A NEW SELECTION PROCEDURE FOR LARGE SCALE PROBLEMS

MOHAMMAD HANI ALMOMANI

UNIVERSITI SAINS MALAYSIA

2012

A NEW SELECTION PROCEDURE FOR LARGE SCALE PROBLEMS

by

MOHAMMAD HANI ALMOMANI

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosphy

April 2012

ACKNOWLEDGEMENTS

I am grateful to my supervisor Dr. Rosmanjawati Abdul Rahman for her guidance, valuable advice, suggestions and encouragement during the preparation of this work. My thanks also go to Associate Professor Adam Baharum for his support and valuable contributions as co-supervisor.

I would also like to take this opportunity to thank the Dean of the School of Mathematical Sciences, Universiti Sains Malaysia, Professor Ahmad Izani Md. Ismail, lecturers and staff of the department for their kind advice and support which has helped me throughout the course of my work.

I would like to extend special thanks to my family for their support and prayers at all times, which were particularly important, which provided a constant source of inspiration to me. In addition, I also wish to dedicate this thesis to my beloved wife Wesam and my son Saif. I am greatly indebted to their enthusiasm and strong support.

This work would not have come to exist without the full support of Dr. Rosmanjawati Abdul Rahman and her valuable advice at all stages. She taught me how to write an academic paper, inspirited good ideas in me, and enhanced my confidence facing challenges. I would like to thank her again for the time she spent guiding me, and for all her efforts.

I thank Associate Professor Mahmoud Alrefaei for his support and help in the buffer allocation problem. Lastly, thanks to all my friends, especially those who have provided helpful suggestions and encouragements. But most of all, I would like to thank God for all his help and blessings.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DEDS Discrete Event Dynamic Systems

CS Correct Selection

P(CS) Probability of Correct Selection

E(OC) Expected Opportunity Cost of a Potentially Incorrect Selection

MR New Proposed Selection Approach

OO Ordinal Optimization

OCBA Optimal Computing Budget Allocation

SS Subset Selection

IZ Indifference-Zone

MCA All Pairwise Multiple Comparisons

MCC Multiple Comparisons with a Control

MCB Multiple Comparisons with the Best

BAP Buffer Allocation Problem

WIP Average Work in Process

VOO Vector Ordinal Optimization

MOCBA Multi Objective Computing Budget Allocation

LIST OF SYMBOLS

Θ Feasible solution set

n Number of elements in feasible solution set

$\bar{y}^{(1)}$ First sample mean

$\bar{y}^{(2)}$ Second sample mean

λ Mean arrival rate

μ Mean service rate

$|A|$ Number of elements in the set A

δ^* Indifference zone

t_0 Initial sample size

Δ Increment in simulation samples

B Total simulation budget

T Elapsed time

S Sequential stopping rule

$P(GS)_{\delta^*}$ Probability of good selection stopping rule

Q Buffer spaces

q Number of intermediate buffers

$P(x)$ Production rate

LIST OF PUBLICATIONS

Chapter 3:

Almomani, M. H. and Abdul Rahman, R. (2012). Selecting a good stochastic system for the large number of alternatives, *Communications in Statistics–Simulation and Computation* **41**(2): 222–237.

Almomani, M. H. and Abdul Rahman, R. (2010). Selecting a good enough simulated design with opportunity cost, *International Journal of Mathematics and Computation* **7**(J10): 114–124.

Almomani, M. H. and Abdul Rahman, R. (2011). Sequential selection approach for selecting the best system, *International Conference on Modeling, Simulation and Applied Optimization (ICMSAO2011)*, pp. 1056–1060.

Chapter 4:

Almomani, M. H. and Abdul Rahman, R. (2011). Efficiency of different stopping rules for selecting a good system, *World Applied Science Journal* **12**(8): 1327–1336.

Almomani, M. H. and Abdul Rahman, R. (2011). The effect of increment in simulation samples on a combined selection procedure, *International Conference on Computer Mathematics and Natural Computing (ICCMNC 2011)*, Vol. 74, pp. 463–467.

Almomani, M. H. and Abdul Rahman, R. (2011). An adequate choice of initial sample size for selection approach, *International Conference on Operations Research (ICOR2011)*, Vol. 60, pp. 799–803.

Almomani, M. H., Abdul Rahman, R. and Baharum, A. (2011). Four–stage selection approach with the initial sample size, *The 2nd symposium of the USM fellowship holders 2011*.

Chapter 5:

Almomani, M. H., Abdul Rahman, R. and Baharum, A. (2011). Selecting a good enough stochastic design, *Far East Journal of Applied Mathematics* **53**(2): 123–131.

Almomani, M. H., Baharum, A. and Abdul Rahman, R. (2011). Three–stage selection approach with the initial simulation sample size, *International Journal of Pure and Applied Mathematics* **72**(2): 159–172.

Almomani, M. H. and Abdul Rahman, R. (2012). Selecting the best system using the three–stage and the four–stage selection approaches, *Applied Mathematical Sciences* **6**(40): 1955–1972.

Almomani, M. H., Baharum, A. and Abdul Rahman, R. (2010). Three–stage procedure and expected opportunity cost for selecting a good simulated design, *Research Journal of Applied Sciences* **5**(6): 417–423.

Almomani, M. H., Abdul Rahman, R. and Baharum, A. (2010). Selecting a good enough stochastic design, *International Conference on Fundamental & Applied Sciences (ICFAS2010)*.

Almomani, M. H., Baharum, A. and Abdul Rahman, R. (2010). The opportunity cost and three–stage approach, *International Conference on Mathematical Applications in Engineering (ICMAE2010)*.

Almomani, M. H., Baharum, A. and Abdul Rahman, R. (2010). Three–stage approach and the initial sample size, *Association of Asian Pacific Operational Research Societies Conference (APORS2010)*.

Chapter 6:

Almomani, M. H., Abdul Rahman, R., Baharum, A. and Alrefaei, M. H. (2012). A selection approach for solving the buffer allocation problem, *International Journal of the Physical Sciences* **7**(3): 413–422.

PROSEDUR PEMILIHAN BARU BAGI MASALAH BERSKALA BESAR

ABSTRAK

Tesis ini mempertimbangkan masalah pemilihan sistem stokastik yang mempunyai jangkaan ukuran prestasi terbaik (maksimum atau minimum) apabila bilangan alternatif adalah terhingga dan besar. Prosedur isih dan pilih telah jayanya digunakan untuk penyelesaian masalah dengan bilangan alternatif yang kecil. Untuk tujuan mengurangkan masalah pengiraan, idea pengoptimuman ordinal digunakan dengan objektif untuk menemui satu sistem yang memadai baik dengan tidak semestinya mencari sistem yang terbaik. Dalam tesis ini, satu pendekatan prosedur baru dicadangkan untuk memilih satu sistem yang baik apabila bilangan alternatif adalah sangat besar. Pendekatan ini mengandungi empat peringkat; Dalam peringkat pertama, prosedur pengoptimuman ordinal digunakan untuk memilih satu subset kecil yang bertindan dengan set yang terkandung $m\%$ sistem yang sebenar baik dengan kebarangkalian yang tinggi. Kemudian, prosedur peruntukan pengkomputeran bajet optimum dilaksanakan dalam peringkat kedua untuk peruntukan bajet pengkomputeran yang ada. Ini diikuti dengan prosedur pemilihan subset bagi mendapatkan subset lebih kecil yang mengandungi sistem terbaik daripada subset yang dipilih sebelumnya dengan kebarangkalian yang tinggi. Akhirnya, prosedur zon tiada-beza digunakan bagi memilih sistem terbaik daripada subset sebelumnya dengan kebarangkalian yang tinggi. Kecekapan pendekatan pemilihan yang dicadangkan disemak dari dua sudut berbeza. Pertama, berdasarkan perubahan sebahagian pembolehubah seperti saiz sampel awal, peningkatan sampel simulasi, jumlah bajet dan masa yang berlalu. Kedua, berdasarkan tiga set petua henti seperti berjujukan, jangkaan kos peluang dan kebarangkalian pemilihan aturan pem-

berhentian yang baik. Sebagai tambahan, perbandingan di antara pendekatan pemilihan yang dicadangkan dengan pendekatan pemilihan Tiga-Peringkat juga dipersembahkan. Akhirnya, salah satu masalah yang paling sukar dalam rekabentuk baris pemasangan yang dikenali sebagai masalah peruntukan penimbal dibentang sebagai aplikasi sebenar bagi pendekatan yang dicadangkan. Secara amnya keputusan menunjukkan pendekatan prosedur pemilihan yang dicadangkan dapat melakukan pemilihan betul dengan kebarangkalian yang tinggi.

A NEW SELECTION PROCEDURE FOR LARGE SCALE PROBLEMS

ABSTRACT

This thesis considers the problem of selecting the stochastic system that has the best (maximum or minimum) expected performance measure when the number of alternatives is finite but large. Ranking and selection procedures have been used successfully for solving problems with small number of alternatives. In order to reduce the computational problem, the idea of ordinal optimization is being used with the objective of finding a good enough system instead of looking for the best system. In this thesis, a new selection approach is proposed for selecting a good system when the number of alternatives is very large. This approach contains four stages; In the first stage, the ordinal optimization procedure is used for selecting a small subset that overlaps with the set that contains the actual best $m\%$ systems with high probability. Then, the optimal computing budget allocation procedure is applied in the second stage to allocate the available computing budget. This is followed with the subset selection procedure to get a smaller subset that contains the best system among the subset that was selected before with high probability. Finally, indifference-zone procedure is used to select the best system from the previous subset with high probability. The efficiency of the proposed selection approach is being examined from two different points. First, based on some parameters changing such as the initial sample size, increment in simulation samples, total budget, and the elapsed time. Secondly, based on three sets of the stopping rules such as sequential, expected opportunity cost and probability of good selection of the stopping rule. In addition, comparisons between the proposed selection approach and the Three-stage selection approach are also presented. Finally, one of the most

difficult problem in designing of production lines, which is known as buffer allocation problem is presented as a real application for the proposed approach. The implementations of our approach are presented with some numerical examples. The results show that in general, the proposed selection approach made the correct selection with high probability.

CHAPTER 1

INTRODUCTION

Selecting the best stochastic system of a large set of alternatives is one of the most important optimization problems. This research provides a combination between the ordinal optimization and cardinal optimization for selecting the best system in large scale problems.

This thesis introduces a new proposed selection approach for selecting a good system when the number of alternatives is very large. It is a combination of the cardinal optimization (ranking and selection procedures) and the ordinal optimization. The advantage of the new proposed selection approach is that, it can be used to select the best system from a very large number of alternatives, because it uses the ordinal optimization method in order to decrease the number of the competing alternatives, to be appropriate for the cardinal optimization methods.

This chapter discusses the background of the problem and the problem statement for the selection problems. The goals and constraints are stated and the research objectives are refined. The significance of the research and research methodology are also discussed in this chapter. Also, in this chapter a background for $M/M/1$ queuing system are discussed. Finally, organization of the thesis is presented.

1.1 Problem Background

There are many problems in a real world that cannot be solved until the discovery of calculus, which enable mathematical modeling for a large number of physical problems. However, since the optimization technique emerged in 1950's, there are many algorithms developed, which

were previously thought to be infeasible for numerical solution in difficult optimization problems, now become possible. However, optimization is an old idea, seeking improvement.

Computational load is one of the important problems in optimization that is still being trying to solve. Actually, this problem does not always necessarily arise due to the size's problem. Instead, the complexity of a problem can also caused infeasible computational load. Simulation models are used to solve many complex problems such as large scale electric power grids, air and land traffic control systems, the Internet and other communication networks. Such systems work with evolve in time via human-made rules of operation, and are difficult to describe by mathematical models such as differential equations for physical systems. These systems are known as discrete event dynamic systems (*DEDS*).

Simulation model is important for designing the problem specially when the real system does not exist or, when it is impossible to do experiment on the real system. Unfortunately, having a simulation model is not the end of the problem difficulty. Usually, it involves with time consuming model.

In this thesis, we consider techniques that are designed to solved the problem of selecting the system that has the best performance measure. This system is known as the best system, where the best system is defined as the one that has the (maximum or minimum) performance measure (mean). Moreover, an experimenter might wish to select several systems from a set of best systems. Selection of the systems according to their ordered means is the problem in ranking the systems.

In fact, when the number of alternative systems is small, there are many techniques can be used to achieve this goal, and these techniques are known as the ranking and selection procedures. Some of these techniques are designed to select a single best system as in indifference-

zone procedures, and some are designed to screen out a set of systems before choosing a random size subset of systems containing the best one as in subset selection procedures. The problem arises for a large scale problem because it needs a huge computational time. For each sample of the performance value requires one simulation run. Therefore, for a large scale problem will need a large number of samples which is very time consuming and may be impossible, especially when we are dealing with a huge number of alternatives in the solution set, as in the engineering applications. In this situation, we would change our objective to finding good systems rather than estimating accurately the performance value for these systems, which is the idea of ordinal optimization.

To selecting the best alternative systems, the experimenter needs two things. First, a statistical procedure that will tell the experimenter which system to select. Second, an operating characteristic function for that procedure based on the probability of making a correct selection. On the other hand, in any experiment there are several different types of decisions that are required to investigate a research hypothesis. These types are characterized (Bechhofer et al., 1995) as follows:-

- The experimenter must determine what characteristics are to be measured and then what is the system design to be used. The system design should include the specification of system combinations to be used.
- The experimenter must determine the number of times each system is to be observed. That is, how many replications need to be performed. Adequate replication will ensure the experimenter be able to achieve the desired design requirement for whatever goal and statistical procedure is to be employed.
- "Blocking" is the third aspect of the experimental design; it means that, system comparisons must be made across homogeneous experimental units.

- The final aspect of the experimental design is the determination of the manner in which the system combinations are randomized to the experimental units. Randomization prevents unrecognized factors from systematically confounding the results.

1.2 Problem Statement

We consider the following optimization problem,

$$\min_{\theta \in \Theta} f(\theta) \tag{1.1}$$

where Θ is the feasible solution set and it is an arbitrary, has no structure, finite but a huge set. Let f be the expected performance measure of some complex stochastic system,

$$f(\theta) = E[L(\theta, Y)]$$

where θ is a vector represents the system design parameters, Y represents all the random effect of the system and L is a deterministic function that depends on θ and Y. Unfortunately, it is not easy to solve this problem because the numerical estimation of the expected performance measure is only available for a limited class of performance measures, and to estimate this value we use simulation which is costly and consume time especially when dealing with a huge number of alternatives.

The goal of the selection procedure is to identify the best n simulated systems. Assume the best system is defined as the system with the (largest or smallest) mean, which is unknown and to be inferred from simulation. Suppose that there are n systems, and let Y_{ij} (observation) represent the j^{th} output from the system i , and let $\mathbf{Y}_i = \{Y_{ij}, j = 1, 2, \ldots\}$ denote the output sequence from the system i. Also assume that Y_{ij} are independent and identically distributed

$(i.i.d.)$ normal with unknown means $\mu_i = E(Y_{ij})$ and variances $\sigma_i^2 = Var(Y_{ij})$. In practice the σ_i^2 are unknown, so we estimate it using the sample variances s_i^2 for Y_{ij}. Also, assume that the $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ are mutually independent.

1.3 The Normality and Independent Assumptions

The raw data in the real world problems are rarely normally distributed, and are not independent. For example, waiting times for customers in a queuing system are usually dependent because a long delay for one customer tends to increase the delays to the following customers. To solve these problems we hope to approximate these raw data to be normally distributed and independent.

To achieve the normality assumption we need to transform the raw data into batch means, which are the averages of a large number of raw data, (Law and Kelton, 2000; Kim and Nelson, 2006a,b). However, the batch means are far less dependent and non normal than the raw data if the batch size is large enough. Another solution for the normality problem is to make multiple replications of each alternative. Then using within-replication averages as the basic observations, and prolong the replications will make the within-replication averages are approximately normal distribution. In both solutions the normality assumption is hold by using the Central Limit Theorem, that suggest the replication averages will be normally distributed approximately. On the other hand, the independent assumption is valid in the simulation experiments when using a different sequence of random numbers to derive the simulation of each system.

1.4 Research Objectives

Selecting the best stochastic system of a huge set of alternatives is one from the most important optimization problems. This thesis provides an useful combination between the ordinal optimization and cardinal optimization for selecting the best system in large scale problems. The objectives of this thesis are as follows:

- To propose a new selection approach that can be used to find the best stochastic system from a large number of alternative set. This new approach is based on the idea of using the ordinal optimization with optimal computing budget allocation procedures to reduce the number of systems in the search space, in order to be suitable for the ranking and selection procedures.
- To study the efficiency of the new proposed selection approach from two points; some changes in the parameter settings and when applying three different stopping rules.
- To compare between the new proposed selection approach with the Three-stage selection approach that have been proposed by Alrefaei and Almomani (2004), to select the best system for large scale problems.
- To solve one of the real life problems, which is known as the buffer allocation problem, by using the new proposed selection approach.

In order to achieve these objectives, the research framework of this study is designed as summarized in Figure 1.1. In this figure, $P(CS)$ represents the probability of correct selection, $E(OC)$ is the expected opportunity cost of a potentially incorrect selection, t_0 is the initial sample size, Δ is the increment in simulation samples, B is the total simulation budget, T is the elapsed (execution) time, S is the sequential stopping rule, $E(OC)$ is the expected opportunity cost stopping rule and $P(GS)_{\delta^*}$ is the probability of good selection stopping rule.

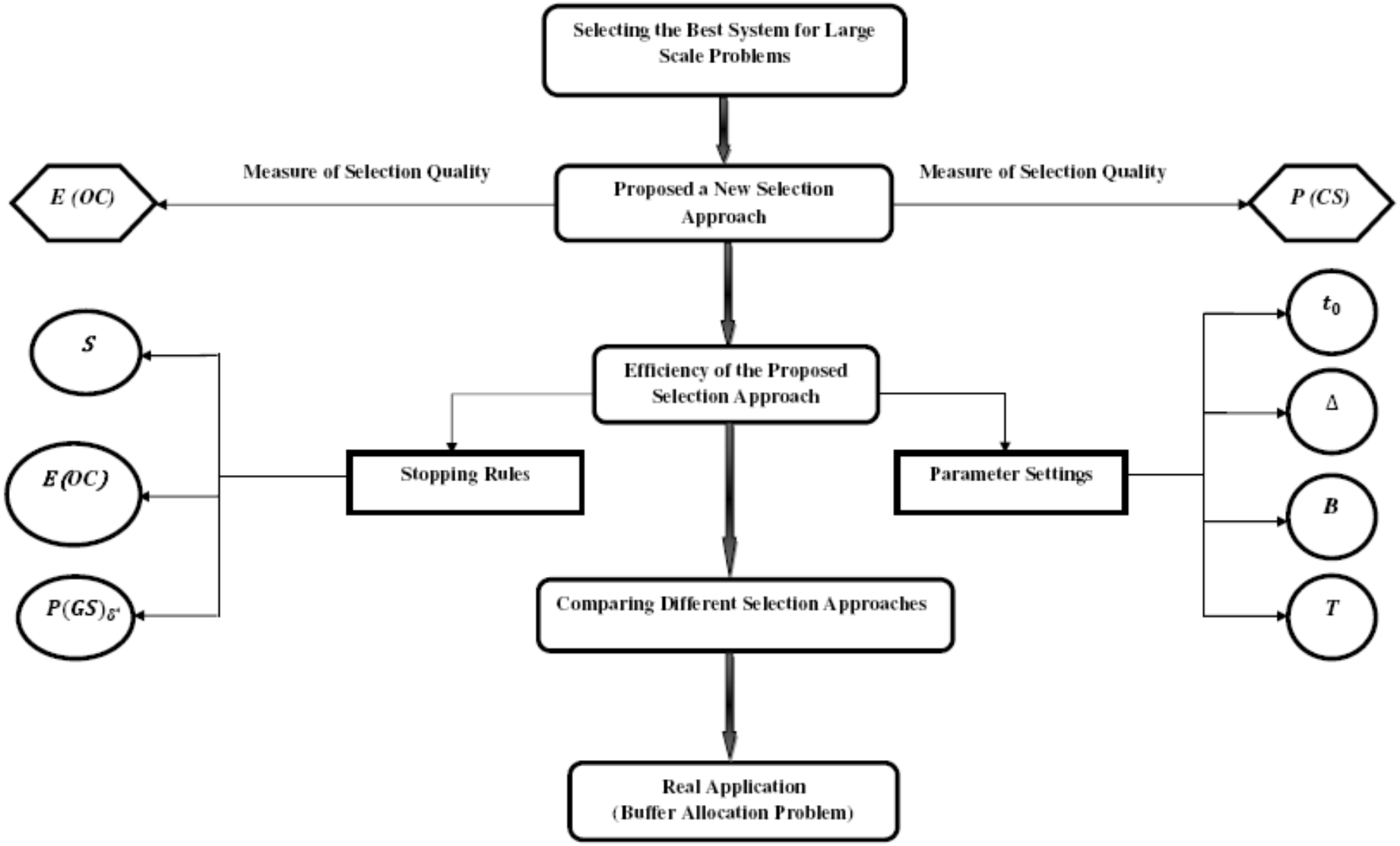

Figure 1.1: A diagrammatical summary of the problem

1.5 Significance of the Research

Selection problems affect our life everyday. The statistician is called upon to help in providing a rational procedure for selecting the best from several alternatives. Thus, an agronomist studying a number of varieties of wheat may seek to select the variety that will produce the largest number of bushels per acre. A clinician may be interested in which type of drug is most effective in treating a certain disease. A manufacturing engineer may wish to find a number of possible assembly line configurations with minimum cost. A polling agency may wish to determine which candidate is most preferred by electorate.

Many of selection problems are dealing with a large number of alternatives, like engineering application. Therefore, existence of a selection approach that can be used to solve these problems is very important. However, selecting the best system from a feasible solution set is not enough here, although it is the main target. On the other hand we want to achieve the target with a high probability of correct selection, small expected opportunity cost of a potentially incorrect selection, minimum simulation samples and minimum elapsed time. This thesis will achieve the target to select the best system under these constraints, by proposing a new selection approach. The proposed selection approach will select the best system from a large feasible solution set with high probability of correct selection and relatively small in expected opportunity cost of a potentially incorrect selection, simulation samples and elapsed time.

Finally, to solve a buffer allocation problem using the proposed selection approach as an important contribution in this thesis since the buffer allocation problem in a finite production line is one of the most difficult problems in modeling performance and in designing a production line. Of course this study can be further extended to solve another problems, such as an inventory system, a different type of queuing systems, and other management problems.

1.6 Methodology

The existing methods for solving the selection problems are used in developing and constructing a new selection approach. It involves with two kinds of optimization; the ordinal and cardinal. It is a combination of ordinal optimization, optimal computing budget allocation, subset selection and indifference-zone procedures. Initially, using ordinal optimization procedure, a small subset is randomly selected from a feasible solution set that overlaps with the set that contains the actual best $m\%$ systems with high probability. Then optimal computing budget allocation procedure is used to allocate the available computing budget. This is followed with subset selection procedure to get a smaller subset with high probability, that contains the best system among the previous selected subset. Finally, indifference-zone procedure is applied to select the best system from previous set.

Every procedure that involved here has a specific use with a clear goal. Note that, the proposed algorithm is appropriate to select the best stochastic system for large scale problems. This is because the ordinal optimization with the optimal computing budget allocation procedures is used to reduce the number of systems in the feasible solution set. Therefore, it will be suitable for applying the subset selection and indifference-zone procedures. Furthermore, using the idea of optimal computing budget allocation in the proposed algorithm will improve the performance of the ordinal optimization which will lead to the improvement in the whole performance of the proposed selection approach.

The efficiency of the proposed selection approach under different parameter settings and different stopping rules are then being studied by applying it on a $M/M/1$ queuing system. This is follows with a study on efficiency of the proposed selection approach in solving a real problem, specifically, to solve the buffer allocation problem with a specific setting.

Statistical selection approaches is used to select the best system from a finite set of alternatives, when the stochastic simulation is used to indicate the performance measure for each alternative system. Since, the used of the simulation methods have a potential for incorrect selection, so we need measures to determine the quality of selection. There are two measures of selection quality; first, the probability of correct selection, where the correct selection means that the selected system is the system that belong to the actual $m\%$ best subset. Secondly, is the expected opportunity cost of a potentially incorrect selection, where the opportunity cost is defined as the difference between the unknown means of the selected system and the actual best system. To estimate the probability of correct selection, we need to count the number of times we successfully find the best systems that belong to the actual $m\%$ best subset out of 100 independent replications.

In this thesis, we use "java" as the programming language in the numerical examples and was run using a computer model of $Optiplex380$, manufactured by *Dell* with installed memory (RAM) 2.00 GB and the Processor $Intel(R)\ Core(TM)2\ Duo\ CPU\ E7500$ @ $2.93GH_z$ $2.93GH_z$.

1.7 *M/M/1* Queuing System

The analysis of queues (waiting lines) is called queuing theory. It is applied to any situation in which customers arrived at a system, waited, and received a service. The queuing theory is used to solve the queuing problem where the problem is about a balance between average waiting time for one customer and an idle time for a server. The main objectives of queuing theory are to improve customer service and to reduce operating costs. Applications of queuing theory are found in various fields, such as in time-shared computer system design, traffic control and hospital management. In a business world, more customers mean more business transactions. Out of the many ways to attract customers, an efficient queuing system plays an important role as it reduces the customers waiting time. Of course, this will make customers happy, and as a

result the customer will come back for doing business again.

In particular, all queuing systems have three elements:

1. Customers that waiting for a service. They can be people, machines waiting to get repaired, a telephone calls and trucks waiting for loading.

2. Servers that provide service. They can be people, telephone center, washing bays, ATM's and computers.

3. A waiting line or queue. The queue is the set of customers waiting for getting the service. It may be a physical line, or an invisible one such as information to be processed in a computer.

As example, in the mini-market, there is a cashier counter, and customers reached the counter in random. The customer will pay off immediately and leave the line, provided that the customer reaches the cashier while it is free from other customer. If the cashier is busy as the customer reaches the counter, the customer will have to wait in the line. Once another customer enters the queue, the customer will receive a service according to the queuing rule, then departs after receiving service. A simple queuing system is shown in Figure 1.2.

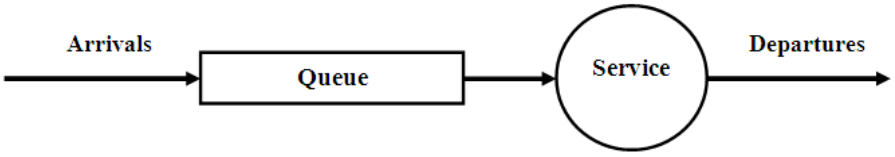

Figure 1.2: A simple queuing system

A queuing system provides measures for a system performance. The first measure is, the quality of service provided to the customers, which is can be measured by waiting time in the queue, time in the system (waiting time plus service time), and completion by a deadline. The

second measure is, the efficiency of the service operation and the cost in providing the service, in which can be measured by an average queue length, average number of customers in the system (queue plus customers in service), throughput (the rate at which customers are served), server utilization (percentage of time servers are busy), and percentage of customers who balk.

In simulation of the queuing system, there are three perceptions usually being used:

1. The arrival process.
2. The service mechanism, such as the number of servers and the service-time distribution.
3. The queuing rule and the criteria of the queuing system such as, first come first served ($FCFS$), last come first served ($LCFS$), random served and priority served.

The notation $A/S/N/C$ is used to classify a queuing system, where A specifies the type of arrival process, S denotes the service-time distribution, N specifies the number of servers, and C denotes the capacity of the system, that is, the maximum number of customers that can be accommodated (including number in queue and number being served). If C is not specified, it is assumed that the capacity of the system is unlimited. For example, an $M/M/2$ queuing system, M represented the memoryless, exponential interarrival times, and a Poisson arrival process, with 2 servers. An $M/G/1$ queuing system has Poisson arrivals, general service-time distribution, and a single server. An $M/D/1$ queuing system, D stands for constant (deterministic) service time. However, there are many examples of queuing systems with limited capacity among other are hospital emergency rooms with limited beds and telephone systems with limited trunks. Our concern in this thesis is $M/M/1$ queuing system, where the two M's refer to the fact that both the interarrival and the service distributions are exponential, and there is one server.

In the $M/M/1$ queuing system, the customers arrive at a single-server service station follows the Poisson probability distribution with λ is the mean arrival rate. Note that, if the number of arrivals follow a Poisson process with λ mean arrival rate, then the time between arrivals has an exponential distribution with mean interarrival time is $1/\lambda$. It is clear that, each customer upon arrival goes directly into service if the server is ideal and if not, the customer will have to join the queue. When the server finishes serving a customer, the customer leaves the system, and the next customer in line, if there is any, enters service. The successive service times are assumed to follow an exponential distribution with mean service rate, μ. Then, the average service time will be $1/\mu$. The performance measures of the system are the customers waiting time, the length of queue, and the idle time of the server. In this thesis our goal is to minimize the average waiting time that a customer spends in the system. So, if we assume W represents the average amount of time that a customer spends in the system, then it can be shown that $W = \frac{1}{\mu-\lambda}$. More details on the $M/M/1$ queuing system can be found in (Ross, 2007; Hsu, 1997; Lian and Wan, 2007).

1.8 Organization of the Thesis

This thesis is organized in the following manner. The introduction and some background to the selection problems was explained in the beginning of Chapter 1, followed by the problem statement, the objectives, the significance of research and the methodology of the study. It also gives some discussions on the normality and independent assumptions with a brief background for the $M/M/1$ queuing system. In Chapter 2, a background of statistical selection procedures are presented together with the ordinal optimization procedure. This chapter is divided into two parts. The first part deals with the statistical selection procedures which include, indifference-zone, subset selection and multiple comparisons procedures. While the second part explains the idea of ordinal optimization and optimal computing budget allocation. Also in this chapter

the measures of selection quality will be discussed.

The new proposed selection approach is presented in Chapter 3 together with a performance evaluation and an examples to illustrate how the approach works. The efficiency of the proposed selection approach under different parameter settings and different stopping rules are presented in Chapter 4 with a series of numerical illustrations. In this chapter, there are four parameter settings are being discussed; they are the initial sample size, increment in simulation samples, total budget and the elapsed time. We also study in case of three stopping rules; the sequential, the expected opportunity cost and the probability of good selection.

Chapter 5 presents a comparison between different selection approaches that have been used to solve the large scale problems. For the sake of the comparisons, in this chapter we argue the Three-Stage selection approach in context of the expected opportunity cost as a measure of selection quality and also we discuss the effect of the initial sample size on the performance of the Three-Stage selection approach. Chapter 6 presents how the new proposed method works in buffer allocation problem. Finally, conclusions, contribution of the thesis and suggestions for further research are presented in Chapter 7.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

We consider the problem of selecting the system that has the best performance measure with a pre-specified high probability. This system is defined as the best system, where the best system might be the one that has maximum or minimum performance measure (mean). We assume that all systems in the search space are normally distributed with unknown means and unknown variances.

Some of the statistical selection procedures that will be discussed in this chapter, are designed to select a single best system. Whereas, other procedures are designed to screen a set of systems by choosing a (random size) subset of the systems containing the best one. In the other hand, some procedures are used to construct simultaneous confidence intervals for specific set of mean system. However, all these methods are effective when the number of alternatives is relatively small.

In the real world problems, the number of alternatives is very huge, so we cannot use statistical selection procedures to select the best systems. To solve this problem, usually we change our concern from finding the best system to finding the good enough system with high probability. In this situation, the ordinal optimization becomes more important from cardinal optimization.

Statistical selection procedures are used to select the best system from a finite set of alternatives, when the stochastic simulation is used to indicate the performance measure for each alternative in the simulation. Since the used of the simulation methods have a potential for incorrect selection, so we need a measure to determine the quality of selection. In fact, there are two measures of selection quality. The first measure is the probability of correct selection, and the second one is the expected opportunity cost of a potential incorrect selection.

In this chapter, we argue the statistical selection procedures (the indifference-zone, the subset selection and multiple comparisons) and the ordinal optimization procedure with the budget allocation problem. Then, we present several combined procedures that are used to select the best system, when the number of alternatives is large. Finally, we discuss two measures of selection quality; the probability of correct selection and the expected opportunity cost of a potential incorrect selection.

2.2 Statistical Selection Procedures

The problem of selection of some or all the systems according to their ordered means, can be solved by ranking and selection procedures, which involved two different procedures; the indifference-zone and the subset selection. A natural strategy of ranking and selection procedures for dealing with this problem is to make comparisons between different systems whose response is normally distributed when the number of systems in the search space is small. These procedures usually are used to select the best system or a subset that contain the best systems when the number of systems is small, with a pre-specified significance level.

In this section, we consider three basic problem formations for selecting, screening and multiple comparison problems. We discuss three different statistical selection procedures that work in case of a single normal random effect of the system with n systems. These procedures

are, the indifference-zone, the subset selection and multiple comparisons procedures.

2.2.1 The Indifference-Zone Procedure

Consider the problem of selecting the best system among n systems when n small ($n \geq 2$), by using the indifference-zone selection procedure. Suppose that we are dealing with systems that are normally distributed with unknown means $\mu_1, \mu_2, \ldots, \mu_n$ and unknown variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$. We assume that a largest mean is better, therefore if the ordered μ_i-values are denoted by $\mu_{[1]} \leq \mu_{[2]} \leq \ldots \leq \mu_{[n]}$, then the system having mean $\mu_{[n]}$ is refer to as the best system. In the indifference-zone procedure, the indifference zone is the interval $[\mu_{[n]} - \delta^*, \mu_{[n]}]$, where δ^* is a small positive real number predetermined by the examiner. The objective is to selecting the system i^* such that $\mu_{i^*} \in [\mu_{[n]} - \delta^*, \mu_{[n]}]$. Let the correct selection (CS) here is defined as, selecting the system whose mean belongs to the indifference zone. We would like the correct selection to take place with high probability, say with a probability no smaller than P^*, where P^* is predetermined by the examiner. In mathematical notation, we seeks to achieve $P(CS) \geq P^*$ whenever $\mu_{[n]} - \mu_{[n-1]} \geq \delta^*$, where $0 < \delta^* < \infty$ and $1/n < P^* < 1$, i.e. $P($ select $[n] \mid \mu_{[n]} - \mu_{[n-1]} \geq \delta^*) \geq P^*$.

The δ^* is called the indifference zone, which is the difference between the best system and the favorable system. It represents the smallest difference that we want to achieve. Therefore, we must be very careful in determining the value of δ^*, because the system requires to implicitly specifying the common single stage sample size that required for the n competing systems. If δ^* is too small, then the required number of simulation samples or replications to guarantee the probability requirement is expected to be large. Therefore, δ^* can be choose as the smallest difference of $\mu_{[n]} - \mu_{[n-1]}$ and it is worth to detect. In the other hand, large P^* may require a larger number of simulation samples or replications to achieve the experimenter request.

In the indifference-zone procedures, we must use at least two-stage procedure when the common variances are unknown to select the best system. Indifference-zone procedures are differ in both, the standard used for selecting the best system, and the choice of the sample size in the second stage. The first stage involves obtaining a fixed number of simulation samples t_0, to get the first stage sample means and sample variances for each system in the feasible solution set Θ. Then using the sample mean and variance from the first stage, the additional simulation samples needed from each system in Θ is determined in order to get the sample means of the second stage. Then a weighted average of the first and second stage sample mean is computed and the system with the best weighted average estimated mean will be selected as the best system.

The first work that presented indifference-zone formulation was proposed by Bechhofer (1954). It involved with a single-stage procedure for ranking means of normal systems with known variances, and suggested the probability requirement for the indifference-zone procedure. Hayter (1989) has proposed a single-stage procedure when the variances are unknown and bounded. In this procedure, the variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ are unknown but $\max\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2\} \leq \sigma_v^2$, where σ_v^2 is known. The Hayter (1989) procedure is as follows, (Bechhofer et al., 1995):

1. For the given n and specified δ^* and P^*, choose $Z_{n-1,1/2}^{(1-P^*)}$ from Table A.1 in Appendix A, corresponding to the n and P^* of interest.

2. Calculate $t = \lceil 2(\sigma_v Z_{n-1,1/2}^{(1-P^*)}/\delta^*)^2 \rceil$, where $\lceil x \rceil$ denotes the smallest integer greater than or equals to x.

3. Take random samples of t observations y_{ij} $(j = 1, \ldots, t)$ for each system i, and calculate the n sample means $\bar{y}_i = \frac{\sum_{j=1}^{t} y_{ij}}{t}$, where $i = 1, \ldots, n$.

4. Select the system i^* that satisfies $\bar{y}_{i^*} = \max\{\bar{y_1}, \bar{y_2}, \ldots, \bar{y_n}\}$ as an estimate of $\mu_{[n]}$.

Remark:- The value of the constant $Z_{n-1,1/2}^{(1-P^*)}$ that have been used to implement the procedure of Hayter (1989), is a special case of the upper equicoordinate point of the multivariate normal distribution. If vector $(W_1,\ldots,W_{n-1})$ has the $(n-1)$-variate multivariate normal distribution with mean vector zero, a unit variances, with a common correlation 1/2, then

$$P(W_1 \leq c, W_2 \leq c, \ldots, W_{n-1} \leq c) = \int_{-\infty}^{\infty} \Phi^{n-1}\left(z + c\sqrt{2}\right) d\Phi(z)$$

where $c = Z_{n-1,1/2}^{(1-P^*)}$, and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, (Bechhofer et al., 1995).

If no information about the variances is available, then the two-stage procedure of Rinott (1978) can be used. This procedure is applicable when the data are normally distributed and all systems are simulated independently of each others. However, Rinott (1978) has proposed a sequential procedure, that consists a two-stage procedure for the case when the variances are completely unknown. In fact, Rinott (1978) procedure is one of the simplest and the most well known ranking and selection procedures. In the other hand, most other ranking and selection procedures are directly or indirectly based on this procedure. The Rinott (1978) procedure is as follows:

For the given n and a specified δ^* and P^*, fix a number of simulation samples $t_0 \geq 2$ to be taken in Stage 1. Choose the constant $g = g(n, P^*, v)$ from Table A.2 in Appendix A, where $v = t_0 - 1$.

Stage 1:

1. Take random samples of $t_0 \geq 2$ observations y_{ij} $(j = 1,\ldots,t_0)$ for each system i, where $i = 1,\ldots,n$.

2. Calculate the first sample mean $\bar{y}_i^{(1)} = \frac{\sum_{j=1}^{t_0} y_{ij}}{t_0}$, and sample variance $s_i^2 = \frac{\sum_{j=1}^{t_0} (y_{ij} - \bar{y}_i^{(1)})^2}{v}$ as an unbiased estimator of σ_i^2 based on $v = t_0 - 1$ degrees of freedom.

   Stage 2:

3. Compute $T_i = \lceil (g s_i / \delta^*)^2 \rceil$ for all $i = 1, \ldots, n$. Take random samples of $\max\{0, T_i - t_0\}$ additional observations from each system i, where $i = 1, \ldots, n$.

4. Calculate the overall sample means $\bar{y}_i = \frac{\sum_{j=1}^{T_i} y_{ij}}{T_i}$ for $i = 1, \ldots, n$.

5. Select the system i^* that satisfies $\bar{y}_{i^*} = \max\{\bar{y_1}, \bar{y_2}, \ldots, \bar{y_n}\}$ as an estimate of $\mu_{[n]}$.

Remark:- In the procedure of Rinott (1978), constant $g = g(n, P^*, v)$ is the solution of

$$\int_0^\infty \int_0^\infty \left[\Phi\left(\frac{g}{\sqrt{v(1/x + 1/y)}} \right) f_v(x) \right]^{n-1} f_v(y)\, dy\, dx = P^*$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $f_v(\cdot)$ is the probability density function of the Chi-squared distribution with v degrees of freedom, (Bechhofer et al., 1995).

Rinott (1978) procedure has two properties. First, it does not eliminate any system at the end of stage one, but instead it uses the data of each system in stage one to estimate the variance of that system. Second, the total number of simulation samples or replications from each system is a random variable, so it is possibly different for each system.

Kim and Nelson (2007) present a brief proof sketch of the statistical validity of Rinott (1978) procedure. They assume that the variances are known and equal across all systems. Therefore, Rinott (1978) procedure does not need two stages and it becomes a one-stage procedure. In addition, Kim and Nelson (2007) give an overview on recent advances made in indifferent-zone ranking and selection procedures. Kim and Nelson (2006a) described the ba-

sic rule for the ranking and selection, and discussed many other types of ranking and selection problems. Furthermore, they gave useful theorems to handle the problems with the unknown and unequal variances.

Recently, Kim and Nelson (2001) proposed a fully sequential procedure, denoted by (*KN*) procedure. The goal of this sequential procedure is to eliminate, at an early stage of simulation process, those simulated systems that are apparently inferior, and thus reduce the overall computational effort. The *KN* procedure is fully sequential since it takes only a single basic output from each system that is still in competition at each stage. However, if there exists evidence that a system is inferior, then it will be eliminated immediately from consideration. This procedure requires independent and identically normal distribution data, to find the best system. The *KN* procedure is as follows:

1. Select confidence level $1-\alpha$, indifference zone δ^*, and first stage sample size $t_0 \geq 2$. Calculate φ and c as described below.

2. Let $I = \{1,2,\ldots,n\}$ be the set of systems still in competition, and let $h^2 = 2c\varphi(t_0-1)$.

3. Obtain t_0 of observations y_{ij}, $(j = 1,\ldots,t_0)$ from each system i, where $i = 1,\ldots,n$. For all $i \neq l$, compute $s_{il}^2 = \frac{1}{t_0-1}\sum_{j=1}^{t_0}(y_{ij}-y_{lj}-[\bar{y}_i(t_0)-\bar{y}_l(t_0)])^2$ as the sample variance of the difference between systems i and l.

4. Let $T_{il} = \left\lfloor \left(\frac{hs_{il}}{\delta^*}\right)^2 \right\rfloor$, where $\lfloor x \rfloor$ refereing to the greatest integer less than or equal to the real number x, and let $T_i = \max_{i\neq l} T_{il}$. Here T_i+1 is the maximum number of observations that can be taken from system i.

5. If $t_0 > \max_i T_i$, then stop and select the system with the largest $\bar{y}_i(t_0)$ as the best system. Otherwise, set the observation counter $d = t_0$ and go to the next step.

6. Set $I^{old} = I$. Let $I = \{i : i \in I^{old} \text{ and } \bar{y}_i(d) \geq \bar{y}_l(d) - B_{il}(d),\ \forall l \in I^{old}, l \neq i\}$, where $B_{il}(d) = \max\left\{0, \frac{\delta^*}{2cd}\left(\left(\frac{hs_{il}}{\delta^*}\right)^2 - d\right)\right\}$.

7. If $|I| = 1$, then stop and select the system whose index is in I as the best system. Here $|I|$ is the number of elements in the set I. Otherwise, take one additional observation $y_{i,d+1}$ from each system $i \in I$, and set $d = d + 1$.

8. If $d = \max_i T_i + 1$, then stop and select the system whose index is in I and has the largest $\bar{y}_i(d)$ as the best system. Otherwise, go to step 6.

Remark:-

- $\bar{y}_i(d) = \frac{1}{d}\sum_{j=1}^{d} y_{ij}$ denote the sample mean of the first d observations from system i.

- Constant c is a nonnegative integer, with standard choices being $c = 1, 2$. It is easy to compute φ when $c = 1, 2$.

- Constant φ is the solution to the equation

$$f(\varphi) \equiv \sum_{k=1}^{c} (-1)^{k+1}\left(1 - \frac{1}{2}\zeta(k = c)\right)\left(1 + \frac{2\varphi(2c-k)k}{c}\right)^{-(t_0-1)/2} = \frac{\alpha}{n-1}$$

  where ζ is the indicator function. In the special case that $c = 1$, we have the closed form solution $\varphi = \frac{1}{2}\left[\left(\frac{2\alpha}{n-1}\right)^{-2/(t_0-1)} - 1\right]$.

To prove the validity of the *KN* procedure, Kim and Nelson (2001) purposed the following theorem:

Theorem 2.2.1 *Suppose that $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ are independent and identically distributed multivariate normal with unknown mean vector μ, that is arbitrary except for the condition that $\mu_n \geq$*

$\mu_{n-1} + \delta^*$, and unknown and arbitrary positive definite covariance matrix Σ. Then with probability $\geq 1 - \alpha$ the fully sequential indifference-zone procedure selects system n.

proof:- See (Kim and Nelson, 2001).

Kim and Nelson (2006b) present two fully sequential procedures for steady-state simulation that are designed to select the best system of a finite number of simulated systems. These two procedures are extensions of *KN* to steady-state simulation. The first procedure is called (*KN*+) procedure in which the variance estimators in this procedure are calculated once from the first stage sample, and the second procedure is called (*KN*++) procedure with the variance estimators in this procedure are allows to updated.

The *KN*+ procedure does not update the variance estimators after the first stage of sampling. The goal of this procedure is finding the system with the largest steady-state mean. However, the goal can be modified to select the system with the smallest mean. Kim and Nelson (2006b) proved the validity of this procedure. The *KN*+ procedure is as follows:

1. Select confidence level $1/n < 1 - \alpha < 1$, indifference zone δ^*, first stage sample size $t_0 \geq 2$, and batch size $m_0 < t_0$. Calculate φ and c as described below.

2. Let $I = \{1, 2, \ldots, n\}$ be the set of systems still in competition, and let $h^2 = 2c\varphi z$, where the degrees of freedom z is determined by which variance estimator is used.

3. Obtain t_0 observations y_{ij} $(j = 1, \ldots, t_0)$ from each system $i = 1, \ldots, n$. For all $i \neq l$, compute the estimator $m_0 V_{il}^2$ the sample asymptotic variance of the difference between systems i and l.

4. Let $T_{il} = \left\lfloor \frac{h^2 m_0 V_{il}^2}{(\delta^*)^2} \right\rfloor$, where $\lfloor x \rfloor$ is the greatest integer less than or equal to the real number x, and let $T_i = \max_{i \neq l} T_{il}$. Here $T_i + 1$ is the maximum number of observations that can

be taken from system i.

5. If $t_0 \geq \max_i T_i + 1$, then stop and select the system with the largest $\bar{y}_i(t_0)$ as the best system. Otherwise, set the observation counter $d = t_0$ and go to the next step.

6. Set $I^{old} = I$. Let $I = \{i : i \in I^{old} \text{ and } \bar{y}_i(d) > \bar{y}_l(d) - B_{il}(d),\ \forall l \in I^{old}, l \neq i\}$, where $B_{il}(d) = \max\left\{0, \frac{\delta^*}{2cd}\left(\frac{h^2 m_0 V_{il}^2}{(\delta^*)^2} - d\right)\right\}$.

7. If $|I| = 1$, then stop and select the system whose index is in I as the best system. Otherwise, take one additional observation $y_{i,d+1}$ from each system $i \in I$ and set $d = d + 1$, and go to step 6.

Remark:-

- Constant c may be any nonnegative integer.

- Constant φ is the solution to the equation

$$f(\varphi) = \sum_{k=1}^{c} (-1)^{k+1} \left(1 - \frac{1}{2}\zeta(k = c)\right) \left(1 + \frac{2\varphi(2c-k)k}{c}\right)^{-z/2} = 1 - (1-\alpha)^{1/(n-1)}$$

where ζ is the indicator function.

The variance estimators used in $KN+$ depend only on the first stage sample. Nevertheless, $KN++$ procedure has improved the $KN+$ by updating the variance estimators as getting more samples. Kim and Nelson (2006b) prove the validity of this procedure. The $KN++$ procedure is as follows:

1. Select confidence level $1/n < 1 - \alpha < 1$, indifference zone δ^*, first stage sample size $t_0 \geq 2$, and batch size $m_0 < t_0$. Calculate φ and c as described below.

2. Let $I = \{1, 2, \ldots, n\}$ be the set of systems still in contention, and let $h^2 = 2c\varphi$.

3. Obtain t_0 observations y_{ij}, $j = 1, \ldots, t_0$ from each system $i = 1, \ldots, n$.

4. Set the observation counter $d = t_0$ and $m_d = m_0$.

5. If m_d has changed since the last update, then for all $i \neq l$, compute estimator $m_d V_{il}^2(d)$ the sample asymptotic variance of the difference between systems i and l based on b_d batches of size m_d.

6. Let $T_{il}(d) = \left\lfloor \frac{h^2 m_d V_{il}^2(d)}{(\delta^*)^2} \right\rfloor$, where $\lfloor x \rfloor$ is the greatest integer less than or equal to the real number x, and let $T_i(d) = \max_{i \neq l} T_{il}(d)$.

7. If $d \geq \max_i T_i(d) + 1$, then stop and select the system with the largest $\bar{y}_i(d)$ as the best system. Otherwise, go to the next step.

8. Set $I^{old} = I$. Let $I = \{i : i \in I^{old} \text{ and } \bar{y}_i(d) > \bar{y}_l(d) - B_{il}(d)\ \forall l \in I^{old}, l \neq i\}$, where $B_{il}(d) = \max\left\{0, \frac{\delta^*}{2cd}\left(\frac{h^2 m_d V_{il}^2(d)}{(\delta^*)^2} - d\right)\right\}$.

9. If $|I| = 1$, then stop and select the system whose index is in I as the best system. Otherwise, take one additional observation $y_{i,d+1}$ from each system $i \in I$ and set $d = d + 1$, and go to step 5.

Remark:-

- Constant c may be any nonnegative integer.

- Constant φ is the solution to the equation

$$f(\varphi) = \sum_{k=1}^{c} (-1)^{k+1} \left(1 - \frac{1}{2}\zeta(k = c)\right) exp\left(-\frac{\varphi}{c}(2c - k)k\right) = 1 - (1 - \alpha)^{1/(n-1)}$$

where ζ is the indicator function.