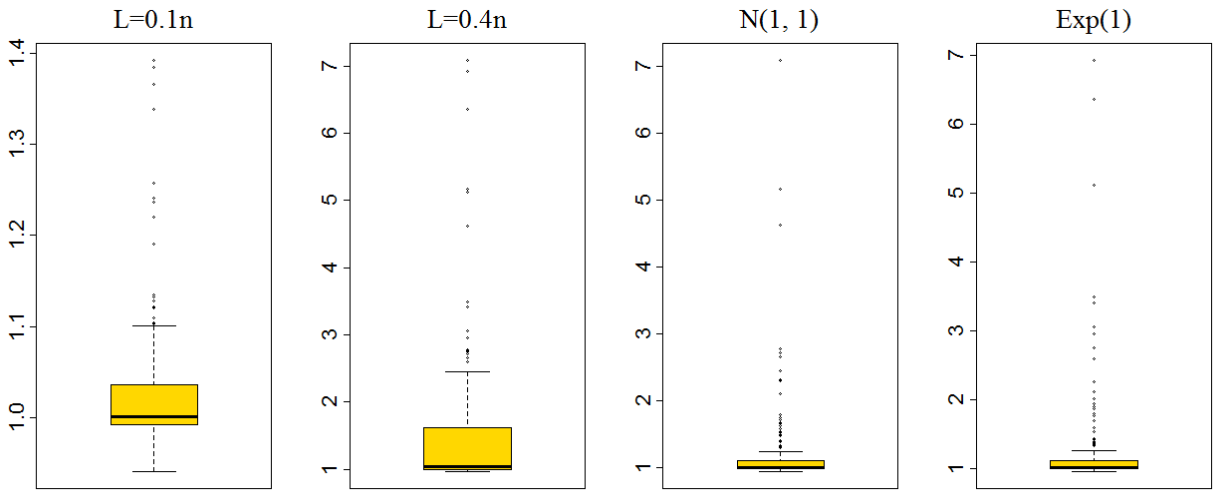

Figure 2.3 Boxplots of MSE ratio between MSER and cusum-rule grouped by L or noise distribution when n=4000

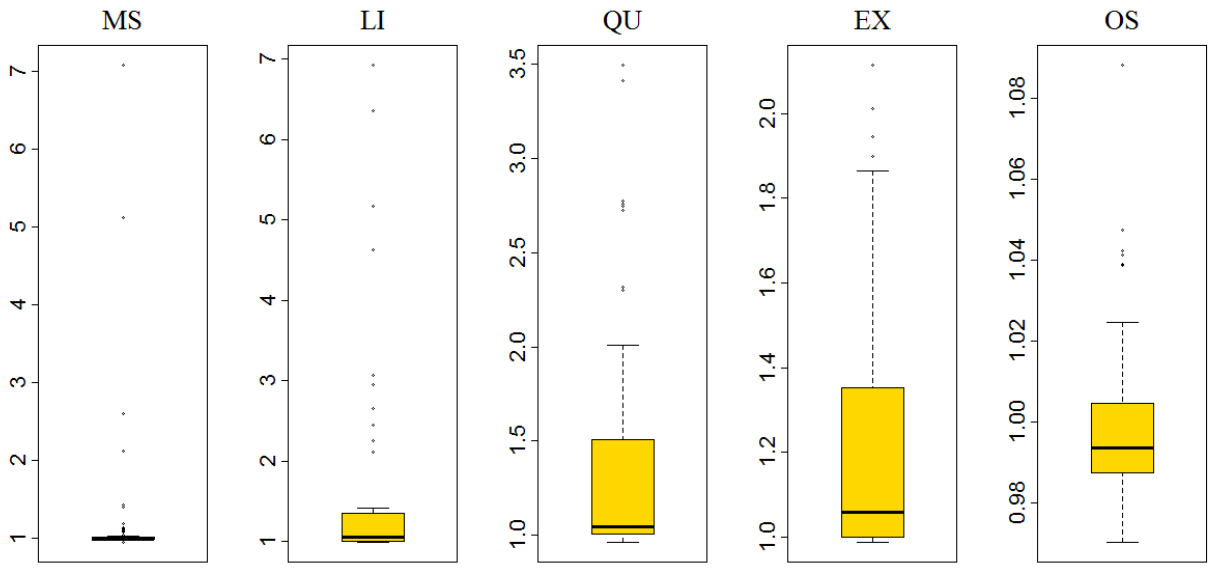

Figure 2.4 Boxplots of MSE ratio between MSER and cusum-rule grouped by bias shape when n=4000

2.3.2 ARMA Model with *L*=0 and Initialization Bias Caused by Different Initial Values

In this section, we will show the comparisons of performance between the cusum-rule and MSER on auto-correlated data with initialization bias caused by different initial values. First, we show the SSMSE of the two methods in Table 2.12. It can be seen that the SSMSE of the cusum rule is still very close to that of the MSER-5 auto-correlated data and non-normal (exponential) noise distribution, even though the parameter *k* in the cusum-rule is selected based on the independent data with normal noise distribution. This result further justifies the procedure we propose for the selection of the parameter *k* for the cusum-rule.

Table 2.13 shows the performance of MSER and cusum-rule for difference error distributions, severity of initialization bias caused by initial values, and auto-correlation

structures. In addition to report the MSE value, we also report the bias and standard deviation of the steady-state mean estimation. We can find that the results are very close between the two methods for AR=1 and 2. When AR=5, it appears that the MSE of the cusum-rule is significantly smaller than MSE of MSER-5.

Table 2.12 SSMSE for MSER and cusum-rule for ARMA data

	Error	N(1, 1)			
	AR	0	1	2	5
n=250	$SSMSE_m$	0.0043	0.478	0.00769	15.7
	$SSMSE_c$	0.00433	0.507	0.00785	16.6
n=1000	$SSMSE_m$	0.001	0.103	0.0018	4.87
	$SSMSE_c$	0.00101	0.105	0.00183	4.77
n=4000	$SSMSE_m$	0.00025	0.0255	0.000447	1.06
	$SSMSE_c$	0.000252	0.0259	0.000451	1.06
	Error	Exp(1)			
	AR	0	1	2	5
n=250	$SSMSE_m$	0.00451	0.476	0.00813	15.9
	$SSMSE_c$	0.00441	0.501	0.00805	17
n=1000	$SSMSE_m$	0.00102	0.11	0.00185	4.86
	$SSMSE_c$	0.00102	0.11	0.00185	4.72
n=4000	$SSMSE_m$	0.000252	0.0256	0.000442	1.07
	$SSMSE_c$	0.000255	0.0257	0.00045	1.06

Table 2.13 Results for MSER-5 and cusum-rule for ARMA data with different initial values when n=1000

Error	C	AR	MSE_m	$bias_m$	std_m	MSE_c	$bias_c$	std_c
Exp(1)	10	1	0.112	-0.00371	0.334	0.112	0.0056	0.334
Exp(1)	5	1	0.107	-0.00997	0.327	0.107	-0.00196	0.327
Exp(1)	2	1	0.105	-0.0134	0.324	0.105	-0.00579	0.325
Exp(1)	1	1	0.108	-0.0129	0.329	0.108	-0.00606	0.329
N(1, 1)	10	1	0.105	0.0124	0.323	0.106	0.0126	0.325
N(1, 1)	5	1	0.106	0.0156	0.326	0.108	0.0143	0.329
N(1, 1)	2	1	0.105	0.0125	0.323	0.106	0.0107	0.326
N(1, 1)	1	1	0.106	6.60E-05	0.326	0.108	-0.00057	0.329
Exp(1)	10	2	0.00183	-0.00212	0.0427	0.00182	-0.00162	0.0427
Exp(1)	5	2	0.00183	-0.00228	0.0427	0.00183	-0.00208	0.0427
Exp(1)	2	2	0.00185	-0.00184	0.043	0.00185	-0.00153	0.043
Exp(1)	1	2	0.00185	-0.00163	0.0429	0.00186	-0.00118	0.0431
N(1, 1)	10	2	0.0018	0.000773	0.0424	0.00182	0.000645	0.0427
N(1, 1)	5	2	0.00181	3.32E-05	0.0425	0.00182	-0.00034	0.0427
N(1, 1)	2	2	0.00179	0.000713	0.0423	0.0018	0.000486	0.0424
N(1, 1)	1	2	0.00176	0.000687	0.042	0.00178	0.000538	0.0422
Exp(1)	10	5	4.97	-0.00707	2.23	4.86	0.0923	2.2
Exp(1)	5	5	4.93	-0.0165	2.22	4.82	0.0412	2.2
Exp(1)	2	5	4.74	-0.101	2.17	4.71	-0.0522	2.17
Exp(1)	1	5	4.86	-0.115	2.2	4.79	-0.0719	2.19
N(1, 1)	10	5	5.12	0.214	2.25	5.03	0.258	2.23
N(1, 1)	5	5	4.81	0.0526	2.19	4.76	0.0631	2.18
N(1, 1)	2	5	4.78	0.00413	2.19	4.72	0.00225	2.17
N(1, 1)	1	5	4.66	0.00134	2.16	4.56	-0.00524	2.14

Table 2.14 Results of MSER-5 and cusum-rule for M/M/1 data

Traffic intensity	n	MSE_m	$bias_m$	std_m	MSE_c	$bias_c$	std_c
0.8	2000	0.105	-0.063	0.318	0.104	-0.0538	0.318
	4000	0.0525	-0.0369	0.226	0.0525	-0.0323	0.227
0.9	2000	2.05	-0.29	1.40	1.97	-0.305	1.37
	4000	0.983	-0.211	0.969	0.962	-0.196	0.961

2.3.3 M/M/1 Models

In this section, queuing system M/M/1 is applied to compare the two methods. The arrival rate is set as 0.8 and 0.9 and the service rate is set as 1. Thus, the corresponding traffic intensity is 0.8 and 0.9. The initial condition of the queue is empty. The steady state of customer's waiting time is 4 and 9. Data length *n* here is set as 2000 and 4000. Each data sample we applied methods on are mean series of five simulated series of M/M/1. The results are shown in Table 2.14. Cusum-rule still holds very close MSE with MSER and in most cases smaller than MSER. This comparison shows the cusum-rule could behave as good as MSER for data generated from basic queuing system M/M/1.

2.3.4 Discussion

From the comprehensive experiments we conduct, we can find that the cusum-rule is a much more robust method than MSER-5, the benchmark heuristic rule for IB elimination. The cusum-rule is never 10% worse than MSER-5 in terms of the MSE for the data sets we tested, while in many cases it performs much better than MSER-5, with as much as 8 times smaller MSE.

Moreover, the adjustable parameter *k* in the cusum-rule makes it possible for the method to adapt to different requirements of SSMSE, while MSER-5 cannot be adjusted to change its SSMSE. In Section 2.2.3, we have explained that, in this chapter, the *k* value is selected so that the cusum-rule has a similar SSMSE with that of MSER-5 in order to have a fair comparison with MSER-5. When we use the cusum-rule in practice, however, the *k* value does not have to be chosen exactly as Table 2.2. For example, in some cases, the users can tolerate more of the variance in the steady-state mean estimation and want to reduce the bias of the estimation further. This can be achieved by reducing the value of *k* from what is given in Table 2.2. Therefore, by adjusting the value of parameter *k* in the cusum-rule, it can achieve any desired level of trade-off between the bias and variance.

2.4 Summary

In this chapter, we propose a new heuristic method, the cusum-rule, motived by the offline change detection methods, to eliminate IB in steady-state simulation. Extensive experiments are conducted to compare the cusum-rule with the SPC method for IB elimination, the cusum plot method, and the benchmark MSER-5 method. MSE is used as the main performance criterion to compare the methods. Through all the experiments, we demonstrate that the performance of the cusum-rule is at least comparable to that of MSER-5 in all cases we have tested, and is much better in many cases, which shows that the cusum-rule is more robust than MSER-5. In addition, the parameter k value enables the cusum-rule to tune the tradeoff between the bias and variance of the steady-state mean estimation.

CHAPTER 3

FULLY SEQUENTIAL MULTI-BEST RANKING AND SELECTION FOR TERMINATING SIMULATION

3.1 Problem Framework

In this section, we formulate the multi-best ranking and selection problem. Suppose we compare *K* alternative systems through discrete event simulations. Let $X_{ij}, i = 1, \dots, K,\ j = 1, \dots, N_i$ represent the j^{th} observed simulation output of the i^{th} system, where N_i is the total number of outputs observed from system *i*. These outputs are assumed to be i.i.d. normal random variables for a fixed *i*, as they often represent either the average outputs of independent replications of simulations or approximately independent batch means from a single replication of steady-state simulation. Let $\mu_i = E[X_{ij}]$ represent the true mean of the i^{th} alternative and $\sigma_i^2 = Var[X_{ij}]$ represent its variance. Suppose the ordering of the true means of the *K* systems is as follows,

$$\mu_1 \leq \mu_2 \leq \cdots \leq \mu_K \tag{3.1}$$

This true ranking of the systems is unknown to the user. Without loss of generality, we assume that a better system has larger mean output. From (3.1), system *K* is the best system and system 1 is the worst. The goal of the *k*-best ranking and selection procedure is to select a subset $I \subseteq \{1, 2, \dots, K\}$ such that

$$\Pr\{I = \mathcal{K} | \mu_{K-k+1} - \mu_{K-k} \geq \delta\} \geq 1 - \alpha.$$

where $\mathcal{K} = \{K, K-1, \dots, K-k+1\}$ is the *k* best systems and $1-\alpha$ is the user-specified confidence level, δ is the indifference zone parameter, which is the smallest difference between the mean outputs of two systems that the experimenter feels worth detecting. In this chapter, we call the *k* best systems, that is, systems *K-k+1* to *K*, as superior systems and the other systems as inferior systems.

In this chapter, we focus on the development of a fully sequential procedure to solve the *k*-best ranking and selection problem. A fully sequential procedure consists of

multiple stages. At each stage, the procedure takes one or more basic observations from each alternative system in play. Based on the observations collected at and prior to the current stage, apparently inferior systems will be eliminated and apparently superior ones will be pre-selected. The procedure will terminate either when k systems are selected or the total number of stages has reached a pre-determined limit. In either case, the k-best systems can be selected with the level of confidence specified by the user. Comparing with procedures with small number of stages (e.g. two or three stages), a fully sequential procedure has more opportunities to eliminate inferior systems or select superior systems in early stages. Therefore, the fully sequential procedure is potentially more efficient in terms of the average number of basic samples needed per system (ANS). It is recommended when the overhead of switching among stopping and restarting alternatives to obtain additional data at each stage is low, which is expected to be the case when modern computing environment is used.

3.2 Multi-best R&S Fully Sequential Procedure

In this section, we describe a fully sequential procedure, called Procedure FS-k, in this chapter, to solve the multi-best R&S problem. This procedure is an extension of the single-best (k=1) R&S fully sequential procedure from Kim and Nelson (2001) (KN). The extension of Procedure KN is mainly based on the following two ideas: (1) In the single-best R&S problems, only the elimination of apparently inferior systems will be considered. For the multi-best R&S problems, when k>1, we consider not only elimination of inferior systems, but also the selection of apparently superior systems, which will lead to greater saving of samples. (2) We introduce a directed graph to represent the systems and their pairwise comparison results. The apparently superior and inferior systems are identified based on the number of predecessors and successors of the corresponding node in the graph.

At the beginning of our procedure, we have a graph with *m*=*K* nodes and no arcs, each node representing a system/alternative. We have $g = 0$ preselected systems. And at each stage we have m systems in contention and $k - g$ system to be selected. The basic idea to screen apparently inferior systems and preselect superior systems at each stage of the procedure is described as follows. At each stage, after collecting an additional observation from each remaining system, we first compare all the systems in pairs. If we can assure system *i* is better than system *i* at certain probability, we will add a directed arc from system *i* to *l* in the graph. Once an arc is created, it will not change direction at later stages. When all the comparisons are finished, we will calculate the number of predecessors and the number of successors for each node/system. For system *i*, its predecessors are the nodes that can reach node *i* by a directed path. For example, if we have *s*→*t*→*i* in the graph, this means that systems *s* and *t* are both predecessors of system *i*. Similarly, system *t* and system *i* are considered as the successors of system *s*. If system *i* has $k - g$ predecessors, it means that there are at least $k - g$ systems in contention that are better than system *i*. And system *i* should be eliminated from contention and node *i*, as well as all the arcs connected with *i*, need to be removed from the graph. If system *i* has $K - (k - g)$ inferior systems, it means that there are at most $k - g - 1$ systems that are better than system *i* so that system *i* can be determined as one of the $k - g$ best systems. Thus, system *i* can be preselected and removed from the graph. Also all the arcs connected with system *i* need to be removed from the graph.

The complete procedure is given as follows.

Procedures FS-*k*:

Initialization: Select confidence level $1 - \alpha$, indifference zone δ and first-stage sample size n_0. Let $I_0 = \{1, \dots, K\}$ be the set of alternative designs. Create a directed graph $V = (I_0,\ E)$, where *E* represents the set of directed edges between nodes in set I_0. A directed edge in *E* from node *i* to node *j* is denoted by $e_{ij}: i \to j, i \in I_0, j \in I_0$. Set $E = \emptyset$. Node i_1 is considered a predecessor (successor) of node i_2 if there is a directed path in *V*

from i_1 (i_2) to i_2 (i_1). Let A_i be the set of all the predecessors of node *i* and B_i the set of all the successors of node *i*. And let $G = \emptyset$ be the set of selected systems and $S = \emptyset$ be the set of eliminated systems. Let $h^2 = 2c\eta(n_0 - 1)$.

Obtain n_0 observations $X_{ij},\ j = 1, 2, \dots, n_0$, from each system *i* = 1,2,…, *K*. For all $i \neq l$, *i* = 1,2,…, *K* and *l* = 1,2,…, *K*, compute $S_{ij}^2 = \frac{1}{n_0}\sum_{j=1}^{n_0}(X_{ij} - X_{lj} - (\bar{X}_i(n_0) - \bar{X}_l(n_0)))^2$, where $\bar{X}_i(n_0) = \frac{1}{n_0}\sum_{j=1}^{n_0} X_{ij}$, i = 1,2,…, *K*.

S_{ij}^2 is the sample variance of data series $X_{ij} - X_{lj}$ where $j = 1, 2, \dots, n_0$. It's the difference between alternatives *i* and *l*. Let $N_{il} = \left\lfloor \frac{h^2 S_{il}^2}{\delta^2} \right\rfloor$, where $\lfloor \cdot \rfloor$ indicates truncation of any fractional part, and let $N_i = \max_{l \neq i} N_{il}$.

Here $N_i + 1$ is the maximum number of observations that can be taken from system *i*. If $n_0 > \max_i N_i$, then stop and select the *k* systems with the largest $\bar{X}_i(n_0)$ as the best *k* systems.

Otherwise, set $I = I_0$, where *I* represents the set of systems still in contention, and the observation counter $r = n_0$. Go to **Screening**.

Screening: Let $W_{il}(r) = \max\{0, \frac{\delta}{2cr}(\frac{h^2 S_{il}^2}{\delta^2} - r)\}$, $\forall i, l \in I, i \neq l$. Let $m = |I|$ and $g = |G|$.

For $\forall i, l \in I, i \neq l$, if $\bar{X}_i(r) - \bar{X}_l(r) \geq W_{il}(r)$ and $e_{il} \notin E, e_{li} \notin E$, then $E = E \cup e_{il}$, $A_l = A_l \cup \{i\}$ and $B_i = B_i \cup \{l\}$. Then for $\forall i, l \in I, i \neq l$, define $S_{ij}^A = \{l: l \in A_i\} \cup \{i\}$, $S_{ij}^B = \{l: l \in B_j\} \cup \{j\}$. For $\forall l \in S_{ij}^A, t \in S_{ij}^B$, let $A_t = A_t \cup S_{ij}^A \backslash \{t\}$ and $B_l = B_l \cup S_{ij}^B \backslash \{l\}$.

For $\forall i \in I$, if $|B_i| \geq m - (k - g)$, then $G = G \cup \{i\}$. If $|A_i| \geq k - g$, then $S = S \cup \{i\}$. If either of these two conditions is true, then $I = I \backslash \{i\}, A_l = A_l \backslash \{i\}$ and $B_l = B_l \backslash \{i\}\ \forall l \in I$.

Stopping Rule: If $|G| \geq k$, stop the procedure and select *k* systems in *G* that have the largest overall sample means $\bar{X}_i(r) = \sum_{j=1}^{r} X_{i,j}$ as the best *k* systems;

If $|G| + |I| \leq k$, stop the procedure and select all the systems in *G* and *I*, together with $k - (|G| + |I|)$ system(s) in *S* that have the largest overall sample means as the best *k* systems;

Otherwise, take one additional observation $X_{i,r+1}$ from each alternative $i \in I$. Set $r = r + 1$ and go to **Screening**.

Constants: The constant *c* may be nonnegative integer. As the analysis made in Kim and Nelson (2001), $c = 1$ is a good compromise solution. The constant η is the solution to the equation

$$g(\eta) \equiv \sum_{l=1}^{c}(-1)^{l+1}\left(1 - \frac{1}{2}I(l = c)\right)\left(1 + \frac{2\eta(2c-l)l}{c}\right)^{-\frac{n_0-1}{2}} = \frac{\alpha}{k(K-k)},$$

where $I(\cdot)$ is the indicator function. When $c = 1$, we have the closed-form solution

$$\eta = \frac{1}{2}\left[\left(\frac{2\alpha}{k(K-k)}\right)^{\frac{-2}{n_0-1}} - 1\right] \quad (3.2)$$

.

The following result shows that the proposed procedure has a probability of correction selection of at least $1 - \alpha$.

Theorem 3.1 Suppose that $\mu_{K-k+1} - \mu_{K-k} \geq \delta$ and the covariance matrix of the observations from the *K* systems is an unknown $K \times K$ positive definite matrix Σ. Then procedure FS-*k* selects the best *k* systems with probability of at least $1 - \alpha$.

Proof:

First we will show that the procedure always stops. When $I = \emptyset$, obviously at least one of stopping criteria is satisfied and the procedure is stopped. If the procedure has not stopped when $r = N_i + 1$, for any $i_1, i_2 \in I$ and $i_1 \neq i_2$, we have $W_{i_1 i_2}(r) = 0$ and either $e_{i_1 i_2} \in E$ or $e_{i_2 i_1} \in E$. Then for any node $i \in I$, $I\{i\} \subseteq A_i \cup B_i$ and $|A_i| + |B_i| \geq |A_i \cup B_i| \geq m - 1$. Therefore, either $|A_i| \geq k - g$ or $|B_i| \geq m - (k - g)$. So for each system/node $i \in I$, it must be in *G* or *S* (or both) after the screening step when $r = N_i + 1$. Then we have $I = \emptyset$ at the end of this iteration and the procedure will stop.

If $n_0 > \max_i N_i$, let ICS denote the event that the k best systems are not selected correctly by the procedure.

$$\Pr\{ICS\} = \Pr\left\{\bigcup_{1\le l\le K-k}\ \bigcup_{K-k+1\le i\le K} \bar{X}_l(n_0) \ge \bar{X}_i(n_0)\right\}$$

$$\le \sum_{1\le l\le K-k}\ \sum_{K-k+1\le i\le K} \Pr\{\bar{X}_l(n_0) \ge \bar{X}_i(n_0)\}$$

$$\le \sum_{1\le l\le K-k}\ \sum_{K-k+1\le i\le K}\ \sum_{l=1}^{c} (-1)^{l+1}\left(1-\frac{1}{2}I(l=c)\right)\left(1+\frac{2\eta(2c-l)l}{c}\right)^{-\frac{n_0-1}{2}}$$

$$= \alpha$$

The last inequality is based on the proof of Theorem 1 by Kim and Nelson (2001).

If $n_0 \le \max_i N_i$, define event $\mathcal{A}$ = {for all $i \in \{K, K-1, \dots, K-k+1\}$ and $l \in \{1, \dots, K-k\}$, $e_{li} \notin E$ when the procedure stops}. Now we will show that if $\mathcal{A}$ is true, at any iteration,

$$G \subseteq \{K, K-1, \dots, K-k+1\} \text{ and } S \subseteq \{1, 2, \dots, K-k\}. \tag{3.3}$$

Before the first screening step of the procedure, condition (3.3) is obviously satisfied since $G = S = \emptyset$. Now assume that when $r = t-1$, (3.3) is satisfied. Then when r=t and $\mathcal{A}$ is true, it can be seen that $|A_i| < k-g$ for all $i \in I \cap \{K, K-1, \dots, K-k+1\}$ and $|B_i| < m-(k-g)$, for all $l \in I \cap \{1, \dots, K-k\}$. So (3.3) will be still satisfied when r=t. By induction, if $\mathcal{A}$ is true, (3.3) is satisfied when the procedure stops. Next we will show that when the procedure stops, the best k systems, systems $K, K-1, \dots, K-k+1$, will be selected if $\mathcal{A}$ is true. This is because if $|G| \ge k$ when the procedure stops, from (3.3), $G = \{K, K-1, \dots, K-k+1\}$ and the best k systems will be selected. If $|G| + |I| \le k$ when the procedure stops, $|S| \ge K - |G| - |I| \ge K-k$. From (3.3), $S = \{1, 2, \dots, K-k\}$ and $G \cup I = \{K, K-1, \dots, K-k+1\}$. Then the best k systems will be selected. Therefore, $CS \supseteq \mathcal{A}$, where CS represents the event that the k best systems are selected correctly. Then we have

$$\Pr\{CS\} \geq \Pr\{\mathcal{A}\}$$

$$= 1 - \Pr\left\{ \bigcup_{1 \leq j \leq K-k} \bigcup_{K-k+1 \leq i \leq K} e_{ji} \in E \right\}$$

$$\geq 1 - \sum_{1 \leq j \leq K-k} \sum_{K-k+1 \leq i \leq K} \Pr\{e_{ji} \in E\}$$

$$\geq 1 - k(K-k) \sum_{l=1}^{c} (-1)^{l+1} \left(1 - \frac{1}{2} I(l = c)\right) \left(1 + \frac{2\eta(2c-l)l}{c}\right)^{-\frac{n_0-1}{2}}$$

$$= 1 - \alpha.$$

The last inequality is based on the proof of Theorem 1 in Kim and Nelson (2001).■

When the observations from the *K* alternatives are uncorrelated, procedure FS-*k* can be improved by using a different equation to solve the constant η, as given in the following theorem.

Theorem 3.2. Under the same assumptions as Theorem 3.1, except that Σ is a diagonal matrix, the procedure FS-*k* selects *k* systems with probability of at least $1 - \alpha$ when η solves $g(\eta) = 1 - (1-\alpha)^{\frac{1}{k(K-k)}}$.

And when *c*=1,

$$\eta = \frac{1}{2}\left[\left(2 - 2(1-\alpha)^{\frac{1}{k(K-k)}}\right)^{\frac{-2}{n_0-1}} - 1\right] \tag{3.4}$$

Proof:

Let CS_{il} denote the event $\{e_{li} \notin E$ when the procedure stops$\}$, $i \in \{K, K-1, \ldots, K-k+1\}$, $l \in \{1, \ldots, K-k\}$, and $\mathcal{A}$ as defined in the proof of Theorem 3.1. Then we have

$$\Pr\{CS\} \geq \Pr\{\mathcal{A}\}$$

$$= \Pr\left(\bigcap_{l=1}^{K-k} \bigcap_{i=K-k+1}^{K} CS_{il} \right)$$

$$= E\left[\Pr\left(\bigcap_{i=K-k+1}^{K}\bigcap_{l=1}^{K-k} CS_{il}\,\middle|\, -X_{l1},\dots,-X_{l,N_l+1},X_{i1},\dots,X_{i,N_i+1},S_{il}^2\right)\right]$$

$$= E\left[\prod_{i=K-k+1}^{K}\prod_{l=1}^{K-k}\Pr\left(CS_{il}\middle|-X_{l1},\dots,-X_{l,N_l+1},X_{i1},\dots,X_{i,N_i+1},S_{il}^2\right)\right]$$

The last equality follows because the events are conditionally independent. Notice that $\Pr\left(CS_{il}\middle|X_{l1},\dots,X_{l,N_l+1},X_{i1},\dots,X_{i,N_i+1},S_{il}^2\right)$ is nondecreasing in $-X_{lj}$ and X_{ij}. Therefore, by Lemma 2 in Kim and Nelson (2001),

$$E\left[\prod_{i=K-k+1}^{K}\prod_{l=1}^{K-k}\Pr\left(CS_{il}\middle|X_{l1},\dots,X_{l,N_l+1},X_{i1},\dots,X_{i,N_i+1},S_{il}^2\right)\right]$$

$$= E_{S_{il}^2}\left[E_{X_{l1},\dots,X_{l,N_l+1},X_{i1},\dots,X_{i,N_i+1}}\left[\prod_{i=K-k+1}^{K}\prod_{l=1}^{K-k}\Pr\left(CS_{il}\middle|X_{l1},\dots,X_{l,N_l+1},X_{i1},\dots,X_{i,N_i+1},S_{il}^2\right)\right]\right]$$

$$\geq E_{S_{il}^2}\left[\prod_{i=K-k+1}^{K}\prod_{l=1}^{K-k}E_{X_{l1},\dots,X_{l,N_l+1},X_{i1},\dots,X_{i,N_i+1}}\left[\Pr\left(CS_{il}\middle|X_{l1},\dots,X_{l,N_l+1},X_{i1},\dots,X_{i,N_i+1},S_{il}^2\right)\right]\right]$$

$$= E_{S_{il}^2}\left[\prod_{i=K-k+1}^{K}\prod_{l=1}^{K-k}\left[1-\Pr\left(ICS_{il}\middle|S_{il}^2\right)\right]\right]$$

$$\geq \left(1-g(\eta)\right)^{k(K-k)}$$

$$= \left[1-\left(1-(1-\alpha)^{\frac{1}{k(K-k)}}\right)\right]^{k(K-k)} = 1-\alpha,$$

where the last inequality follows from the proof of Theorem 3.1 in Kim and Nelson (2001).■

3.3 Experimental Results

In this section, we conduct experiments to test the procedure FS-*k*. The alternative systems are represented by different configurations of *K* random variables generated from normal distribution. In all cases, system *K-k+1* to system *K* have the largest true means

and are considered as superior systems. We evaluate the procedure on different variance configurations of the systems, and also examine the effect of parameters including the number of alternative systems, *K*, the number of best systems required to be selected, *k*, the true means, $\mu_1, \dots, \mu_K$, and the true variances, $\sigma_1^2, \dots, \sigma_K^2$. We describe the configurations of the simulated datasets in Section 3.3.1, experimental results for evaluating the benefit of the pre-selection of superior systems in Section 3.1.2 and experimental results for comparisons between FS-*k* and three other methods.

3.3.1 Configurations and Experiment Design

As our procedure is developed as an extension of the method by Kim and Nelson (2001), we test our procedure on similar datasets from Kim and Nelson (2001) and select the same parameters for the following experiments. The first-stage sample size is selected to be $n_0^{raw} = 24$. The number of systems in each experiment is set as 30 and 10, and the number of best systems *k* is set as all the odd numbers in [1, *K*]. The indifference zone parameter is set as $\delta = \frac{1}{\sqrt{n_0^{raw}}}$.

Two system configurations are applied. The first one is the slippage configuration (SC), in which $\mu_K = \mu_{K-1} = \cdots = \mu_{K-k+1} = \delta$ and $\mu_{K-k} = \cdots = \mu_1 = 0$. The slippage configuration is also known as the least-favored configuration for ranking and selection, since the differences of the true means between each superior and inferior systems are exactly equal to the indifference zone value. To investigate the effectiveness of the procedure in selecting superior systems and eliminating inferior systems, the monotone decreasing means (MDM) configuration is also used. The differences between any two consecutive means are the same in MDM configuration. Thus, the true means of each system are set as $\mu_i = \delta * (i - K + 1)$for $i = 1, 2, \dots, K$.

For each configuration, the effect of increasing variances, decreasing variances and equal variances are examined. In the increasing-variance case, the variances of the superior systems are lower than the variances of inferior systems and they are set as

$\sigma_i^2 = |\delta(i-K)| + 1$ for $i = 1, 2, \ldots, K$; in the decreasing-variance case, the variances of the superior systems are higher than the variances of the inferior systems and they are set as $\sigma_i^2 = \frac{1}{|\delta(i-K)|+1}$ for $i = 1, 2, \ldots, K$; in the equal-variance case, $\sigma_i^2 = 1$.

Thus, we have six configurations in total: SC with increasing variances (SCin), SC with decreasing variances (SCde), SC with equal variances (SCeq), MDM with increasing variances (MDMin), MDM with decreasing variances (MDMde) and MDM with equal variances (MDMeq). When *K* and *k* values are fixed, for each configuration, 500 replications of experiments are performed. To evaluate the performance of our procedure, we use the average number of observations per system (ANS) as a measure of the computational cost of each procedure and PCS refers to the probability of correct selection, used to estimate the accuracy of the procedure. The nominal probability of correct selection is set as 0.95.

3.3.2 Experimental Results for Evaluating Benefits of Pre-selection of Superior Systems

Table 3.1 shows the experimental results when the number of alternative systems *K* = 10 with *k*=1, 3, 5, 7 and 9 best systems. Table 3.2 shows the experimental results when the number of alternative systems *K* = 30 with *k*=1, 3, 5, 7…, 27, 29. To test the benefit of the pre-selection of superior systems, we show and compare the results from both the procedure with pre-selection of superior systems performed and the procedure without the pre-selection using the same data sets in each configuration and *k* value. The ANS, total number of stages, as well as the estimated probability of correct selection (PCS), are shown in Table 3.1 and Table 3.2.

3.3.2.1 Effect of Pre-selection of Superior Systems

Our procedure can pre-select obviously superior systems in order to save the observations generated from these superior systems. We conduct experiments to compare

the ANS and the number of stages between the procedures with and without the pre-selection of the superior systems.

Table 3.1 shows the experiment results when the number of alternatives, K=10. And Table 3.2 shows the results when K=30. From Table 3.1 and Table 3.2, we can find the procedure with or without selection makes no difference when we select the single best system. This is not out of expectation because if the (single) superior system is selected, it means at the same time all the other systems can be eliminated. Therefore, pre-selection of superior systems does not save any additional sampling from the procedure with only elimination of inferior systems (without the pre-selection of superior systems).

When $k > 1$, the procedure with the pre-selection of superior systems can save more observations than the procedure without pre-selection. The number of stages is only affected by the system that needs the most samples. So when k>1, the pre-selection of superior systems may reduce the number of samples needed from many systems, but may have slight effect on the system that requires the most samples. As k increases, we can find that the ANS increases first and then decreases. The location of the peak depends on the type of the data. In equal-variance cases, for both SC and MDM configurations, the ANS value is symmetric when the procedure includes pre-selection of superior systems. This is because we can consider the pre-selection of superior systems and elimination of inferior systems as symmetric procedures for selecting systems with larger means as the best or selecting systems with smaller means as the best.

Table 3.1. ANS and number of stages for procedures with and without selection during process when the number of alternatives is K=10.

C		Without selection			With selection		
		A	S	P	A	S	P
S		1	1	0.	11	1	0.
		1	2	0.	17	2	0.
		2	2	0.	17	2	0.
		2	1	0.	14	1	0.
		1	1	0.	89	1	0.
S		1	2	0.	17	2	0.
		2	3	0.	27	3	0.
		3	3	0.	29	3	0.
		3	3	0.	27	3	0.
		2	2	0.	17	2	0.
S		2	4	0.	28	4	0.
		5	6	0.	47	6	0.
		6	7	0.	56	7	0.
		7	7	0.	55	8	0.
		6	5	0.	36	6	0.
M		6	1	0.	63	1	0.
		8	1	0.	74	1	0.
		9	1	1.	66	1	1.
		9	7	0.	53	7	0.
		6	4	0.	35	4	0.
M		7	1	0.	74	1	0.
		1	2	1.	11	2	1.
		1	2	1.	12	2	1.
		2	2	1.	11	2	1.
		1	1	0.	74	1	0.
M		9	1	0.	99	1	0.
		2	3	1.	17	3	1.
		3	4	0.	23	4	0.
		4	5	0.	25	5	0.
		4	4	0.	18	4	0.

Table 3.2. ANS and number of stages for procedures with and without selection during process when the number of alternatives is K=30

Config.	k	Without selection			With selection		
		A	S	P	AN	S	P
SCde	1	87	2	0	87.	2	0
	3	14	3	0	133	3	0
	5	16	3	0	149	3	0
	7	17	2	0	155	2	0
	9	19	2	0	157	2	0
	11	20	2	0	158	2	0
	13	21	2	0	155	2	0
	15	21	2	0	151	2	0
	17	21	2	0	144	2	0
	19	22	2	0	138	2	0
	21	22	2	0	131	2	0
	23	22	2	0	122	2	0
	25	22	2	0	110	2	0
	27	20	1	0	92.	1	0
	29	16	1	0	60.	1	0
SCeq	1	20	3	0	204	3	0
	3	34	5	0	333	5	0
	5	42	6	0	395	6	0
	7	47	6	0	432	6	0
	9	52	7	0	461	7	0
	11	56	7	0	478	7	0
	13	59	7	0	490	7	0
	15	62	7	0	495	7	0
	17	64	7	0	491	7	0
	19	65	7	0	480	7	0
	21	65	6	0	461	7	0
	23	65	6	0	435	6	0
	25	62	6	0	396	6	0
	27	56	5	0	334	5	0
	29	40	3	0	209	3	0
SCin	1	62	1	0	623	1	0
	3	98	1	0	943	1	0
	5	12	2	0	117	2	0

	7	15	2	0	132	2	0
	9	18	2	0	150	2	0
	11	20	2	0	162	2	0
	13	22	2	0	172	2	0
	15	24	3	0	181	3	0
	17	27	3	0	187	3	0
	19	28	3	0	191	3	0
	21	30	3	0	189	3	0
	23	31	3	0	185	3	0
	25	31	3	0	172	3	0
	27	29	3	0	148	3	0
	29	22	2	0	934	2	0
MDMde	1	43	2	1	43.	2	1
	3	55	1	1	49.	1	1
	5	60	1	1	47.	1	1
	7	62	1	1	44.	1	1
	9	63	1	1	41.	1	1
	11	62	9	1	37.	9	1
	13	62	8	1	34.	8	1
	15	61	6	1	32.	6	1
	17	61	6	0	31.	6	0
	19	59	5	1	29.	5	1
	21	57	4	1	28.	4	1
	23	56	4	1	27.	4	1
	25	52	3	1	26.	3	1
	27	46	2	1	26.	2	1
	29	37	1	0	24.	1	0
MDMeq	1	49	2	1	49.	2	1
	3	80	2	1	69.	2	1
	5	10	3	1	80.	3	1
	7	13	3	1	87.	3	1
	9	15	3	1	92.	3	1
	11	18	3	0	95.	3	0
	13	20	3	1	97.	3	1
	15	23	3	0	97.	3	0
	17	25	3	1	96.	3	1
	19	27	3	1	95.	3	1
	21	28	3	1	92.	3	1

	23	30	3	1	88.	3	1
	25	30	3	1	80.	3	1
	27	29	2	0	70.	2	0
	29	24	2	0	49.	2	0
MDMin	1	76	2	0	76.	2	0
	3	14	4	1	133	4	1
	5	23	6	0	181	6	0
	7	33	7	1	223	7	1
	9	45	9	1	269	9	1
	11	59	1	1	309	1	1
	13	76	1	0	351	1	0
	15	92	1	1	384	1	1
	17	11	1	1	421	1	1
	19	12	1	0	436	1	0
	21	14	1	1	454	1	1
	23	16	1	1	454	1	1
	25	18	1	1	445	1	1
	27	18	1	1	395	1	1
	29	16	1	0	264	1	0

3.3.2.2 Effect of Mean/Variance Configurations

In Figure 3.1, we show the plots of the ANS versus the *k* value and the number of stages versus the *k* value for all six system configurations when *K*=30. We can find that for the MDM configuration, the difference of the number of stages between the procedure with or without pre-selection of superior systems is too small to be distinguished in the plot. For SC data, when *k* is small, the difference is also not significant. The three different kinds of variance settings have different influences for the number of stages. When the variance is set larger for a system with a larger true mean, the number of stages peaks when *k* is around 5. When the systems have equal variances, the number of stages peaks in the middle where *k* is around 15. When the variance is set larger for a system with a smaller true mean, the number of stages peaks when *k* is around 25. For the

method with pre-selection of superior systems, ANS has the same trend as the number of stages. For the method without selection, the ANS and number of stages do not follow the same trend. When pre-selection of superior systems is not used, the ANS may still increase even after the number of stages has decreased. For all configurations except MDMde, the ANS peaks around 25; while MDMde peaks around 5. This is because without selection during processing, we need to generate observations for the superior systems until the end of the procedure. The number of superior systems is more dominant than the number of stages required to remove all the inferior systems for the five configurations SCin, SCeq, SCde, MDMeq and MDMin; while for MDMde data, it is easier to remove the inferior systems and thus the number of stages required to remove all the inferior systems is more dominant than the number of superior systems.

3.3.2.3 Effect of Wrong Arc

Among all the experiments, we also record the estimated probability that a *wrong arc*, which is the event defined as $\{e_{ij} = 1, \exists 1 \leq K-k, K-k+1 \leq j \leq K\}$, is added to the graph before the procedure stops. The experiment results show that the probability of correct selection is always greater than or equal to one minus the probability of wrong arc. This is expected based on the proof of Theorem 3.1. However, a wrong arc will not always lead to incorrect selections. We illustrate this by using a simple example as shown in Figure 3.2. In the system in Figure 3.3, we want to select 2 systems from 4 alternatives represented by the squares. Systems 4 and 3 are superior systems and system 1 and 2 are inferior ones. Suppose in one stage (stage *i*) of the procedure, a wrong arc 1→3 is added and at a later stage (stage *j*>*i*), three other arcs 3→4, 4→2 and 2→1 are added. In this case, according to our procedure, all the four systems would be selected to set *G* defined in Procedure FS-*k*. Then the two systems with larger sample means would be selected as the best systems. Thus, as long as in stage *j*, system 4 and 3 have larger sample means, we can still obtain a correct selection.

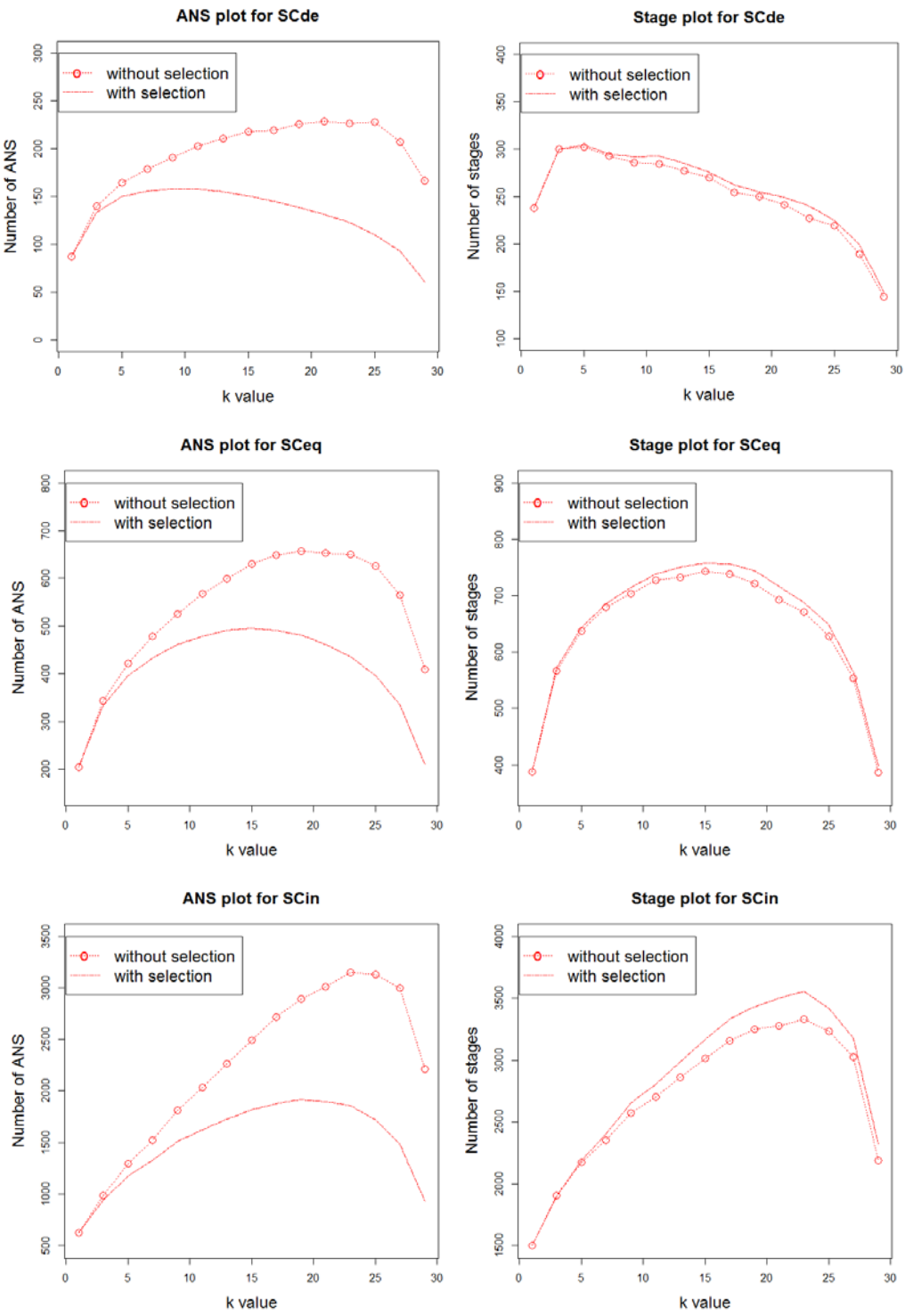

Figure 3.1. Plots of ANS and stages for SC data

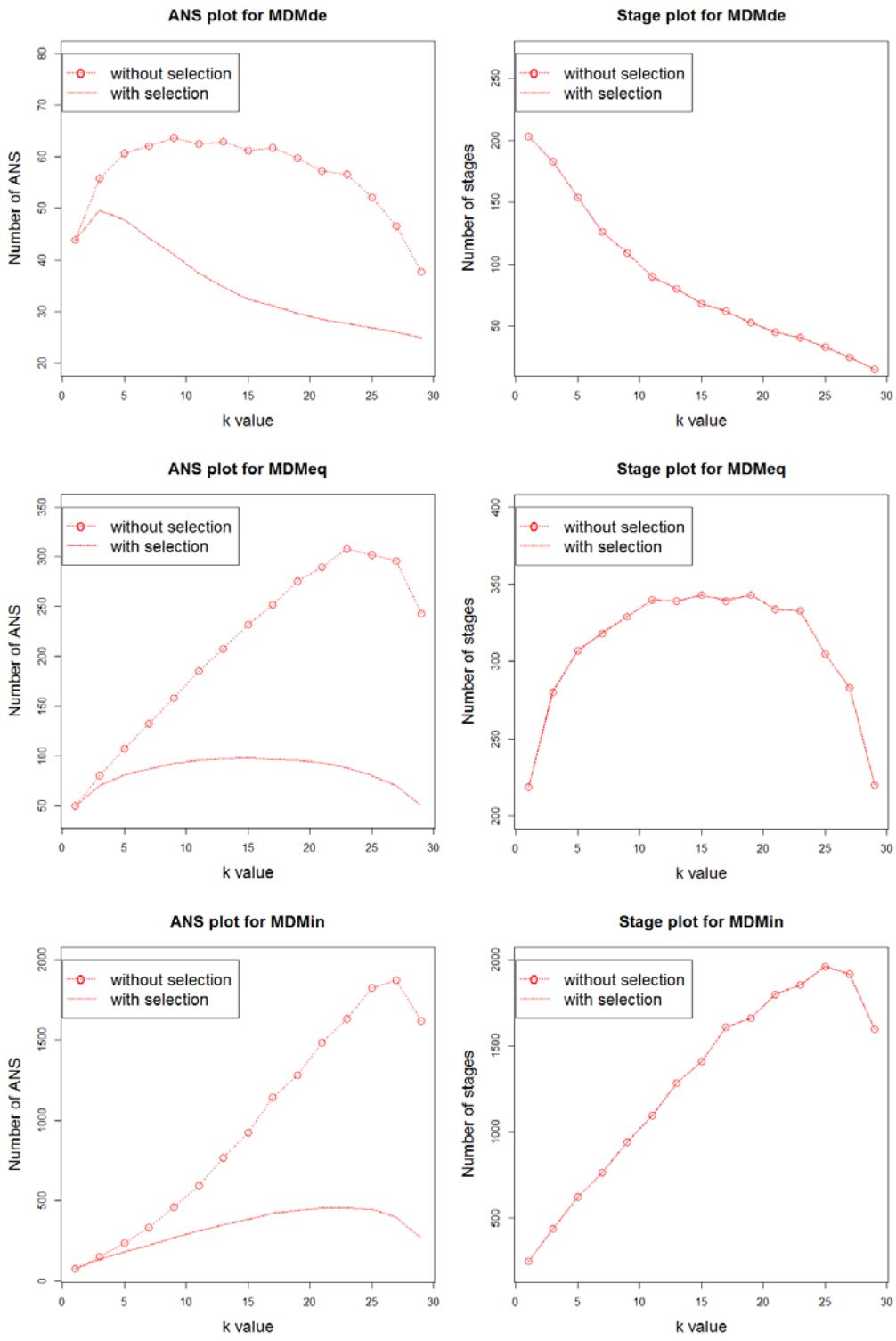

Figure 3.2. Plots of ANS and stages for MDM data

Figure 3.3. Example for wrong arc but correct selection

3.3.3 Experimental Results for Comparisons with Existing Methods

3.3.3.1 Existing Methods and Common Random Number

To solve the single best ranking and selection problem, Dudewicz and Dalal (1975) proposes a two-stage method, Rinott (1978) proposes another two-stage method and Nelson, Swann et al. (2001) proposes the combined procedure. Kiekhaefer (2011) makes extension of these three methods for selecting multi-best systems from alternatives. These extended methods are referred to as P_E, $P_{R'}$ and $P_{SP''}$ in Kiekhaefer (2011) and also applied in this thesis. We conduct experiments to compare the performance between our FS-*k* method and these three methods.

Also, experiments are performed to test the effect of dependence across systems' outputs due to the use of *common random numbers* (CRN). The effect of CRN is to reduce the value of σ_{il}^2. The reduction of σ_{il}^2 will tend to reduce S_{il}^2 which would reduce the value of N_{il} and could eliminate alternatives earlier in the sampling process. On the other hand, the use of CRN would increase the value of η and this tends to reduce the efficiency of the procedure (due to the different ways of calculation of η in (3.2) and

(3.4)). In this section, we show that the decrease in S_{il}^2 due to the use of CRN can make the procedure more efficient despite of the increase in η.

3.3.3.2 Experimental Results

Experiments are conducted for configurations MDMde, MDMeq, MDMin, SCde, SCeq and SCin. *K* value is set as 10 and 30. And in each case, *k* = 1, 2, 3, 4, 5 best systems are selected. When CRN is employed, ρ represents the correlation between each pair of systems. And we consider ρ= 0.25, 0.5, 0.75.

Table 3.3 shows the comparison results for FS-*k* method and P_E, $P_{R'}$ and $P_{SP''}$ among 10 alternative systems (*K*=10). Table 3.4 shows the comparison results among 30 systems (*K*=30). Correlation ρ=0 indicates the systems are independent with each other. From the two tables, we can find that the FS-*k* method can save much more observations than the three other methods. The trend is clear that FS-*k* methods have universally smaller ANS than the others among all the configuratoins. And the PCS are all over the pre-specified value 0.95. It is shown in Table 3.5 that how much more efficient FS-*k* is than the other three methods.

Table 3.3. Comparison experiments between FS-*k* method and other methods among *K*=10 alternatives

MDMde *K*=10										
	k=1		*k*=2		*k*=3		*k*=4		*k*=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-*k* ρ=0	330.4	0.994	353.2	0.994	331.1	1	305.4	0.998	287.3	1
FS-*k* ρ=0.25	302.7	0.994	299.9	0.998	287.2	0.998	271.1	1	260.1	0.996
FS-*k* ρ=0.5	263.6	0.99	262.3	0.996	253.3	0.996	246.9	0.994	244	1

FS-k ρ=0.75	242.4	0.998	241.2	0.996	240.3	1	240.1	1	240	1
$P_{R'}$	724.2	1	856.3	1	909.9	1	932.7	1	941.2	1
$P_{SP''}$	497.8	1	526.2	1	437.2	1	368.4	1	330.3	1
P_E	655.5	1	784.4	1	837	1	858.9	1	867.1	1
MDMeq K=10										
	k=1		k=2		k=3		k=4		k=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-k ρ=0	37.20	0.998	44.21	0.998	49.38	0.998	51.41	0.998	52.08	0.996
FS-k ρ=0.25	32.21	1	36.83	0.996	38.81	0.994	40.08	1	40.78	1
FS-k ρ=0.5	27.41	1	29.46	0.998	30.15	0.996	31.11	1	30.64	1
FS-k ρ=0.75	24.36	0.998	24.62	0.998	24.82	0.998	24.85	1	24.85	1
$P_{R'}$	146.5	0.996	170.8	0.996	185.7	0.998	191	1	193	0.996
$P_{SP''}$	57.22	0.998	77.76	0.998	90.66	1	93.16	1	95.40	0.996
P_E	132.3	0.994	156.4	0.998	170.7	0.996	175.8	1	177.7	0.998
MDMin K=10										
	k=1		k=2		k=3		k=4		k=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-k ρ=0	48.41	0.996	71.30	1	90.88	0.996	109.1	0.996	124	0.998
FS-k ρ=0.25	38.60	0.994	54.85	0.996	70.15	0.998	82.59	0.996	92.41	0.998
FS-k ρ=0.5	29.84	1	39.39	0.998	48.32	1	57.31	1	63.44	1
FS-k ρ=0.75	24.78	0.998	26.82	1	29.64	1	32.54	0.998	35.39	0.998
$P_{R'}$	355	0.996	415.3	1	445.7	1	462.8	1	469.3	0.996
$P_{SP''}$	90.53	0.998	166.1	1	252.8	1	344.7	1	411.5	1
P_E	320.8	0.996	380.1	0.996	409.6	0.996	426.1	0.998	432.1	0.996
SCde K=10										
	k=1		k=2		k=3		k=4		k=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-k ρ=0	44.98	0.966	57.26	0.966	61.09	0.98	61.84	0.968	59.94	0.984

FS-k ρ=0.25	36.66	0.986	44.14	0.98	46.91	0.97	47.54	0.97	46.45	0.98
FS-k ρ=0.5	29.31	0.984	32.57	0.98	34.19	0.978	34.15	0.978	33.68	0.978
FS-k ρ=0.75	24.68	0.98	25.19	0.996	25.36	0.996	25.60	0.982	25.45	0.984
$P_{R'}$	72.60	0.952	84.95	0.962	91.01	0.964	94.03	0.978	94.68	0.96
$P_{SP''}$	66.14	0.962	88.94	0.976	98.40	0.988	101.1	0.982	100.3	0.976
P_E	65.72	0.946	77.83	0.94	83.71	0.96	86.61	0.952	87.23	0.952
SCeq K=10										
	k=1		k=2		k=3		k=4		k=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-k ρ=0	72.79	0.96	100.3	0.966	116.3	0.964	124.4	0.956	124.9	0.956
FS-k ρ=0.25	55.92	0.968	76.62	0.96	87.78	0.956	93.61	0.978	96.31	0.948
FS-k ρ=0.5	39.45	0.98	51.90	0.962	58.67	0.964	62.76	0.972	63.17	0.96
FS-k ρ=0.75	25.88	0.986	29.09	0.954	31.93	0.98	32.91	0.982	33.25	0.966
$P_{R'}$	145.7	0.958	170.8	0.952	185.6	0.97	191	0.95	192.4	0.958
$P_{SP''}$	142.3	0.966	190.2	0.966	214	0.98	221.2	0.98	223.6	0.978
P_E	131.6	0.932	156.3	0.94	170.6	0.962	175.8	0.94	177.2	0.936
SCin K=10										
	k=1		k=2		k=3		k=4		k=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-k ρ=0	287.1	0.948	408.3	0.938	473.6	0.966	538.1	0.962	562.9	0.956
FS-k ρ=0.25	213.3	0.968	308.2	0.95	361.7	0.966	405.6	0.96	424.8	0.948
FS-k ρ=0.5	153.9	0.972	210.9	0.968	250	0.956	272.4	0.97	288.4	0.976
FS-k ρ=0.75	83.60	0.966	115.7	0.962	134.7	0.972	147.6	0.978	156.5	0.96
$P_{R'}$	676.6	0.964	794.2	0.954	858.5	0.97	883.4	0.946	886.8	0.976
$P_{SP''}$	782.7	0.962	938.1	0.97	1012	0.986	1037	0.974	1042	0.98
P_E	611.2	0.938	727	0.95	789.1	0.962	813.1	0.942	816.6	0.954

Table 3.4. Comparison experiments between FS-k method and other methods among K=30 alternatives

MDMde K=30										
	k=1		k=2		k=3		k=4		k=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-k ρ=0	29.14	0.996	29.87	0.998	29.22	0.998	28.22	1	27.52	1
FS-k ρ=0.25	27.38	0.998	27.59	1	26.95	0.998	26.37	0.998	25.79	1
FS-k ρ=0.5	25.58	0.998	25.46	1	25.10	0.998	24.77	1	24.55	1
FS-k ρ=0.75	24.18	1	24.15	1	24.06	1	24.04	1	24.01	1
$P_{R'}$	54.03	1	62.73	1	66.42	1	69.90	1	70.76	1
$P_{SP''}$	37.69	1	41.62	1	37.22	1	33.29	1	30.87	1
P_E	48.60	1	56.91	1	60.51	1	63.78	1	64.69	1
MDMeq K=30										
	k=1		k=2		k=3		k=4		k=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-k ρ=0	31.19	1	34.68	0.998	37.11	1	38.92	1	40.18	1
FS-k ρ=0.25	28.26	1	30.86	1	32.14	1	33.56	1	34.29	1
FS-k ρ=0.5	26.04	0.998	27.20	1	27.95	1	28.31	1	28.82	1
FS-k ρ=0.75	24.32	1	24.58	1	24.69	1	24.79	1	24.92	1
$P_{R'}$	199.2	0.998	230.4	1	249.3	1	262.5	0.998	270.2	1
$P_{SP''}$	42.89	0.996	54.05	1	61.28	1	65.50	1	67.04	1
P_E	174.9	0.996	206.6	1	225.1	0.998	237.9	1	245.5	1
MDMin K=30										
	k=1		k=2		k=3		k=4		k=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-k ρ=0	41.32	1	56.75	0.998	71.04	1	84.58	1	99.38	0.998

FS-k ρ=0.25	34.41	1	45.11	0.998	55.81	1	66.41	1	76.20	1
FS-k ρ=0.5	28.57	0.998	34.30	0.996	40.59	0.998	46.69	0.998	54.07	1
FS-k ρ=0.75	24.71	1	26.19	1	27.86	1	30.33	0.998	33.00	1
$P_{R'}$	1111	0.998	128	0.998	1388	1	1461	0.998	1510	0.998
$P_{SP''}$	64.87	0.998	106.9	1	157.1	1	212.1	1	280.3	0.998
P_E	975.7	0.994	1153	0.998	125	1	1324	0.998	1371	0.998
SCde K=30										
	k=1		k=2		k=3		k=4		k=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-k ρ=0	35.66	0.988	42.53	0.982	46.59	0.984	48.97	0.988	50.11	0.988
FS-k ρ=0.25	31.14	0.996	35.65	0.99	38.11	0.986	39.27	0.996	40.12	0.978
FS-k ρ=0.5	27.51	0.99	29.59	0.994	30.45	0.986	31.47	0.998	31.57	0.992
FS-k ρ=0.75	24.78	0.99	25.42	0.992	25.73	0.988	25.97	0.994	26.07	0.994
$P_{R'}$	53.80	0.974	61.65	0.974	66.38	0.97	70.26	0.98	72.47	0.984
$P_{SP''}$	47.62	0.984	60.75	0.974	68.31	0.986	73.75	0.992	76.40	0.988
P_E	48.44	0.952	55.97	0.952	60.47	0.946	64.11	0.958	66.23	0.968
SCeq K=30										
	k=1		k=2		k=3		k=4		k=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-k ρ=0	86.70	0.97	119.4	0.966	139.3	0.958	154	0.958	166.4	0.964
FS-k ρ=0.25	65.63	0.968	90.92	0.966	106.0	0.97	116.7	0.962	125.9	0.976
FS-k ρ=0.5	45.59	0.962	60.58	0.96	70.56	0.978	78.43	0.974	84.20	0.976
FS-k ρ=0.75	27.69	0.974	32.43	0.984	36.90	0.978	40.47	0.97	42.76	0.972
$P_{R'}$	198.9	0.974	229.7	0.976	249.3	0.962	261.3	0.962	271.3	0.974
$P_{SP''}$	199.2	0.972	254.1	0.988	280.5	0.972	295.5	0.982	306.9	0.982
P_E	174.6	0.956	205.9	0.944	225.1	0.942	236.9	0.948	246.5	0.956

SCin K=30										
	k=1		k=2		k=3		k=4		k=5	
	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS	ANS	PCS
FS-k ρ=0	617.2	0.962	821.9	0.96	953.3	0.952	1072	0.964	1169	0.97
FS-k ρ=0.25	497.3	0.95	643.7	0.972	748.4	0.966	847	0.958	922.8	0.962
FS-k ρ=0.5	373.5	0.954	482	0.958	552.7	0.958	603.1	0.962	655.6	0.97
FS-k ρ=0.75	235.8	0.962	305.5	0.97	348.3	0.958	374.5	0.972	404.5	0.952
$P_{R'}$	1886	0.976	2194	0.962	2359	0.966	2490	0.978	2588	0.964
$P_{SP''}$	2165	0.97	2524	0.976	2700	0.986	2838	0.984	2943	0.986
P_E	1655	0.962	1966	0.948	2130	0.948	2257	0.97	2351	0.956

Table 3.5 shows the maximum and minimum percentages of observations FS-k saves compared to other three methods when ρ=0 (CRN is not used), as well as P_E, $P_{R'}$ and $P_{SP''}$, for all six configurations. From the results, we can find that FS-k can save 13.02%-86.37% observations over all these cases than P_E, $P_{R'}$ and $P_{SP''}$ when K=10. It can save 10.85%-96.28% when K=30. The smaller k value is, the larger fraction of observations tends to be saved by FS-k. Table 3.6 shows the maximum and minimum percentages of observations saved by CRN compared with ρ=0 for the FS-k procedure (CRN is not used) among all six configurations. The larger correlation is applied, the more observations are saved. And we cannot see a clear pattern of influence from k value. The results demonstrate that a higher correlation among system can save more observations.

Table 3.5. Maximum and minimum percentages of observations FS- k ρ=0 saves compared to other three methods among all six configurations.

		k = 1	k = 2	k = 3	k = 4	k = 5
K=10	Max	86.37%	82.83%	79.61%	76.41%	73.57%
	Min	31.56%	26.43%	24.27%	17.10%	13.02%
K=30	Max	96.28%	95.59%	94.89%	94.21%	93.42%
	Min	22.69%	24.01%	21.49%	15.23%	10.85%

Table 3.6. Maximum and minimum percentages of observations saved by CRN compared with ρ=0 for FS-k among all six configurations.

			k = 1	k = 2	k = 3	k = 4	k = 5
K =10	ρ =0.25	Max	25.68%	24.51%	24.57%	24.81%	25.48%
		Min	8.38%	15.09%	13.26%	11.23%	9.47%
	ρ =0.5	Max	46.38%	48.34%	49.58%	49.59%	49.44%
		Min	20.22%	25.74%	23.50%	19.16%	15.07%
	ρ =0.75	Max	70.89%	71.66%	72.56%	73.56%	73.39%
		Min	26.63%	31.71%	27.42%	21.38%	16.46%
K =30	ρ =0.25	Max	24.30%	23.89%	23.89%	24.18%	24.35%
		Min	6.04%	7.63%	7.77%	6.56%	6.29%
	ρ =0.5	Max	47.42%	49.29%	49.36%	49.08%	49.43%
		Min	12.22%	14.76%	14.10%	12.23%	10.79%
	ρ =0.75	Max	68.06%	72.85%	73.52%	73.73%	74.32%
		Min	17.02%	19.15%	17.66%	14.81%	12.75%

3.4 Summary

In this chapter, we described a fully sequential procedure, which pre-selects superior systems and eliminate inferior systems at each stage, to determine multi-best systems from alternative systems. We proved that, by using our procedure, we can correctly select a pre-specified number (*k*) of best systems from *K* alternative systems with probability of at least $1-\alpha$, when the indifference zone parameter δ is known. We also conducted experiments to support the theoretical results and tested the influences of the number of systems to be selected, the pre-selection of superior systems, and different mean and variance configurations respectively. Compared to three existing methods, FS-*k* was shown to be able to significantly save computational load. We also conducted experiments to study the influence of common random numbers on the performance of the algorithm. CRN is demonstrated to be helpful in saving observations for the FS-*k* procedure to do multi-best ranking and selection. The larger the correlation is set, the more observations are saved.

CHAPTER 4

RANKING AND SELECTION FOR STEADY-STATE SIMLATION

4.1 Problem Framework

In Chapter 3, we introduced a *k*-best ranking and selection method (FS-*k*) for terminating simulation. In this chapter, we will describe how to extend this fully sequential method to solve the multi-best R&S for steady-state simulation. The main issue to adapt FS-*k* to steady-state simulation is that FS-*k* is designed for independent and identically distributed data outputs. In other words, outputs from a single system are i.i.d. However, data outputs from one steady-state simulation are usually auto-correlated. To solve this problem, there are two simple methods discussed in Section 1.3. First we can generate *m* replications and consider the steady-state sample mean in each replication as one observation for the ranking and selection procedure. Another method is to generate a single long replication and use the batch mean method to generate approximately independent observations, which are the batch means. There are drawbacks in both methods. As we stated before, steady-state simulation focuses on the long-run performance of a system. The initial trend data in the transient state can cause trouble when we include them to estimate the steady-state mean. If we use the first method, i.e., to generate multiple replications, we have to remove the samples from the initial transient state for each replication, which causes a huge waste of samples. If we use the second method, i.e., to generate a single replication with batching, although we only need to do initialization bias elimination once, the batch size is hard to control. If the batch size is too large, it will reduce the opportunities to eliminate/screening inferior systems or preselect superior systems, which will cause inefficiency in the ranking and selection procedure. Thus, we will study another method which can maintain the efficiency of the fully-sequential R&S procedure and solve the i.i.d problem at the same time.

The main reason we need to assume that the outputs are i.i.d for the FS-*k* method in Chapter 3 is that we can then compute the sample variance of the difference between system *i* and *l*, simply by

$$S_{il}^2 = \frac{1}{n_0 - 1}\sum_{j=1}^{n_0}(X_{ij} - X_{lj} - (\bar{X}_i(n_0) - \bar{X}_l(n_0)))^2 \qquad (4.1)$$

For autocorrelated data from a single replication of steady-state simulation, however, the estimator in (4.1) cannot be used to estimate the variance of the sample mean. Therefore, an important task in this chapter is to study and use valid variance estimator for autocorrelated data. In Goldsman et al. (2002) and Kim and Nelson (2006), it has been shown both empirically and theoretically (in asymptotic sense) that modification of the fully sequential procedures for terminating simulation to select the single best alternative also work well in the single-replication steady-state simulation environment. In this chapter, we will follow the similar ideas to modify the FS-*k* procedure to select multiple best systems for single replications of steady-state simulation and test its performance.

In Section 4.2, we will first introduce the idea of asymptotic variance estimators. Then in Section 4.3 we present how to extend the FK-*k* method for ranking and selection problems in steady-state simulation. We will also describe the procedure and provide the theoretical proof for it in terms of the asymptotic lower bounds of its probability of correct selection. Then we will show the design of simulation experiments to test the efficiency of our procedure in Section 4.4 and provide the experimental results in Section 4.5 to see how the procedure behaves for finite samples.

4.2 Asymptotic Analysis and Asymptotic Variance Estimators

Asymptotic analysis, especially in this chapter, is to analyze the performance of a procedure as the simulation run length approaches infinity. It can establish some large-sample validity if used appropriately. To estimate the variance of sample means for

autocorrleated data, a key concept is the *variance parameter* (or *asymptotic variance constant*), $v_i^2 \equiv \lim_{r\to\infty} r\,Var(\bar{X}_i(r))$ where $\bar{X}_i(r) = \frac{1}{r}\sum_{j=1}^{r} X_{ij}$. There are many methods to estimate the variance parameter from the literature. Goldsman et al. (2002) review the variance estimators that have been developed. Goldsman (1983) suggests batch mean estimators. Nakayama (1995) applies the batch mean estimator to select the best system. Goldsman (1985) revises Dudewicz and Dalal's (1975) method by applying the theory of standardized time series to estimate the variance parameter. Dudewicz and Zaino (1977) estimate v_i^2 specifically for AR(1) process. Goldsman, Kim et al.(2002) and Kim and Nelson (2006) validated the usage of certain variance estimator in single-best fully sequential selection procedures for steady-state simulation empirically.

There are a group of variance estimators based on the batching method. As we mentioned in the previous chapter, batching is to partition *r* output observations into *b* batches so that each batch has *m* observations. So we have $r = mb$. We will apply two batching-based variance estimators in this chapter and compare the performance between them. Different from the batching method mentioned in the previous section, when using batching to estimate the variance parameter, it's not strongly required that the batch means are i.i.d. Thus the selection of batch size is much more flexible.

The two asymptotic variance estimators we use are the batch means (BM) estimator and the weighted area (AREA) estimator. Suppose we have *r* observations $X_{i1}, \dots, X_{ir}$ from alternative i. The number of batches is set as *b* and the batch size is *m*. For BM estimator, the batch means are calculated as

$$\bar{X}_{i,j,m} = \frac{1}{m}\sum_{l=1}^{m} X_{i,(j-1)m+l}\,,\ j = 1, 2, \dots, b.$$

Then the BM estimator recommended by Goldsman (1983) and Glynn and Whitt (1991) is

$$mV_B^2 = \frac{m}{b-1}\sum_{j=1}^{b}(\bar{X}_{i,j,m} - \bar{X}_i(r))^2 \to \frac{v_i^2\chi^2(b-1)}{b-1}, \text{ where } \bar{X}_i(r) = \frac{1}{r}\sum_{j=1}^{r} X_{ij} \text{ and } r \to \infty.$$

And the degree of freedom of the chi-square distribution $\chi^2(d)$ is *d*=*b*-1.

The weighted area (AREA) estimator comes from a different methodology known as the standardized time series introduced by Goldsman (1990) and Schmeiser and Song (1989). First, a *cumulative mean* is calculated for each element in each batch of system *i* as:

$$\bar{X}_{i,j,h} = \frac{1}{h}\sum_{l=1}^{h} X_{i,(j-1)m+l} \text{ for } j=1, 2, \ldots, b \text{ and } h=1, 2, \ldots, m.$$

Then the *standardized time series* from batch *j* of alternative *i* is defined as:

$$T_{i,j,m}(t) = \frac{\lfloor mt \rfloor (\bar{X}_{i,j,m} - \bar{X}_{i,j,\lfloor mt \rfloor})}{v_i\sqrt{m}}.$$

The *weighted area* from batch *j* of system *i* is defined as:

$$A_{i,j} = \frac{v_i}{m}\sum_{l=1}^{m} \omega\,(l/m) T_{i,j,m}(l/m),$$

where $\omega(\cdot)$ is pre-specified weighting function.

Then the weighted area estimator can be written as:

$$mV_A^2 = \frac{1}{b}\sum_{j=1}^{b} A_{i,j}^2 \rightarrow \frac{v_i^2 \chi^2(\mathrm{b})}{b}.$$

The approximate distribution of estimator mV_A^2 is proved by Goldsman et al. (1990). The degree of freedom of the chi-square distribution $\chi^2(d)$ is $d = b$.

From the description above about the two variance estimators, we can see that they both satisfy the Assumption 4.1, stated in the next section:

4.3 FS-*k*+ Procedure

In this section, we will first describe the general idea of our approach and also introduce the related assumptions and background knowledge. And then we will present the procedure FS-*k*+, which is extended from FS-*k* in Chapter 3 and designed to solve multi-best ranking and selection for steady-state simulation. We will also give the theoretical proof to justify the procedure asymptotically.

Similar to the proof of Theorem 3.1, a lemma about Brownian motion process will be used.

Lemma 4.1 (Fabian 1974)

For a Brownian Motion Process $B(\cdot,\Delta)$ on $[0,+\infty)$, with $E[B(t,\Delta)] = \Delta \mathrm{t}$ and $\mathrm{Var}[B(t,\Delta)] = \mathrm{t}$ where $\Delta > 0$. Then set

$U(t) = \mathcal{A} - \mathcal{B}t,$

$L(t) = -\mathcal{A} + \mathcal{B}t$

for some $\mathcal{A} > 0$ and $\mathcal{B} = \Delta/(2c)$ for some positive integer *c.* And let $R(t)$ represent the interval $(L(t), U(t))$ and let T^* be the first time that $B(t,\Delta) \notin R(t)$. Finally, let $\mathcal{G}$ be the event that $B(T^*,\Delta) \leq L(T^*)$. Then, we have

$$\Pr\{\mathcal{G}\} = \sum_{l=1}^{c} (-1)^{l+1} \left(1 - \frac{1}{2} I(l = c)\right) \exp\{-2\mathcal{A}\mathcal{B}(2c - l)l\}.$$

Remarks: here function $I(event) = 1$, if $event$ is true, otherwise equal to 0. The event $\mathcal{G}$ in the lemma can correspond to the situation that an incorrect selection is made between two systems.

Lemma 4.1 (Fabian 1974) constructs a "continuation region" $R(t)$ for a specified Brownian motion process $B(\cdot,\Delta)$. The very first time $B(\cdot,\Delta)$ moves out of the region $R(t)$ (from either the upper side or lower side of the region) is denoted by T^*. And the probability that at T^* it moves out from the lower side can be calculated mathematically as in the lemma. When we compare two systems, e.g. systems *i* and *l* with $\mu_i > \mu_l$, in the ranking and selection procedure, we can construct an approximate Brownian motion process $B(\cdot,\Delta)$ based on the cumulative difference between these two systems, $S(r) = \sum_{j=1}^{r}(X_{ij} - X_{lj}), r = n_0, \ldots, N$. Because $\mu_i > \mu_l$, the process is more likely to leave from the region from the upper boundary, which is corresponding to correct selection between systems *i* and *l*. And leaving the region first from the lower boundary corresponds to incorrect selection. By adjusting the upper and lower limits appropriately, we are able to control the probability of incorrect selection, which is the event $\mathcal{G}$ in the Lemma 4.1.

The next problem we are facing is how to approximate the cumulative difference between two systems $S(r)$ by a Brownian motion process. The following assumption can help us accomplish this:

Assumption 4.1

Functional Central Limit Theorem (FCLT) : For data series $X_{i1}, X_{i2}, \dots, X_{ir}$ from system *i*, there exist constants μ_i and $v_i^2 > 0$, such that we can construct the standard partial sum, defined as $C_i(t,r) \equiv \frac{\sum_{j=1}^{\lfloor rt \rfloor} X_{ij} - rt\mu_i}{v_i\sqrt{r}}, 0 \le t \le 1$. The probability distribution of $C_i(t,r)$ over [0, 1] converges to that of a standard Brownian motion process, for *t* on the unit interval, as *r* increases, *i.e.* $C_i(\cdot,r) \overset{d}{\Rightarrow} B(\cdot), r \to \infty$.

Under Assumption 4.1, μ_i can be identified as the steady-state mean for system *i* and v_i^2 is the asymptotic variance. The assumption also implies that for systems *i* and *l*, we have $C_i(\cdot,r) \overset{d}{\Rightarrow} B(\cdot)$, $C_l(\cdot,r) \overset{d}{\Rightarrow} B(\cdot), r \to \infty$. Billingsley (1968) provides some conditions under which Assumption 4.1 is satisfied and demonstrates that many stochastic processes satisfy Assumption 4.1.

In our procedure, we are calculating the variance estimator for the sample differences between the two systems. Thus we need to prove that for data series $Z_{il}(j) = X_{ij} - X_{lj}$ for $j = 1,2, \dots, r, i \neq l$, its standard partial sum $C_{il}(t,r) = \frac{\sum_{j=1}^{\lfloor rt \rfloor} Z_{il}(j) - rt(\mu_i - \mu_l)}{v_{il}\sqrt{r}}$ also approximates the Brownian motion process.

Lemma 4.2

For $i \neq l$, if system *i* and *l* satisfy Assumption 4.1 and are independent with each other, then there exists a constant $v_{il}^2 > 0$ such that $C_{il}(t,r) \overset{d}{\to} B(\cdot)$ as $r \to \infty$.

Proof:

$v_{il}^2 = \lim_{r\to\infty} rVar\left(\bar{Z}_{il}(r)\right) = \lim_{r\to\infty} rVar(\bar{X}_i - \bar{X}_l) = v_i^2 + v_l^2$ due to the independence between system *i* and *l*. We set $C_i(t,r)$ and $C_l(t,r)$ represent the standard partial sum for system *i* and *l*. Then the standard partial sum for the difference between these two systems is

$$C_{il}(t,r) = \frac{\sum_{j=1}^{\lfloor rt \rfloor} Z_{il}(j) - rt(\mu_i - \mu_l)}{v_{il}\sqrt{r}}$$
$$= \frac{\sum_{j=1}^{\lfloor rt \rfloor} (X_{ij} - X_{lj}) - rt(\mu_i - \mu_l)}{v_{il}\sqrt{r}}$$
$$= \frac{\sum_{j=1}^{\lfloor rt \rfloor} X_{ij} - rt\mu_i}{v_{il}\sqrt{r}} - \frac{\sum_{j=1}^{\lfloor rt \rfloor} X_{lj} - rt\mu_l}{v_{il}\sqrt{r}}$$
$$= \frac{v_i}{v_{il}} C_i(t,r) - \frac{v_l}{v_{il}} C_l(t,r)$$

Because systems *i* and *l* satisfy Lemma 4.2, we have $C_i(t,r) \xrightarrow{d} B_i(\cdot)$ and $C_l(t,r) \xrightarrow{d} B_l(\cdot)$. Also the two systems are independent with each other, so $B_i(\cdot)$ and $B_l(\cdot)$ are independent Brownian motion processes. Thus, according to (Billingsley 1968), $\big(C_i(t,r), C_l(t,r)\big) \to (B_i(\cdot), B_l(\cdot))$. By Continuous Mapping Theorem(CMT), we have $C_{il}(\cdot,r) \xrightarrow{d} \frac{v_i}{v_{il}} B_i(\cdot) - \frac{v_l}{v_{il}} B_l(\cdot)$ which is a standard Brownian motion process.

Remarks: v_{il}^2 is the asymptotic variance for data series $X_{ij} - X_{lj},\ j = 1, 2, \ldots$.

Lemma 4.2 demonstrates that the distribution of the cumulative differences between the sample outputs of two systems is approximately Brownian motion process. This fact will be used to justify our procedure in Theorem 4.1.

To set up the limiting environment required by Lemma 4.2, we can make the indifference zone parameter $\delta \to 0$, which leads to the hardest situation for ranking and selection problem. And naturally, the required number of observations $r \to \infty$.

While this PCS can be attained when sample size goes to infinity, we cannot guarantee it for finite sample sizes. That is the reason we will do empirical experiments to study how the procedure performs when the sample size is finite.

Following the above idea, we can assure the asymptotic probability of correct selection when comparing between two systems. This result can then be extended to selection of *k* best systems from *K* systems. The general idea is similar to that of Chapter 3. For system *i*, if we are confident that there are at least *k* systems better than it, it should be eliminated from the candidate pool right away to save observations and the problem becomes selection of *k* systems from *K*-1 systems; if we are confident there are at least *K*-*k* systems worse than system *i*, then it should be preselected as one of the best *k* systems and the problem becomes selecting *k*-1 systems from *K*-1 systems. The complete fully sequential procedure to select *k* best systems from *K* systems for steady-state simulation is given below.

Procedure (FS-*k*+):

Initialization: Select confidence level $1-\alpha$, indifference zone δ, batch size m_0 and first-stage sample size n_0. The first-stage number of batches is $b = \left\lfloor \frac{n_0}{m_0} \right\rfloor$.

Obtain n_0 observations $X_{ij},\ j = 1, 2, \ldots, n_0$, from each system *i* = 1,2,…, *K*. For all $i \neq l$, *i*, *l* = 1, 2,…, *K*, compute $m_0 V_{il}^2$, which is the sample asymptotic variance of the difference between alternatives *i* and *l*. And $m_0 V_{il}^2 \to v_{il}^2 \frac{\chi^2(d)}{d}$ as $n_0 \to \infty$, where *d* is the degree of freedom of the asymptotic variance estimator.

For batch mean estimator, $m_0 V_{B,il}^2 = \frac{m_0}{b-1}\sum_{j=1}^{b}(\bar{X}_{ij} - \bar{X}_{lj} - (\bar{X}_i(n_0) - \bar{X}_l(n_0)))^2$.

For area estimator, $\bar{X}_{i,j,h} = \frac{1}{h}\sum_{l=1}^{h} X_{i,(j-1)m_0+l}$ for *j=1, 2, …, b* and *h=1, 2, …, m*$_0$, $A_{il,j} = \frac{1}{m_0}\sum_{p=1}^{m_0} \omega\,(p/m_0)\,\frac{\mathrm{p}(\bar{X}_{i,j,m_0} - \bar{X}_{l,j,m_0} - (\bar{X}_{i,j,p} - \bar{X}_{l,j,\mathrm{p}}))}{\sqrt{m_0}}$, where $\omega(t) \equiv \sqrt{840}(3t^2 - 3t + 1/2)$. And asymptotic variance is estimated by $m_0 V_{A,il}^2 = \frac{1}{b}\sum_{j=1}^{b} A_{il,j}^2$.

Let $N_{il} = \left\lfloor \frac{h^2 m_0 V_{il}^2}{\delta^2} \right\rfloor$, where $\lfloor \cdot \rfloor$ indicates truncation of any fractional part, and let $N_i = \max_{l \neq i} N_{il}$.

Here $N_i + 1$ is the maximum number of observations that can be taken from alternative *i*. If $n_0 > \max_i N_i$, then stop and select the *k* systems with the largest $\bar{X}_i(n_0)$ as the best *k* systems.

Otherwise, let $I_0 = \{1, \ldots, K\}$ be the set of alternative designs. Create a directed graph $V = (I_0,\ E)$, where *E* represents the set of directed edges between nodes in set I_0. A directed edge in *E* from node *i* to node *j* is denoted by $e_{ij}: i \to j, i \in I_0, j \in I_0$. Set $E = \emptyset$. Node i_1 is considered a predecessor (successor) of node i_2 if there is a directed path in *V* from i_1 (i_2) to i_2 (i_1). Let A_i be the set of all the predecessors of node *i* and B_i the set of all the successors of node *i*. And let $G = \emptyset$ be the set of selected systems and $S = \emptyset$ be the set of eliminated systems. Let $h^2 = 2c\eta d$. For BM variance estimator *d*=*b*-1 and for AREA variance estimator *d*=*b*.

Set $I = I_0$, where *I* represents the set of alternatives still in contention, and the observation counter $r = n_0$. Go to **Screening and Preselection**.

Screening and Preselection: Let $W_{il}(r) = \max\{0, \frac{\delta}{2cr}(\frac{h^2 m_0 V_{il}^2}{\delta^2} - r)\}$, $\forall i, l \in I, i \neq l$. Let $m = |I|$ and $g = |G|$.

For $\forall i, l \in I, i \neq l$, if $\bar{X}_i(r) - \bar{X}_l(r) \geq W_{il}(r)$ and $e_{il} \notin E, e_{li} \notin E$, then $E = E \cup e_{il}$, $A_l = A_l \cup \{i\}$ and $B_i = B_i \cup \{l\}$. Then for $\forall i, l \in I, i \neq l$, define $S_{ij}^A = \{l: l \in A_i\} \cup \{i\}$, $S_{ij}^B = \{l: l \in B_j\} \cup \{j\}$. For $\forall l \in S_{ij}^A, t \in S_{ij}^B$, let $A_t = A_t \cup S_{ij}^A \backslash \{t\}$ and $B_l = B_l \cup S_{ij}^B \backslash \{l\}$.

For $\forall i \in I$, if $|B_i| \geq m - (k - g)$, then $G = G \cup \{i\}$. If $|A_i| \geq k - g$, then $S = S \cup \{i\}$. If either of these two conditions is true, then $I = I \backslash \{i\}, A_l = A_l \backslash \{i\}$ and $B_l = B_l \backslash \{i\}\ \forall l \in I$.

Stopping Rule: If $|G| \geq k$, stop the procedure and select *k* systems in *G* that have the largest overall sample means $\bar{X}_i(r) = \sum_{j=1}^{r} X_{i,j}$ as the best *k* systems;

If $|G| + |I| \leq k$, stop the procedure and select all the systems in *G* and *I*, together with $k - (|G| + |I|)$ system(s) in *S* that have the largest overall sample means as the best *k* systems;

Otherwise, take one additional observation $X_{i,r+1}$ from each alternative $i \in I$. Set *r* = *r* + 1 and go to **Screening and Preselection**.

Constants: The constant *c* may be nonnegative integer. The same as the last chapter, we still use *c* = 1. The constant η is the solution to the equation

$$g(\eta) \equiv \sum_{l=1}^{c} (-1)^{l+1} \left(1 - \frac{1}{2} I(l = c)\right) \left(1 + \frac{2\eta(2c - l)l}{c}\right)^{-\frac{d}{2}} = 1 - (1 - \alpha)^{\frac{1}{k(K-k)}},$$

where $I(\cdot)$ is the indicator function. When *c* = 1, we have the closed-form solution

$$\eta = \frac{1}{2}\left[\left(2 - 2(1 - \alpha)^{\frac{1}{k(K-k)}}\right)^{\frac{-2}{d}} - 1\right] \tag{4.2}$$

.

The procedure above is an extension of the procedure in Chapter 3 to autocorrelated outputs from steady-state simulation. When simulation outputs are i.i.d,

the two procedures are equivalent if we use BM estimator for the procedure in this chapter and set batch size as 1. Then the BM estimator is exactly calculating sample variance and using it as S_{il}^2. Also the degree of freedom of the variance estimator would be $d = b - 1 = n_0 - 1$. Thus the calculation of parameter h and η will be the same as Chapter 3. Different from the procedure in Chapter 3, we use estimated asymptotic variance of the difference between observations from two systems to replace the sample variance S_{il}^2. And S_{il}^2 follows a chi-square distribution with degrees of freedom $n_0 - 1$. Comparing with Chapter 3, *d* is calculated differently in this chapter according to what variance estimator is used.

Next we will provide the proof on the asymptotic result of our procedure.

Theorem 4.1 Denote $\mathrm{X}_i = \{X_{i1}, X_{i2}, \dots\}$. And $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_K$ represent the outputs from *K* systems. They are independent and satisfy Assumption 4.1 and $m_0 V_{il}^2$ is asymptotically distributed as $v_{il}^2 \chi_d^2 / d$ and is asymptotically independent of $C_{il}(\cdot, r)$, then $\liminf_{\delta \to 0} \Pr\{FK + selects\ the\ best\ k\ systems\} \geq 1 - \alpha$ given $\mu_K \geq \mu_{K-1} \geq \cdots \geq \mu_{K-k+1} \geq \mu_{K-k} + \delta \geq \cdots \geq \mu_1 + \delta$.

Proof:

We first define the same event $\mathcal{A}$ as the proof for Theorem 3.1 $\mathcal{A} = \{\text{for all } i \in \{K, K-1, \dots, K-k+1\} \text{ and } l \in \{1, \dots, K-k\}, e_{li} \notin E \text{ when the procedure stops}\}$. Based on the proof of Theorem 3.1, we have $\Pr\{CS\} \geq \Pr\{\mathcal{A}\}$. As described in Appendix, we can prove that for the ranking and selection between two systems, $\lim_{\delta \to 0} \sup \Pr\{ICS\} \leq 1 - (1 - \alpha)^{1/(k(K-k))}$.

When $K > 2$, and let ICS_{il} represent the event that an incorrect selection happens between a superior system $i, K - k + 1 \leq i \leq K$, and an inferior system $l, 1 \leq l \leq K - k$. Then,

$$\begin{aligned}
&\Pr\{\mathcal{A}\} \\
&\quad = 1 - \Pr\left\{\bigcup_{i=K-k+1}^{K} \bigcup_{l=1}^{K-k} (system\ l\ eliminates\ i)\right\} \\
&\quad \geq 1 - \sum_{i=K-k+1}^{K} \sum_{l=1}^{K-k} \Pr\{ICS_{il}\}
\end{aligned}$$

Then,

$$\lim_{\delta\to 0}\inf \Pr\{CS\}$$

$$\geq \lim_{\delta\to 0}\inf\left(1-\sum_{i=K-k+1}^{K}\sum_{l=1}^{K-k}\Pr\{ICS_{il}\}\right)$$

$$\geq \lim_{\delta\to 0}\inf \Pr\{\bigcap_{i=K-k+1}^{K}\bigcap_{l=1}^{K-k} CS_{il}\}$$

$$=$$

$$E\{\lim_{\delta\to 0}\inf \Pr\{\bigcap_{i=K-k+1}^{K}\bigcap_{l=1}^{K-k} CS_{il}\,|X_{i1},\dots,X_{i(N_i+1)},X_{l1},\dots,X_{l(N_l+1)},m_0V_{il}^2, i=K-k+1,\dots,K,\ l=1,\dots K-k\}\}$$

$$= E\{\lim_{\delta\to 0}\inf\prod_{i=K-k+1}^{K}\prod_{l=1}^{K-k}\Pr\{CS_{il}|X_{i1},\dots,X_{i(N_i+1)},X_{l1},\dots,X_{l(N_l+1)},m_0V_{il}^2\}\}$$

Events $\{CS_{il}|X_{i1},\dots,X_{i(N_i+1)},X_{l1},\dots,X_{l(N_l+1)},m_0V_{il}^2\}$, $i=K-k+1,\dots,K,\ l=1,\dots K-k$ are conditionally independent with each other. By using the argument which compares two systems and the fact that conditional probability is a bounded, continuous function of the condition, we have

$$\lim_{\delta\to 0}\inf \Pr\{CS\}$$

$$\geq E\{\lim_{\delta\to 0}\inf(\Pr\{CS_{il}|X_{i1},\dots,X_{i(N_i+1)},X_{l1},\dots,X_{l(N_l+1)},m_0V_{il}^2\})^{k(K-k)}\}$$

$$\geq (\lim_{\delta\to 0}\inf \Pr(CS_{il}))^{k(K-k)}$$

$$\geq (1-\lim_{\delta\to 0}\sup \Pr(CS_{il}))^{k(K-k)}$$

$$\geq 1-\alpha$$

Remarks: $m_0V_{il}^2$ is the estimated asymptotic variance for the differences between systems *i* and *l* based on the first-stage outputs. Thus, the value of $m_0V_{il}^2$ is dependent on the first n_0 observations. On the other hand, the term $C_{il}(\cdot,r)$ is shown to approximate a standard Brownian motion process, when $r\to\infty$, which means $C_{il}(\cdot,r)$ depends on the initial observations less and less as we collect more observations. Therefore, we can conclude that the asymptotic independence between $m_0V_{il}^2$ and $C_{il}(\cdot,r)$ stands.

4.4 Experiment Settings

We presented the procedure and its proof on asymptotic performance of the algorithm in the previous section. In this section and Section 4.4 we will present the experiment settings and the corresponding results for applying our procedure in finite sample cases.

We conduct a total of 344 different experiments. In each experiment, *K* systems are compared and *k* best systems are selected from those *K* systems. Data are simulated from two models:

AR(1): $X_{i,j} = \mu_i + \phi\left(X_{i,j-1} - \mu_i\right) + Z_{i,j}$, where $Z_{i,j} \overset{i.i.d}{\sim} N(0, 1-\phi^2)$

MA(1): $X_{i,j} = \mu_i + \theta Z_{i,j-1} + Z_{i,j}$, where $Z_{i,j} \overset{i.i.d}{\sim} N(0, \frac{1}{1+\theta^2})$

For these two models, the strength of dependence/correlation among all outputs from one system depend on ϕ and θ. We set parameters ϕ and θ both as 0.9 in our experiments to see how the method works for data with high autocorrelation.

For the configuration of steady state means for all alternative systems, we still use slippage configuration (SC) and monotonically decreasing means (MDM) as Chapter 3. For SC set-up, the true steady-state means among all alternatives are set as $\mu_K = \mu_{K-1} = \cdots = \mu_{K-k+1} = \mu_{K-k} + \delta = \cdots = \mu_1 + \delta$. Here, δ is the indifference zone parameter. SC is the hardest situation to do ranking and selection, because the difference between any pair of superior system and inferior system is the smallest value worth to detect, which is the indifference zone value. For MDM configuration, the true steady-state means are set as $\mu_i = \mu_{i-1} + \delta$. Under our settings, a system with a smaller steady-state mean is considered as better than one with a lager steady-state mean. Also, different from the setting in Chapter 3, the variance of the random noise is only related to the correlation parameter, ϕ for AR(1) model or θ for MA(1) model. Thus, among all the alternative systems in each experiment, all the output variances of the systems are equal.

The indifference zone parameter δ is set as $\sqrt{v_{K-k+1}^2/n_0}$ in this chapter, which is close to the standard deviation of first-stage sample mean of the k^{th} best system. To have

a good set-up of indifference zone parameter and understand completely about the comparison results, we need to know the true value of the variance parameter.

The true variance parameter v_{K-k+1}^2 for AR(1) and MA(1) models can be determined by certain formula described by Sargent (1992) and Batur (2009). Based on our settings of models, the variance parameter is calculated as:

for AR(1) data: $v_1^2 = v_2^2 = \cdots = v_{K-k+1}^2 = \cdots = v_K^2 = \frac{1+\varphi}{1-\varphi}$;

for MA(1) data: $v_1^2 = v_2^2 = \cdots = v_{K-k+1}^2 = \cdots = v_K^2 = 1 + \frac{2\theta}{1+\theta^2}$.

We set φ and θ as 0.9 and thus the variance parameter v_{K-k+1}^2 is known and the corresponding indifference zone parameter can be calculated.

Then we need to set up the number of first stage samples n_0, which is also set differently from Chapter 3. This is because the dataset we used in this chapter are auto-correlated and thus we need more data to have a good estimation of asymptotic variance. The criterion we applied to select the value of n_0 is to make sure that the difference between the asymptotic variance estimation based on the first stage samples $v^2(n_0) = n_0 var(\bar{X}(n_0))$ and the true value of asymptotic variance v^2 is very small, for example, close to $0.01v^2$. Thus for AR(1) we set n_0 as 1000 and for MA(1) we set it as 60.

Another parameter needs to be determined is the batch size m_0 for the first stage. We will look into the values of m_0 that are divisors of n_0 and keep at least four batches. Thus for AR(1) we have $m_0 = 40, 50, 100, 125, 200, 250$. And for MA(1) we have $m_0 = 10, 12, 15$.

For each experiment, 1000 replications are carried out. And the pre-specified probability of correct selection $1 - \alpha$ is set as 0.95.

Table 4.1 Parameters used in the experiments

Parameter	Meaning	Value
K	Number of alternative systems	10;20
k	Number of systems need to be selected	1;3;5
ϕ	Correlation parameter for AR(1)	0.9
θ	Correlation parameter for MA(1)	0.9
n_0	Number of first stage observations	AR(1):1000 MA(1): 60
m_0	Batch size for variance estimation	AR(1):40; 50; 100; 125; 200; 250 MA(a): 10; 12; 15
$1-\alpha$	Nominal probability of correct selection	0.95
v^2	Variance parameter	AR(1): 19 MA(1): 1.99

4.5 Experimental results

The experimental results are shown in this section. There are two criteria applied to compare the performance of procedure FK-*k*+ for different simulation outputs and different parameters used in the procedure. They are probability of correct selection (PCS) and the average number of samples per system (ANS).

We will discuss the performance of FS-*k*+ procedure in the next two subsections for AR(1) model and MA(1) model, respectively.

4.5.1 Experiment Results for AR(1) Model

From Table 4.2 to Table 4.5, all experiments with AR(1) model and *K*=10 are shown. The experiments under SC setting have larger (in most cases more than two times) ANS than those under MDM. Also the PCS is higher for MDM than SC. This is easy to understand because SC is corresponding to the worst-case scenario for the ranking and selection, while MDM represents a much easier case where we can easily eliminate many systems with much inferior means in the early stages. Next, we will present how the FS-k+ procedure performs for SC and MDM separately.

- **SCconfiguration:** Table 4.2 and Table 4.3 show the experiments under SC and the AR(1) model, the BM estimator needs 20%-96% more observations than the AREA estimator. As shown in Figure 4.1, for both estimators, the ANS increases quickly when m_0 is increased from 200 to 250. And ANS increases more slowly when m_0 is smaller than 200. At the same time, from Table 4.3, we can find that PCS is reduced when decreasing m_0 for both variance estimators. And the BM estimator has better PCS than the AREA estimator in most cases. Especially when m_0 is smaller than 50, the PCS of the AREA estimator drops to 0.6 or 0.7 and that of BM estimator is still about 0.9. To explore the reason why the PCS is low for the AREA estimator in some cases, we check values of $m_0 V_{il}^2$ using the sample size of the first-stage observations, based on the two estimators. As stated in the procedure FS-*k*+, $m_0 V_{il}^2 \to v_{il}^2 \chi^2(d)/d$ as $n_0 \to \infty$ and v_{il}^2 is the asymptotic variance for data series $X_{ij} - X_{lj},\ j = 1, 2, \ldots$..Then $E[m_0 V_{il}^2] \to v_{il}^2$. The true value for v_{il}^2 for is 38 based on our settings. We generate 500 replications of data outputs, and calculate the value of $m_0 V_{il}^2$ for each replication and estimate $E[m_0 V_{il}^2]$ by the average of them. We plot this estimation for both BM and AREA estimators for different m_0 values in Figure 4.2. We can find that the AREA estimator fails to have a reasonable estimation when m_0 is small. But both estimators have similarly good estimation when $m_0 \geq 200$. The increase of value k also causes the increase of ANS. When k is increased from 1 to 3 or from 3 to 5, there are both two units increase in k but the latter only cost less than one third of the former case, which indicates that the increasing of ANS is slower than the linear trend.

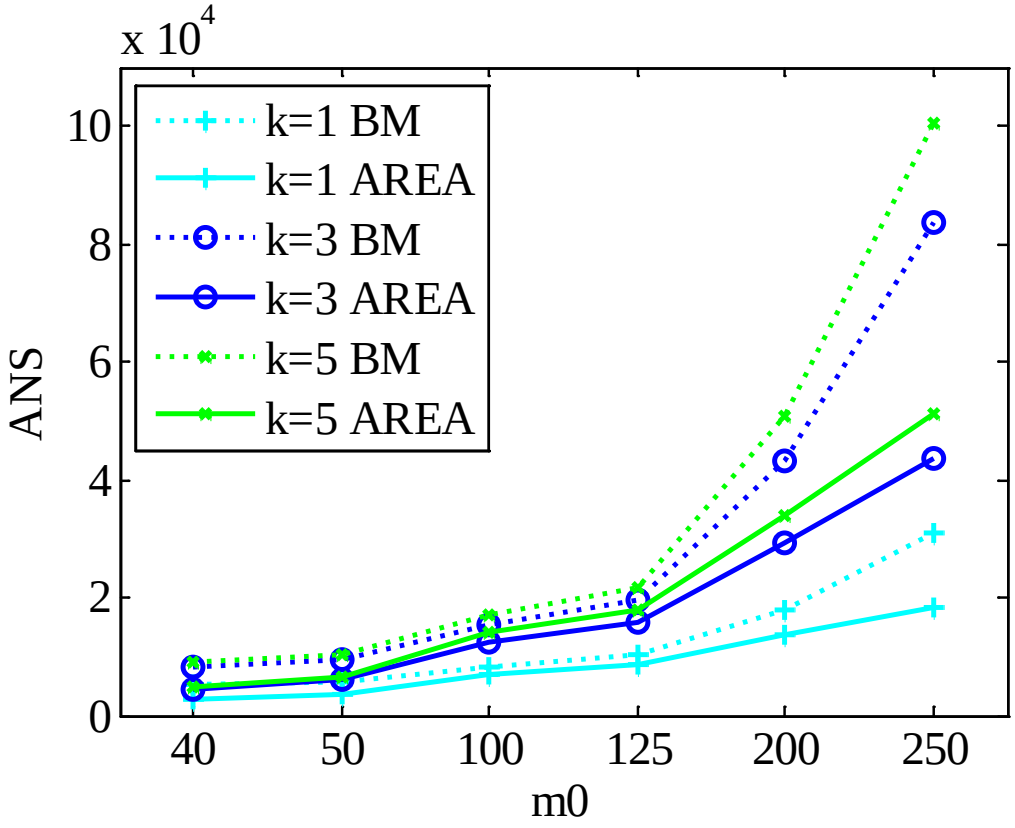

Figure 4.1 Plot of ANS for experiments with K=10, AR(1) model and SC

Table 4.2 Average number of observations per system (ANS) when AR(1) models are tested under SC configuration and *K*=10.

	k=1		*k*=3		*k*=5	
m_0	BM	AREA	BM	AREA	BM	AREA
250	30974.7	18530.32	83815.8	43758.25	100326.6	51044.18
200	18117.66	13600.94	43058.27	29330.91	50761.01	34132.86
125	10130.96	8585.16	19534.58	15675.96	21831.18	17980.31
100	8356.51	6995.63	15199.22	12611.56	17136.9	13961.69
50	5795.18	3542.65	9564.46	6014.01	10513.53	6638.79
40	5259.12	2684.97	8398.36	4536.05	9165.19	4998.41

Table 4.3 Estimated probability of correct selection (PCS) when AR(1) models are tested with SC configuration, K=10.

	k=1		k=3		k=5	
m_0	BM	AREA	BM	AREA	BM	AREA
250	0.958	0.96	0.951	0.938	0.946	0.959
200	0.965	0.961	0.962	0.951	0.942	0.951
125	0.949	0.942	0.949	0.927	0.942	0.941
100	0.953	0.928	0.937	0.921	0.944	0.919
50	0.935	0.798	0.928	0.749	0.926	0.733
40	0.901	0.755	0.912	0.64	0.894	0.619

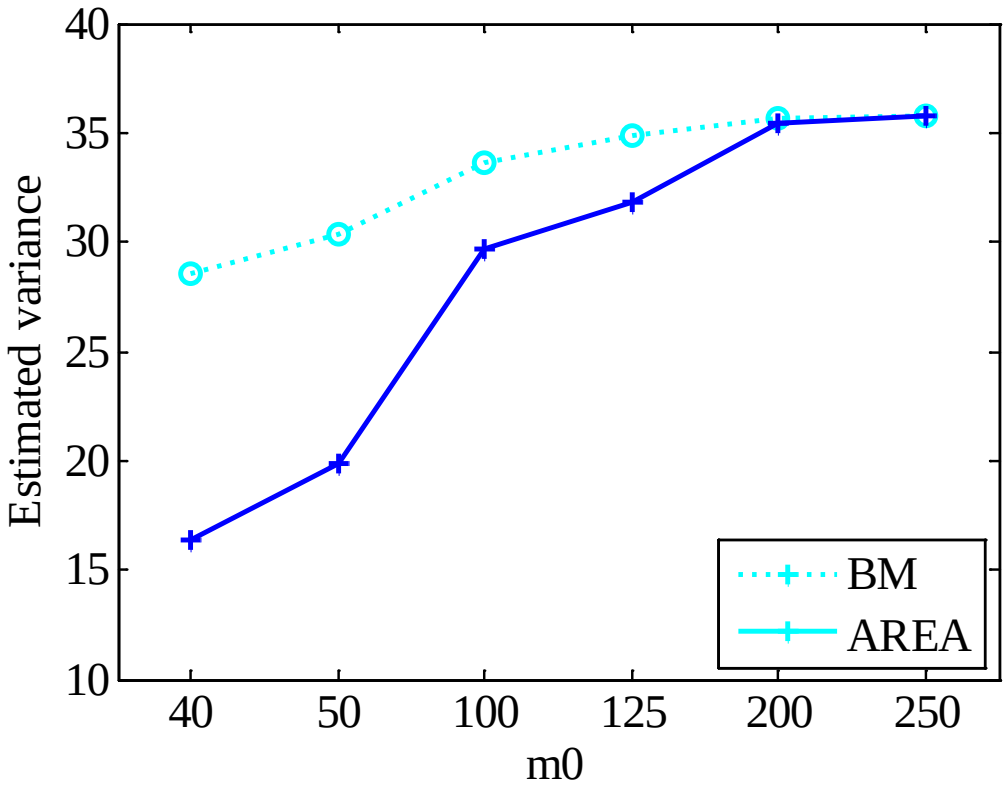

Figure 4.2 The average variance for the difference between two systems among different m0, when AR(1) model are used and K=10, k=3

- **MDM configuration:** Table 4.4 and Table 4.5 show the results for experiments under MDM. In terms of ANS, MDM costs much less than SC, which is about 43%-61% less. And the PCS are all over 0.95, except for one case with AREA estimator. When m_0 is larger than 125, PCS are all larger than 0.99 which is much higher than the nominal PCS 0.95. This is because the setting for our procedure, i.e. the value of η, is based on the SC configuration and is conservative for the MDM configurations. The relationship between ANS and m_0, as well as between ANS and *k* are the same as SC: ANS increases as m_0 increases and as *k* increases. Also, in the MDM cases, although the variance estimation is still not good when m_0 is small, the easier configuration leads to much better PCS.

Table 4.4 Average number of observations per system (ANS) when AR(1) models are tested with MDM configuration, *K*=10.

	k=1		*k*=3		*k*=5	
m_0	BM	AREA	BM	AREA	BM	AREA
250	12741.56	7983.89	33963.54	17571.04	39461.42	19943.48
200	7965.66	6075.01	17268.18	11908.53	20204.87	13607.16
125	4375.09	3830.23	7899.94	6422.52	8833.64	7196.32
100	3722.69	3092.25	6401.15	5143.14	7034.31	5697.31
50	2605.76	1815.61	3952.71	2566.96	4276.37	2750.51
40	2363.29	1536.68	3485.67	2016.19	3805.50	2146.82

Table 4.6 to Table 4.9 show the experiment results under AR(1) model when there are 20 systems total. Although the number of alternative systems is doubled, the estimated ANS is only increased by 27% on average for SC and AR(1) case, and even reduced by 12% for MDM and AR(1) case. This shows the increase of *K* value doesn't always cause extra observations per system. The effects of *k* and m_0 on ANS and the performance of variance estimation when *K*=20 are similar to the cases when *K*=10.

Table 4.5 Estimated PCS when AR(1) models are tested with MDM configuration, *K*=10.

	k=1		*k*=3		*k*=5	
m_0	BM	AREA	BM	AREA	BM	AREA
250	0.994	0.992	0.997	0.995	0.998	0.997
200	0.992	0.997	1	0.998	0.997	0.995
125	0.993	0.992	0.995	0.996	0.998	0.998
100	0.995	0.987	0.999	0.995	0.995	0.998
50	0.986	0.964	0.995	0.984	0.997	0.987
40	0.982	0.942	0.991	0.966	0.996	0.971

Table 4.6 Average number of observations per system (ANS) when AR(1) models are tested with SC configuration, *K*=20.

	k=1		*k*=3		*k*=5	
m_0	BM	AREA	BM	AREA	BM	AREA
250	41245.43	23123.23	128588.9	60010.2	191826.45	83811.45
200	23060.34	16549.24	59948.55	38032.29	83386.65	51526.25
125	11633.47	9774.83	23552.515	19222.28	30215.91	24117.25
100	9738.5	7956.82	18071.065	14617.795	22795.25	18139.46
50	6437.95	3885.13	10931.13	6776.125	13032.31	8178.22
40	5814.41	2926.78	9571.83	5018.245	11324.03	6094.37

Table 4.7 Estimated PCS when AR(1) models are tested with SC configuration, *K*=20.

	k=1		*k*=3		*k*=5	
m_0	BM	AREA	BM	AREA	BM	AREA
250	0.953	0.964	0.954	0.956	0.954	0.945
200	0.954	0.937	0.945	0.961	0.948	0.95
125	0.947	0.941	0.951	0.93	0.932	0.94
100	0.943	0.927	0.942	0.912	0.933	0.911
50	0.932	0.812	0.916	0.731	0.908	0.681
40	0.92	0.728	0.867	0.565	0.885	0.463

Table 4.8 Average number of observations per system (ANS) when AR(1) models are tested with MDM configuration, *K*=20.

	k=1		*k*=3		*k*=5	
m_0	BM	AREA	BM	AREA	BM	AREA
250	12024.02	6769.57	35354.94	16010.89	50461.45	22365.93
200	6741.97	4920.75	16210.31	10305.92	22041.88	13557.46
125	3532.55	3003.16	6438.15	5239.07	8102.53	6437.38
100	2865.89	2495.52	4981.23	4086.46	6118.99	4879.31
50	2081.18	1553.48	3010.06	2058.11	3470.86	2314.67
40	1919.34	1387.51	2680.99	1709.27	3071.5	1870.42

Table 4.9 Estimated PCS when AR(1) models are tested with MDM configuration, *K*=20.

	k=1		*k*=3		*k*=5	
m_0	BM	AREA	BM	AREA	BM	AREA
250	0.996	0.995	0.998	0.999	0.999	0.998
200	0.997	0.996	1	0.999	0.998	0.998
125	0.999	0.995	1	0.998	1	0.998
100	0.999	0.993	0.998	0.999	0.999	0.998
50	0.993	0.978	0.998	0.992	0.999	0.992
40	0.986	0.959	0.996	0.97	0.995	0.978

4.5.2 Experiment Results for MA(1) Model

Table 4.10 to Table 4.13 show the results for MA(1) models with *K*=10. One of the biggest differences in the results between AR(1) and MA(1) models is that the estimated ANS for MA(1) is only 10% of that of AR(1) on average, for both SC and MDM configurations. And all the PCS for MA(1) model are larger than 0.95 even for small m_0 values. In MDM configuration, all the PCS is even larger than 0.99. Unlike AR(1) models, the PCS are close to each other between the two variance estimators. One

of the reasons for the good results is that the true asymptotic variance for the difference between two systems is 3.8, which is much smaller than AR(1). This is also why we only need 60 first-stage observations to do a good estimation of the asymptotic variance. Table 4.14 shows that the estimated asymptotic variance is very close to the true value 3.8. There are also similar trends to AR(1) model. The ANS increases as m_0 increases or as k increases. The AREA estimator costs 50%-80% observations of BM estimator. And the SC configuration takes more observations than the MDM configuration.

Table 4.10 Average number of observations per system (ANS) when MA(1) models are tested with SC configuration, K=10.

	k=1		k=3		k=5	
m_0	BM	AREA	BM	AREA	BM	AREA
15	1873.723	1150.017	5147.33	2709.528	6147.357	3168.944
12	1101.151	868.7633	2567.051	1857.514	3045.074	2128.097
10	821.9294	712.6566	1798.564	1431.298	2027.651	1649.123

Table 4.11 Estimated PCS when MA(1) models are tested with SC configuration, K=10.

	k=1		k=3		k=5	
m_0	BM	AREA	BM	AREA	BM	AREA
15	0.954	0.953	0.958	0.955	0.957	0.957
12	0.961	0.949	0.951	0.952	0.957	0.953
10	0.959	0.958	0.961	0.967	0.955	0.961

Table 4.12 Average number of observations per system (ANS) when MA(1) models are tested with MDM configuration, K=10.

	k=1		k=3		k=5	
m_0	BM	AREA	BM	AREA	BM	AREA
15	775.61	492.3	2025.42	1114.4	2419.48	1207.33
12	476.37	380	1054.33	752.22	1212.62	827.87
10	369.73	312.06	727.71	579.30	823.02	652.4

Table 4.13 Estimated PCS when MA(1) models are tested with MDM configuration, K=10.

	k=1		k=3		k=5	
m_0	BM	AREA	BM	AREA	BM	AREA
15	0.992	0.994	0.995	0.999	0.998	0.995
12	0.994	0.995	0.998	0.995	0.997	0.998
10	0.992	0.994	0.995	0.996	0.998	0.995

Table 4.14 Estimated asymptotic variance for the difference between systems for MA(1) model

m_0	10		12		15	
Estimator	BM	AREA	BM	AREA	BM	AREA
Est. var ($m_0 V_{il}^2$)	3.7766	3.7934	3.7930	3.8325	3.8064	3.8863

Table 4.15 and Table 4.18 show the experiment results under MA(1) model and when there are 20 systems in total. Although the number of alternative systems is doubled, the estimated ANS is only increased by 40% for SC configuration and reduced by 6% for MDM configuration. This is consistent with the result under AR(1) model.

Table 4.15 Average number of observations per system (ANS) when MA(1) models are tested with SC configuration, K=20.

	k=1		*k*=3		*k*=5	
m_0	BM	AREA	BM	AREA	BM	AREA
15	2480.60	1442.34	7727.58	3697.51	11437.86	5176.67
12	1382.14	1035.79	3610.61	2382.46	5070.48	3181.91
10	1008.17	836.51	2322.08	1804.77	3114.37	2333.71

Table 4.16 Estimated PCS when MA(1) models are tested with SC configuration, *K*=20.

	k=1		*k*=3		*k*=5	
m_0	BM	AREA	BM	AREA	BM	AREA
15	0.963	0.942	0.969	0.973	0.95	0.947
12	0.952	0.961	0.951	0.965	0.968	0.955
10	0.958	0.966	0.945	0.952	0.951	0.955

Table 4.17 Average number of observations per system (ANS) when MA(1) models are tested with MDM configuration, *K*=20.

	k=1		*k*=3		*k*=5	
m_0	BM	AREA	BM	AREA	BM	AREA
15	742.56	422.43	2056.8	1007.86	3044.57	1364.16
12	403.32	302.63	984.38	652.62	1323.66	856.49
10	295.23	252.09	626.68	489.29	831.86	621.79

Table 4.18 Estimated PCS when MA(1) models are tested with MDM configuration, K=20.

	k=1		k=3		k=5	
m_0	BM	AREA	BM	AREA	BM	AREA
15	0.999	0.997	0.999	1	0.998	0.999
12	0.999	0.998	0.998	0.999	0.999	1
10	0.996	0.996	0.999	0.999	1	1

4.5.3 Summary of Experiment Results

We conduct hundreds of experiments to test the performance of FS-k+ procedure in this chapter. Parameters such as the variance estimators, the data generating models, the between-system steady-state mean configuration, k, K and m_0 are controlled. We find that our procedure, which is theoretically based on limiting process, can perform well in finite-sample scenario in most cases. BM estimator generally requires more observations than AREA estimator, but can usually assure higher PCS. So if the cost of observations is not as important as high probability of correct selection, BM estimator is preferred to AREA. The first-stage batch size m_0 is also influential to the results. Large m_0 requires more observations, but can assure higher PCS. This is another tradeoff between ANS and PCS.

4.6 Conclusion

In this chapter, we introduce the multi-best ranking and selection problem in steady-state simulation. And then we present the concept of asymptotic analysis and asymptotic variance, based on which we can extend the FS-k procedure for terminating simulation to solve R&S for steady-state simulations. The extended procedure FS-k+ is described and the theoretical proof on its asymptotic performance is provided. Then we

conduct hundreds of experiments to study the performance of FS-*k*+ and the influences of different parameter settings and data types for finite samples.

Based on our results, we can conclude that our procedure is effective in determining the multi-best systems for steady-state simulation. And when we set up the procedure, there is a trade-off between the number of observations and the probability of correct selection. These need to be decided by considering the cost of observations, importance to assure PCS, and other factors.

CHAPTER 5

CONCLUSION AND FUTURE WORK

In this thesis, we studied ranking and selection problems for both terminating and steady-state simulation. A ranking and selection procedure is to find the single one system or multiple systems with the best (largest/smallest) population mean(s). We designed two multi-best R&S procedures based on the fully sequential concept, one for terminating simulation, and the other one for steady-state simulation. We also conducted extensive experiments to study the procedures and compared their performances with existing methods. A good procedure should achieve two goals: assure the probability of correction selection and save number of samples/observations. Thus we consider these two as the criteria in evaluation of our approaches in the experiments of this thesis.

When we apply ranking and selection procedures to compare steady-state performance of systems, we should remove the initialization bias first. Initialization bias is caused by starting the sample generation from non-steady state, or transient state. The system would take a “warm-up” before it hits its steady-state. This “warm-up” period, in other words initial trend, needs to be removed. Otherwise it could cause bias if we estimate the steady-state mean by all the samples. Thus we designed a heuristic initialization bias elimination method called the cusum-rule in Chapter 2, which is motived by the offline change point detection methods in the statistics literature. We showed that the cusum-rule can perform at least as good as, in in a lot cases much better than other existing methods.

In Chapter 3, we proposed the multi-best ranking and selection method FS-*k* for terminating simulations. FS-*k* procedure is a fully sequential procedure that can sequentially select superior systems and eliminate the inferior ones at the same time. We provided theoretical proof on a lower bound of the probability of correc selection based on the proposed procedure. We have shown that this procedure can significantly save

observations compared with other multi-best R&S methods. In Chapter 3, we also consider the effect of the usage of common random numbers. The procedure is still valid when common random numbers are used. Our experiments showed that systems with positive correlations among their outputs can help save observations.

After studying multi-best R&S for terminating simulations, we looked into solving this problem for steady-state simulation. The FS-*k* procedure was extended to steady-state simulation by employing the concept of asymptotic variance. We described the procedure FS-*k*+ in detail in Chapter 4 and also gave the theoretical proof of its asymptotic performance. By conducting experiments with different data types and procedure settings, we could claim that our procedure is also effective in most finite-sample cases. Different variance estimators and different parameter values could influence results significantly. Through numerical experiments, we explained there is a trade-off between the required number of observations and the probability of correct selection for different set-up of the procedure.

In the future, we plan to study how the usage of common random numbers would affect the experiment results for multi-best R&S procedure for steady-state simulation. We need to check that if our FS-*k*+ procedure can be still be effective when common random numbers are used and if its effect is the same as in the terminating simulations. We would like to know how our multi-best ranking and selection procedures, both FS-*k* and FS-*k*+, perform when the data generating models are more complicated than the ones we used in the experiments for this thesis, e.g., when the distribution of the data is significantly deviated from normal distribution. And we would also want to find out how our initialization bias elimination method, the cusum-rule, as well as the ranking and selection algorithms, perform for complex simulation models in various real-world applications.

APPENDIX

In this appendix, we give the details about proof of $\lim_{\delta\to 0}\sup \Pr\{ICS\} \le 1-(1-\alpha)^{1/(k(K-k))}$ under the conditions given in Theorem 4.1.

As in the proof for Theorem 3.2, we still start with the case with two alternatives, system *i* and *l*, and $\mu_i \ge \mu_l + \delta$. The value of η is set as (4.2). And we want to show the probability of incorrect selection (ICS) between two systems satisfies:

$\liminf_{\delta\to 0} \Pr\{ICS\} \le 1-(1-\alpha)^{1/(k(K-k))}$.

When $N_{il}+1 > n_0$, we set $T(\delta) = \min\{r: r \ge n_0 \; and \; |\bar{X}_i(r) - \bar{X}_l(r)| \ge W_{il}(r)\}$. Here $T(\delta)$ represents the stage at which the procedure FS-*k*+ terminates.

$\Pr\{ICS\}$

$$= \Pr\left\{\bar{X}_i(r) - \bar{X}_l(r) \le W_{il}\big(T(\delta)\big)\right\}$$

$$= \Pr\left\{\sum_{j=1}^{T(\delta)}\big(X_{ij} - X_{lj}\big) \le -T(\delta) * W_{il}\big(T(\delta)\big)\right\}$$

$$= \Pr\left\{\sum_{j=1}^{T(\delta)}\big(X_{ij} - X_{lj}\big) \le -T(\delta) * \max\left\{0, \frac{\delta}{2cT(\delta)}\left(\frac{h^2 m_0 V_{il}^2}{\delta^2} - T(\delta)\right)\right\}\right\}$$

$$= \Pr\left\{\sum_{j=1}^{T(\delta)}\big(X_{ij} - X_{lj}\big) \le \min\left\{0, -\frac{h^2 m_0 V_{il}^2}{2c\delta} + \frac{\delta T(\delta)}{2c}\right\}\right\}$$

$$= \Pr\left\{\sum_{j=1}^{T(\delta)}\big(X_{ij} - X_{lj}\big) - (\mu_i - \mu_l)T(\delta) + (\mu_i - \mu_l)T(\delta) \le \min\left\{0, -\frac{h^2 m_0 V_{il}^2}{2c\delta} + \frac{\delta T(\delta)}{2c}\right\}\right\}$$

$$= Pr\{\left(\sum_{j=1}^{T(\delta)}\big(X_{ij} - X_{lj}\big) - (\mu_i - \mu_l)T(\delta)\right)/v_{il}\sqrt{N_{il}+1} + (\mu_i - \mu_l)T(\delta)/v_{il}\sqrt{N_{il}+1} \le \min\left\{0, -\frac{h^2 m_0 V_{il}^2}{2c\delta v_{il}\sqrt{N_{il}+1}} + \frac{\delta T(\delta)}{2cv_{il}\sqrt{N_{il}+1}}\right\}\}$$

$$\le \Pr\{(\sum_{j=1}^{T(\delta)}\big(X_{ij} - X_{lj}\big) - (\mu_i - \mu_l)T(\delta))/v_{il}\sqrt{N_{il}+1} + \delta T(\delta)/v_{il}\sqrt{N_{il}+1} \le \min\{0, -\frac{h^2 m_0 V_{il}^2}{2c\delta v_{il}\sqrt{N_{il}+1}} + \frac{\delta T(\delta)}{2cv_{il}\sqrt{N_{il}+1}}\}\}$$

$$= E\{\Pr\{(\sum_{j=1}^{T(\delta)}\big(X_{ij} - X_{lj}\big) - (\mu_i - \mu_l)T(\delta))/v_{il}\sqrt{N_{il}+1} + \delta T(\delta)/v_{il}\sqrt{N_{il}+1} \le \min\{0, -\frac{h^2 m_0 V_{il}^2}{2c\delta v_{il}\sqrt{N_{il}+1}} + \frac{\delta T(\delta)}{2cv_{il}\sqrt{N_{il}+1}}\}|m_0 V_{il}^2\}\}$$

Then inequality happens because $\mu_i - \mu_l \ge \delta$ and we replace $\mu_i - \mu_l$ by δ.

When $N_{il}+1 \le n_0$, in other words, $\frac{h^2 m_0 V_{il}^2}{\delta^2} < n_0$, we have

$W_{il}(n_0) = \max\left\{0, \frac{\delta}{2cn_0}\left(\frac{h^2 m_0 V_{il}^2}{\delta^2} - n_0\right)\right\} = 0.$

$$\Pr\{ICS\} = \Pr\{\bar{X}_i(n_0) \le \bar{X}_l\}$$
$$= \Pr\{\bar{X}_i(n_0) - \bar{X}_l \le 0\}$$
$$= \Pr\{\sum_{j=1}^{n_0}(X_{ij} - X_{lj}) \le -n_0 * W_{il}(n_0)\}$$

Thus, the result in $N_{il} + 1 > n_0$ case can also stand when $N_{il} + 1 \le n_0$ by making $T(\delta) = n_0$.

Further step, we want to prove that $\lim_{\delta\to 0} \sup \Pr\{ICS\} \le 1 - (1-\alpha)^{1/(k(K-k))}$.

For this purpose, we first construct a standardized partial sum

$$C_{il}(t,\delta) = \frac{\sum_{j=1}^{\lfloor (N_{il}+1)t \rfloor}(X_{ij} - X_{lj}) - (N_{il}+1)(\mu_i - \mu_l)t}{v_{il}\sqrt{N_{il}+1}}, \text{ where } 0 \le t \le 1.$$

C_{il} can be interpreted as a function of δ because N_{il} is a function of δ. Then we define an estimation of $T(\delta)$:

$$\hat{T}(\delta) = \min\{t \in \left\{\frac{n_0}{N_{il}+1}, \frac{n_0+1}{N_{il}+1}, \dots, 1\right\}:$$
$$\left|C_{il}(t,\delta) + \delta t(N_{il}+1)/v_{il}\sqrt{N_{il}+1}\right| \ge \frac{h^2 m_0 V_{il}^2}{2c\delta v_{il}\sqrt{N_{il}+1}} - \frac{\delta t(N_{il}+1)}{2cv_{il}\sqrt{N_{il}+1}}\}.$$

Actually $\hat{T}(\delta) = T(\delta)/(N_{il}+1)$. Also, a stopping time of the corresponding continuous-time process is defined as:

$$\tilde{T}(\delta) = \min\{t \ge \frac{n_0}{N_{il}+1}:$$
$$\left|C_{il}(t,\delta) + \delta t(N_{il}+1)/v_{il}\sqrt{N_{il}+1}\right| \ge \frac{h^2 m_0 V_{il}^2}{2c\delta v_{il}\sqrt{N_{il}+1}} - \frac{\delta t(N_{il}+1)}{2cv_{il}\sqrt{N_{il}+1}}\}.$$

We can find for one fixed δ, $\hat{T}(\delta)(N_{il}+1)$ is the smallest integer greater than or equal to $\tilde{T}(\delta)\ (N_{il}+1)$. Because as $\delta \to 0$, $\frac{1}{N_{il}+1} \to 0$, $\hat{T}(\delta) \to \tilde{T}(\delta)$. Thus in the limit we can focus on continuous version $C_{il}(\tilde{T}(\delta), \delta)$.

According to Lemma 4.1 (Fabian 1974), Lemma 4.2 and assumptions, we can have:

$C_{il}(t,\delta) + \delta t(N_{il}+1)/v_{il}\sqrt{N_{il}+1} \to B(t,\Delta)$ as $\delta \to 0$, conditioning on $m_0 V_{il}^2$

where $\Delta = \lim_{\delta\to 0} \delta(N_{il}+1)/v_{il}\sqrt{N_{il}+1} = \frac{\sqrt{h^2 m_0 V_{il}^2}}{v_{il}}$.

Then to satisfy the requirement of Lemma 4.1 (Fabian 1974), we let

$$\mathcal{A}(\delta) = \frac{h^2 m_0 V_{il}^2}{2c\delta v_{il}\sqrt{N_{il}+1}} \xrightarrow{\delta\to 0} \frac{\sqrt{h^2 m_0 V_{il}^2}}{2cv_{il}} \equiv \mathcal{A},$$
$$\mathcal{B}(\delta) = \frac{\delta(N_{il}+1)}{2cv_{il}\sqrt{N_{il}+1}} \xrightarrow{\delta\to 0} \frac{\sqrt{h^2 m_0 V_{il}^2}}{2cv_{il}} \equiv \mathcal{B}.$$

Thus the continuous stopping time $\tilde{T}(\delta)$ is the smallest *t* subject to

$$\left|C_{il}(t,\delta) + \delta t(N_{il}+1)/v_{il}\sqrt{N_{il}+1}\right| \ge \mathcal{A}(\delta) - \mathcal{B}(\delta)t$$

Define a mapping function $f_\delta: D[0,1] \to \mathcal{R}$ such that

$f_\delta(Y) = Y\left(T_{Y,\delta}\right)$, where $T_{Y,\delta} = \inf\{t: |Y(t)| - \mathcal{A}(\delta) + \mathcal{B}(\delta)t \geq 0\}$ for every $Y(t) \in D[0,1]$ and $\delta > 0$. So function $f_\delta(Y)$ is the value of $Y(t)$ at $\inf\{t: |Y(t)| - \mathcal{A}(\delta) + \mathcal{B}(\delta)t \geq 0\}$. Similarly, define $f(Y) = Y(T_Y)$ where $T_Y = \inf\{t: |Y(t)| - \mathcal{A} + \mathcal{B}t \geq 0\}$ for every $Y(t) \in D[0,1]$ and $\delta > 0$.

Note that

$f_\delta\left(C_{il}(t,\delta) + \delta t(N_{il}+1)/v_{il}\sqrt{N_{il}+1}\right) = C_{il}\left(\tilde{T}(\delta),\delta\right) + \delta\tilde{T}(\delta)(N_{il}+1)/v_{il}\sqrt{N_{il}+1}$

and $f\left(B(\cdot,\Delta)\right) = B(T_{B(\cdot,\Delta)},\Delta)$ by making $Y = B(\cdot,\Delta)$.

Denote $\mathcal{Y}_{il}(t,\delta) \equiv C_{il}(t,\delta) + \delta t(N_{il}+1)/v_{il}\sqrt{N_{il}+1}$ where $t \in [0,1]$ and $\delta > 0$.

Then according to (Kim 2005) and (Billingsley 1968), $f_\delta\left(\mathcal{Y}_{il}(t,\delta)\right) \to f\left(B(\cdot,\Delta)\right)$ as $\delta \to \infty$.

Now unconditioning on $m_0 V_{il}^2$ we have

$$\begin{aligned}
&\limsup_{\delta\to 0} \Pr\{ICS\} \\
&\leq E[\Pr\{B(t,\Delta)\text{exits continuation region through the lower boundary}|m_0V_{il}^2] \\
&= E[\textstyle\sum_{l=1}^{c}(-1)^{l+1}\left(1-\frac{1}{2}I(l=c)\right)\exp\{-2\mathcal{A}\mathcal{B}(2c-l)l\}] \text{ (By Lemma2)} \\
&= E[\textstyle\sum_{l=1}^{c}(-1)^{l+1}\left(1-\frac{1}{2}I(l=c)\right)\exp\left\{-2\frac{h^2m_0V_{il}^2}{(2c)^2v_{il}^2}(2c-l)l\right\}] \\
&= E[\textstyle\sum_{l=1}^{c}(-1)^{l+1}\left(1-\frac{1}{2}I(l=c)\right)\exp\left\{-\frac{\eta d m_0V_{il}^2}{cv_{il}^2}(2c-l)l\right\}] \text{ (By } h^2 = 2c\eta d) \\
&= \textstyle\sum_{l=1}^{c}(-1)^{l+1}\left(1-\frac{1}{2}I(l=c)\right)E[\exp\left\{-\frac{\eta d m_0V_{il}^2}{cv_{il}^2}(2c-l)l\right\}] \\
&= \textstyle\sum_{l=1}^{c}(-1)^{l+1}\left(1-\frac{1}{2}I(l=c)\right)(1+\frac{2\eta(2c-l)l}{c})^{-d/2} \\
&= 1-(1-\alpha)^{1/(k(K-k))}
\end{aligned}$$

The fifth equality comes from the moment generating function of chi-square distribution. $E[\exp(t\chi_v^2)] = (1-2t)^{-v/2}$ for $t < 1/2$ and χ_v^2 is a chi-squared random variable with v degrees of freedom.

We have proved that $\lim_{\delta\to 0}\sup \Pr\{ICS\} \leq 1-(1-\alpha)^{1/(k(K-k))}$.

REFERENCES

Anderson, T.W. and D.A Darling. 1954. A test of goodness of fit. *Journal of the American Statistical Association* 49 (765–769).

Banks, J., J. Carson, B.L Nelson, and D. Nicol. 2004. *Discrete-Event System Simulation*. 4 ed: Prentice Hall.

Barton, R. R. and L. W. Schruben. 1989. Graphical methods for the design and analysis of simulation experiments. Paper read at WSC '89 Proceedings of the 21st conference on Winter simulation

Batur, D., Goldman, D., Kim, SH. 2009. An improved standardized time series Durbin-Watson variance estimator for steady-state simulation. *Operations Research Letters* 37:285-289.

Bechhofer, R.E. 1954a. A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with known Variances. *Annals of Mathematical Statistics* 25 (1):16-39.

Bechhofer, R.E., C.W Dunnett, and M. Sobel. 1954b. A Two-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with a Common Unknown Variance. *Biometrika* 41:170-176.

Billingsley, P. 1968. *Convergence of Probability Measures*. New York: John Wiley and Sons.

Bofinger, E. and K. Mangersen. 1986. Subset selection of the t best populations. *Communications in Statistics - Theory and Methods* 15 (10):3145-3161.

Brooks, S.P. 1998. Quantitative convergence assessment for Markov chain Monte Carlo via cusums. *Statistics and Computing* 8:267-274.

Carroll, R.J, S.S. Gupta, and D.Y. Huang. 1975. On Selection Procedures for the t Best Populations and some Related Problems. *Comm. Statist.* 4:987-1008.

Cash, C.R., D.G. Dippold, J.M. Long, and W.P. Pollard. 1992. Evaluation of tests for initial-condition bias. *Proceedings of the winter simulation conference*:575-585.

Cash, C.R., Dippold, D.G., Long, J.M. and Pollard, W.P. 1992. Evaluation of tests for initial-condition bias. *Proceedings of the winter simulation conference*:575-585.

Damerdji, H., and M.K. Nakayama. 1999. Two-Stage Multiple-Comparison Procedures For Steady-State Simulations. *ACM Transactions on Modeling and Computer Simulation* 9 (1):1-30.

Dudewicz, E.J., and S.R. Dalal. 1975. Allocation of Observations in Ranking and Selection with Unequal Variances. *Sankhyā: The Indian Journal of Statistics* 37 (1):28-78.

Dudewicz, E.J., and N.A. Zaino. 1977. Allowance for correlation in setting simulation run-length via ranking-and-selection procedures. *TIMS Studies in teh Management Sciences* 7:51-61.

Fabian, V. 1974. Note on Anderson's Sequential Procedures with Triangular Boundary. *The Annals of Statistics* 2 (1):170-176.

Fishman, G.S. 1972. Bias considerations in simulation experiments. *Operations Research* 20:785-790.

Fishman, G.S. 1978. Grouping observations in digital simulation. *Management Science* 24:510–521.

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

Glynn, P.W. and W. Whitt. 1991. Estimating the asymptotic variance with batch means. *Operations Research Letters* 10:431-435.

Goldsman, D. 1983. Ranking and selection in simulation. *WSC '83 Proceedings of the 15th conference on Winter Simulation* 2:387-394.

Goldsman, D. 1985. Ranking and selection procedures using standardized time series. *WSC '85 Proceedings of the 17th conference on Winter simulation*:120-124.

Goldsman, D., S. Kim, W.S. Marshall, and B.L. Nelson. 2002. Ranking and Selection for Steady-State Simulation: Procedures and Perspectives. *INFORMS Journal on Computing* 14 (1):2-19.

Goldsman, D., M. Meketon, and L. Schruben. 1990. Properties of Standardized Time Series Weighted Area Variance Estimators. *Management Science* 36 (5):602-612.

Goldsman, D., L.W. Schruben, and J.J. Swain. 1994. Tests for transient means in simulated time series. *Naval Research Logistics* 41:171-187.

Gupta, S.S. 1965. On Some Multiple Decision (Selection and Ranking) Rules. *Technometrics* 7 (2):225-245.

Hoad, K., S. Robinson, and R. Davies. 2008. Automating warm-up length estimation. *Proceedings of the 2008 Winter Simulation Conference*:532 - 540.

Hoad, K., S. Robinson, and R. Davies. 2010. Automating warm-up length estimation. *Journal of the Operational Research Society* 61:1389-1403.

Kiekhaefer, A.P. 2011. Simulation Ranking and Selection Procedures and Applications in Network Reliability Design. PhD thesis.

Kim, S. and B.L. Nelson. 2001. A Fully Sequential Procedure for Indifference-Zone Selection in Simulation. *ACM Transactions on Modeling and Computer Simulation* 11 (3):251-273.

Kim, S. and B.L. Nelson. 2005. Chapter 17: Selecting the Best System. In *Elsevier Science*.

Kim, S. and B.L. Nelson. 2006. On the Asymptotic Validity of Fully Sequential Selection Procedures for Steady-State Simulation. *Operations Research* 54 (3):475-488.

Kim, S., Nelson, B. L., Wilson, J. R. 2005. Some almost-sure convergence properties useful in sequential analysis. *Sequential Anal.* 24 (4):411-419.

Law, A.M. and W.D. Kelton. 2000. Simulation Modeling and Analysis. *third ed. McGraw-Hill, New York.*

Law, A.M. and W.D. Kelton. 2000. *Simulation Modeling and*

Analysis. 3 ed. New York: McGraw-Hill.

Meketon, M.S. and B. Schmeiser. 1984. Overlapping batch means: something for nothing? *WSC '84 Proceedings of the 16th conference on Winter simulation*:226-230.

Montgomery, D.C. 2004. *Introduction to statistical quality control*. 5 ed: Wiley.

Montgomery, D.C., Runger, G.C. 1994. Applied Statistics and Probability for Engineers. *Wiley, New York.*

Nakayama, M.K. 1995. Selecting the best system in steady-state simulations using batch means. *Simulation Conference Proceedings*:362-366.

Nakayama, M.K. 1997. Multiple-comparison procedures for steady-state simulations. *Annals of Statistics* 25 (6):2433-2450.

Nelson, B. L., J. Swann, D. Goldsman, and W. Song. 2001. Simple Procedures for Selecting the Best Simulated System When the Number of Alternatives is Large. *Operations Research* 49 (6):950-963.

Page, E. S. 1955. Control Charts With Warning Lines. *Biometrics* 42:243-257.

Pasupathy, R. and B. Schmeiser. 2010. The initial transient in steady-state point estimation: Contexts, a bibliography, the mse criterion, and the MSER statistic. Paper read at Proceedings of the 2010 Winter Simulation Conference.

Rinott, Y. 1978. On Two-Stage Selection Procedures and Related Probability-Inequalities. *Communications in Statistics-Theory and Methods* 7 (8):799-811.

Robinson, S. 2002. A statistical process control approach for estimating the warm-up period. *Proceedings of the 2002 Winter Simulation Conference* 1:439 - 446.

Robinson, S. 2004. *Simulation: The Practice of Model Development and Use*. Chichester, UK: Wiley.

Robinson, S. 2007. A statistical process control approach to selecting a warm-up period for a discrete-event simulation. *European Journal of Operational Research* 176 (1):332–346.

Sargent, R.G., Kang, K., Goldsman, D. 1992. An investigation of finite-sample behavior of confidence interval estimators. *Operations Research* 40 (5):898-913.

Schmeiser, B. and W. Song. 1989. Inverse-transformation algorithms for some common stochastic processes. *Simulation Conference Proceedings*:490-496.

Schruben, L.W. 1982. Detecting initialization bias in simulation output. *Operations Research* 30 (3):569-590.

Schruben, L.W. 1983. Confidence interval estimation using standardized time series. *Operations Research* 31:1090-1108.

Snell, M. and L. Schruben. 1985. Weighting simulation data to reduce initialization effects. *IIE Transactions* 17 (4):354-363.

Spratt, S. C. 1998. Heuristic for startup problem, Department of Systems Engineering, University of Virginia.

Steiger, N.M., and J.R. Wilson. 2002. An Improved Batch Means Procedure for Simulation Output Analysis. *Management Science* 48 (12):1569-1586.

Swisher, J.R., P.D. Hyden, S.H. Jacobson, and L.W. Schruben. 2004. A survey of recent advances in discrete input parameter discrete-event simulation optimization. *IIE Transactions* 36:591-600.

Swisher, J.R., S.H. Jacobson, and E. Yucesan. 2003. Discrete-event simulation optimization using ranking, selection, and multiple comparison procedures: A survey. *ACM Transactions on Modeling and Computer Simulation* 13 (2):134-154.

Vassilacopoulos, G. 1989. Testing for initialization bias in simulation output. *Simulation* 52 (4):151-153.

White, K.P., Jr. 1995. A Simple Rule for Mitigating Initialization Bias in Simulation Output: Comparative Results. *International Conference on Systems, Man, and Cybernetics*:206-211.

White, K.P., Jr. 1997. An effective truncation heuristic for bias rduction in smulation output. *Simulation* 69 (6):323-334.

White, K.P., Jr., and S.C. Spratt. 2000. A comparison of five steady-state truncation heuristics for simulation. *Proceedings of the 2000 winter simulation conference*:755-760.

Yu, B. 1996. Comment: Extracting More Diagnostic Information from a Single Run Using Cusum Path Plot. *Statistical Science* 10 (1):54-58.

Yu, B. and P. Mykland. 1998. Looking at Markov samplers through cusum path plots: a simple diagnostic idea. *Statistics and Computing* 8:275-286.