ADVANCES IN SIMULATION: VALIDITY AND EFFICIENCY

A Thesis
Presented to
The Academic Faculty

by

Judy S. Lee

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

Georgia Institute of Technology
May 2014

ADVANCES IN SIMULATION: VALIDITY AND EFFICIENCY

Approved by:

Professor Sigrún Andradóttir, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Hayriye Ayhan
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Richard Fujimoto
College of Computing
Georgia Institute of Technology

Professor David Goldsman
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Seong-Hee Kim
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: April 4 2014

To my family

ACKNOWLEDGEMENTS

Even after the strenuous days of writing, I am actually happy to take an opportunity to express my gratitude to those, who have influenced, inspired, taught, and supported my work.

First of all, I wish to thank my advisor, Dr. Sigrún Andradóttir for her insight, support, time and patience. This thesis would not have been existent without her guidance, from the beginning to the end. I would also like to thank the members of my committee, Dr. Hayriye Ayhan, Dr. Richard Fujomoto, Dr. Dave Goldsman, and Dr. Seong-hee Kim, for their valuable and insightful comments and encouragements.

I have been tremendously fortunate to meet all those people to whom I own my time in Georgia Tech, that I feel that I am bound to unintentionally omit someone's name however hard I try, but my thanks go to: Hinkyeol, Jooyoung, Seonghye, Haewon, Milim, Chuljin, Donggu and Mijung. I would also like to acknowledge tennis, that has served to be the only way to move me other than breathing and typing, and those whom I compete with, or against.

Finally and foremost, I cannot thank enough to my parents, sister, and grandparents. Thank you so much for your love and patience that supported this work, and for impatience that finally stimulated to bring this to an end. Hopefully, both my grandfathers, can rejoice the news and my acknowledgment, from wherever their peaceful souls are.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In this thesis, we present and analyze three algorithms that are designed to make computer simulation more efficient, valid, and/or applicable.

The first algorithm uses simulation cloning to enhance efficiency in transient simulation. Traditional simulation cloning is a technique that shares some parts of the simulation results when simulating different scenarios. We apply this idea to transient simulation, where multiple replications are required to achieve statistical validity. Computational savings are achieved by sharing some parts of the simulation results among several replications. We improve the algorithm by inducing negative correlation to compensate for the (undesirable) positive correlation introduced by sharing some parts of the simulation. Then we identify how many replications should share the same data, and provide numerical results to analyze the performance of our approach.

The second algorithm chooses a set of best systems when there are multiple candidate systems and multiple objectives. We provide three different formulations of correct selection of the Pareto optimal set, where a system is Pareto optimal if it is not inferior in all objectives compared to other competing systems. Then we present our Pareto selection algorithm and prove its validity for all three formulations. Finally, we provide numerical results aimed at understanding how well our algorithm performs in various settings.

Finally, we discuss the estimation of input distributions when theoretical distributions do not provide a good fit to existing data. Our approach is to use a quasi-empirical distribution, which is a mixture of an empirical distribution and a distribution for the right tail. We describe an existing approach that involves an exponential tail distribution, and adapt the approach to incorporate a Pareto tail distribution and to use a different cutoff point between the empirical and tail distributions. Then, to measure the impact, we simulate a stable M/G/1 queue with a known inter-arrival and unknown service time distributions, and estimate the mean and tail probabilities of the waiting time in queue using the different approaches. The results suggest that if we know that the system is stable, and suspect that the tail of the service time distribution is not exponential, then a quasi-empirical distribution with a Pareto tail works well, but with a lower bound imposed on the tail index.

CHAPTER I

INTRODUCTION

When it comes to evaluating large and complex systems in real life, finding closed-form solution can be very difficult. In such cases, simulation can come to the rescue. However, despite advances in technology, realistic simulations of real life phenomena are increasingly becoming more challenging. Also, many problems in our modern world require a prompt solution; thus speed of simulation matters. In this thesis, we discuss advances in simulation, with focus on the efficiency and validity of the simulation.

In the first part of the thesis, we seek for efficiency via cloning. In our study, simulation cloning combines conventional cloning and splitting ideas. It achieves computational saving by sharing some parts of the simulation results among several replications. Doing so may lead to a problem of growing variance, because sharing computations induces positive correlation within replications that adds to the variance. To improve performance, we introduce a modified version of simulation cloning utilizing induced negative correlation between replications. We also address implementation issues such as identifying the optimal number of branches sharing a common part of a simulation for both basic and modified cloning. The objective is to achieve maximum efficiency, which is defined to be the reciprocal of the product of the variance and computational effort per replication.

In the second part of the thesis, we focus on ranking and selection. Although ranking and selection has been widely studied, most works have concentrated on choosing the best system(s) based on a single objective. However, in real life, decisions are usually based on multiple attributes. For example, when deciding what to have for lunch, one naturally considers the food itself, price, time, and distance, to name a few. It is not surprising that when the decision to make becomes more significant, then it may not suffice to investigate a single objective. Therefore, we study Pareto set estimation using ranking and selection. Instead of choosing a single best system, the goal is to select a set of Pareto optimal systems. A system is Pareto optimal if there exists no other system that can improve upon it in one objective without hurting some other objective(s).

"Garbage in, garbage out" is one of the best-known words of wisdom in the world of simulation. Thus, estimating input distributions correctly is an essential part of simulation. In the third part of the thesis, we pursue the validity of simulation by identifying better input distributions. In many cases, a set of well-known distributions (e.g., exponential, normal, gamma, etc.) are not sufficient to describe the behavior of stochastic input quantities. When the input distribution is distinctively different from well-known distributions, combining some distributions may help improve the fit. For example, Bratley et al. [15] suggests the use of quasi-empirical distributions, which are mixtures of empirical and exponential distributions. As there may be distributions that cannot be approximated using an exponential distribution, we combine other distributions for a better fit and study the effects of such algorithms via simulation of an M/G/1 queue.

This thesis is organized as following. In Chapter 2, we describe related works for each topic we cover. In Chapter 3, we propose a simulation cloning algorithm to promote efficiency in transient simulation. In Chapter 4, we propose an algorithm that estimates the Pareto optimal set when multiple objectives are present. Finally, Chapter 5 addresses difficult input analysis problem where theoretical distributions may not provide a good fit. We conclude this thesis with a summary and description of future work in Chapter 6. Additional numerical results for Chapter 5 are provided in Appendix A.

CHAPTER II

LITERATURE REVIEW

In this chapter, we review previous works and discuss their contribution and how they motivated our work. In Section 2.1, we describe works related to simulation cloning (see Chapter 3) and in Section 2.2, we review contributions relevant to Pareto set estimation (which we introduce in Chapter 4). Finally in Section 2.3, literature that is pertinent to difficult input analysis problems (see Chapter 5) is provided.

2.1 Simulation Cloning

Hybinette and Fujimoto [60] introduced a cloning mechanism that can be used to simulate different scenarios more efficiently. Their mechanism shares some simulation results to save computer effort in simulating different scenarios. More specifically, if multiple scenarios have common path until a decision point where the scenarios start to differ, then simulation cloning shares simulation results up until that point. This research was continued by Hybinette and Fujimoto [61] who studied the impact of the number of clones for simulation problems of different sizes. Hybinette [59] suggested just-in-time cloning to delay the decision point as far as possible, with the goal of making the computational savings more significant. Chen et al. [23, 24] provided an architecture, mechanism, and design for managing simulation cloning. They also studied an incremental cloning mechanism for distributed simulation based on the HLA (High-Level Architecture) standard.

In a simulation context, cloning resembles the variance reduction approach known as splitting. While splitting was originally suggested in a particle transmission setting [62], Bayes [11] suggested the concept of splitting in simulation while using the term importance sampling, and Hopmans and Kleijnen [56] applied the idea to a complex system but found the result to be disappointing due to the increased net variance.

Other research on splitting includes Villén-Altamirano and Villén-Altamirano [89, 90] who introduced the RESTART (REpetitive Simulations Trials After Reaching Thresholds) method, which is one application of classical splitting. Their study of the RESTART method continued, see, for example, [91, 92] and a recent tutorial on RESTART for applications [93]. Also, Schreiber and Görg [49] modified the RESTART method and successfully applied it to several finite buffered queueing systems. Garvels [39] extended and unified existing splitting methods and analyzed the importance function. The results were extended by Garvels and Kroese [40] who compared different implementations of an existing RESTART method and suggested the best strategy, and by Garvels, Van Ommeran, and Kroese [41] who studied the importance function in splitting. Lagnoux [64] also analyzed an importance splitting model that divides the state space into regions called importance regions. His proposed algorithm minimizes the variance under a fixed budget. Lagnoux studied splitting under a cost constraint in [66, 65].

More research on splitting includes Glasserman, Heidelberger, Shahabuddin, and Zajic [43, 44], who resolved the issue of choosing the number of subpaths to generate when a path is split and provided a proof of the method being optimal in a general setting, and Cérou and Guyader [17] who studied adaptive multilevel splitting to find splitting levels without much advance knowledge about the system. Cérou, LeGland, Del Moral, and Lezaud [18] derived limit theorems for estimating rare-event probabilities. Other variants of multilevel splitting include the work of Del Moral et al. [31, 32] who referred to multilevel splitting as the Feynman-Kac model. L'Ecuyer, Demer, and Tuffin [68, 69] provide an introduction to splitting techniques, introduce some ways to improve their implementation by combining them with randomized quasi-Monte Carlo, and also give examples of application where the techniques can be effective and not.

2.2 Pareto Set Estimation

Choosing the best system(s) from multiple candidate systems has been vastly studied. Among many algorithms that serve the purpose, we will briefly describe the most relevant ones, which include Ranking and Selection, Optimal Computing Budget Allocation (OCBA), and Bayesian methods.

Ranking and Selection (R&S) is one of the best known approaches for choosing the best from a set of systems and a vast amount of research has been done on this topic. To name a few papers, Bechhofer [12] suggested a ranking and selection procedure for systems with known variances. Bechhofer, Elmaghraby, and Morse [13] introduced the problem of selecting the multinomial event with highest probability. They also calculated the probabilities of correct selection when a lower bound on the ratio of the best and the second best performance measures is given. Paulson [80] developed an algorithm to choose the k best systems when systems have normal populations with known or unknown common variances, and the algorithm was improved by Hartmann [50]. Dudewicz and Dalal [33] considered the case with unequal and unknown variances, and Rinott [84] adapted the algorithm for better performance. Miller, Nelson, and Reilly [75] increased the efficiency of the technique by using pseudo replications that are not independent, when the original set of independent samples is not large enough to achieve the desired probability of correct selection. Kim and Nelson [63] presented a fully sequential procedure suited for the case when a small amount of additional sampling can be done repeatedly. Nelson, Swann, Goldsman, and Song [77] provided a two-stage procedure where some candidates are eliminated in the first stage, and a sequential approach is used to select the best system in the second stage.

Chen [19, 20] introduced the OCBA technique whose objective is to maximize the approximated confidence probability of correct selection with a given amount of computational budget by deciding how to divide the budget among the systems. When applying the OCBA technique, solving an optimization problem to decide the sample sizes is crucial. Chen et al. [26] considered a new OCBA technique that solves this optimization problem using a gradient method. Chen et al. [22] developed an asymptotic allocation rule that uses an approximate probability of correct selection to identify the optimal allocation analytically. Glynn and Juneja [46] used large deviation theory on top of the OCBA approach to enable the use of the OCBA when the performance measures of interest are not Gaussian.

Finally, there are some approaches using the Bayesian method. Chick and Inoue [27, 28] suggested two-stage and sequential procedures based on a Bayesian model, and used common random numbers in the latter. Frazier et al. [38] studied a similar approach with a different stopping criterion under slightly more strict assumptions, and found that these modifications can improve the results significantly in many cases.

The literature reviewed so far in this section concentrates on the single objective case. While ranking and selection of systems based on a single objective is a mature field of study, there is less literature covering the multiple objective case. Butler, Morrice and Mullarkey [16] suggested applying multiple utility theory to transform the multi-objective problem into a single objective problem. Baesler aned Sepúlveda [6] also transformed the problem into a single objective problem by using a goal programming framework and applied their algorithm to a simulation model of a new cancer treatment center [7]. Santner and Tamhane [86] solved the problem of achieving maximum mean and minimum variance in an approximate sense by selecting a set with reasonably large means and small variances. Batur and Choobineh [10] also studied the case when both the mean and variance are measures of performance using a fully sequential procedure for comparing systems. Andradóttir and Kim [3] consider two performance measures, with one being the primary objective and the other one imposing a constraint, and suggested fully sequential procedures for comparing such systems. Healey, Andradóttir, and Kim [51] developed the dormancy concept to further enhance efficiency. By making non-promising systems dormant, additional sampling can be avoided unless a dormant system turns out to be more promising than first thought (because the system that dominated it is infeasible). They also expanded the fully sequential procedures to incorporate the multiple constraints case [52].

Several studies have been performed using the OCBA approach with multiple objectives. In Lee et al. [70], a primary performance measure is selected as the objective, with the other performance measures being constraints. Hunter et al. [57, 58] proposed a budget allocation algorithm under stochastic constraints using large deviation theory as in Glynn and Juneja [46]. In other works, the Pareto concept is used to decide what are the best systems in the presence of multiple objectives. In particular, Lee, Chew, and Teng [71] proposed a solution framework that addresses multi-objective problems with huge solution spaces and high uncertainty in performance measures. They integrated heuristic search and OCBA to find non-dominated systems. Chen and Lee [25] suggested a two-stage algorithm under the assumption that the performance measures are independent of each other. They select an incomplete Pareto set that only contains the systems with the most promising performance measures in any objective in the first stage, and complete it by selecting the other non-dominated systems in the second stage. Lee et al. [72] proposed a multi-objective optimal computing budget allocation (MOCBA) algorithm to allocate computing budget to systems to minimize type I and type II errors in selecting non-dominated systems. Teng, Lee, and Chew [88] incorporated an indifference-zone approach into the aforementioned MOCBA algorithm.

2.3 Difficult Input Analysis Problem: Experiments using the M/G/1 Queue

There exist numerous ways to estimate input distributions given sample data points that represent the distribution of interest. Therefore, input analysis has been widely studied. Most standard simulation textbooks now have a good chapter dedicated to the subject, including Law and Kelton [67], Fishman [37], and Banks et al. [8], just to name a few. There are also a number of commercial distribution-fitting software packages. In general, these packages make recommendations on the best fitted theoretical distribution, and also provide graphical views for heuristic decisions. For example, Expertfit, Easyfit, and the Input Analyzer of Arena are widely used.

While aforementioned algorithms and software packages concentrate on fitting existing theoretical distributions, there may be cases where those distributions cannot provide a good fit. For a remedy in such case, Bratley, Fox, and Schrage [15] suggest to use a quasi-empirical distribution. A quasi-empirical distribution is a mixture of an empirical and theoretical distribution, where they used an exponential for the latter.

However, the importance of heavy-tails in real-life data has been prominent, and probably more so after their work. Data is heavy-tailed when the probability of occurrences of extreme values is higher, and the tail behavior of these distributions cannot be well approximated with the exponential. In diverse fields, various phenomena exhibit the behavior that generates the heavy-tailed data. Teletraffic is one of the areas that are known to be rich in heavy-tailed data, as can be seen in Willinger et al. [96], Resnick [82], and references therein, for example. Financial data and insurance risks have also seen a significant amount of heavy-tailed phenomena, as can be seen in Mandelbrot [74] and Müller et al. [76]. Nature and human dynamics also display various heavy-tailed behavior, as can be seen in Barabasi [9] and Newman [78]. Finally, there are books that provide overview, examples, and theory on the subject. Interested readers can start from [34, 35, 79, 83] to list just a few. Heyde and Kou [54] pointed out that the tail of a distribution heavily drives the behavior of many performance measures of probabilistic models, and the distinction of the behavior when the tail is light-tailed and heavy-tailed is significant. Therefore, we would like to study and experiment with whether adding a right tail to an empirical distribution is still efficient when the underlying distribution is suspected to have a heavy-tail, and if so, what type of right tails should be used.

Naturally, as the interest and importance of the heavy-tailed data emerged, fitting distributions to the potentially heavy-tailed data received attention. There is a vast amount of literature on this subject. An approach to estimate a tail index, proposed by Hill [55], is one of the most popular methods. There have been numerous studies that have improved Hill's estimator in some direction under different circumstances. De Haan and Peng [30] compared some of the available approaches in their work. The endeavor to improve Hill's estimator continued, as in Resnick and Stărică [81], Gomes et al. [48] and Alves [2], for example. More recently, Clauset et al. [29] described an approach that fits a power-law distribution to an empirical data set, if applicable, using Hill's estimator.

On the other hand, even if an input distribution is known to be heavy-tailed, simulation or analysis of a system involved can be challenging, and also has been worked on widely. We conclude this chapter by describing various works on the single-server M/G/1 queueing system when the service time is heavy-tailed. Abate et al. [1] approximated the steady-state waiting time distribution when the service time distribution is heavy-tailed and the Laplace transforms of the inter-arrival and service time distributions are known. Boxma and Cohen [14] also provide an asymptotic series for the tail probabilities of the waiting time in the system. Feldmann and Whitt [36] used mixtures of exponential distributions to predict the performance of a M/G/1 queue when the service time is heavy-tailed. Asmussen and Kroese [5] suggest two simulation estimators that have bounded relative error and involve using importance sampling and Monte Carlo conditioning to estimate the tail probability of system times when the service time distribution is heavy-tailed.

CHAPTER III

SIMULATION CLONING WITH INDUCED NEGATIVE CORRELATIONS

3.1 Introduction

Simulation cloning is an algorithm to enhance efficiency by sharing some calculations. Hybinette and Fujimoto [60] introduced an efficient cloning mechanism. They concentrated on the simulation of complex systems with multiple “decision points” where clones are generated to simulate different sample paths. A decision point is an instance in time where it is of interest to consider multiple scenarios. It may not necessarily involve any decisions, but rather a deviation, such as a change of environment. Computation is shared until the decision points and cloning is done by utilizing a construct called virtual logical processes (LPs). A (physical) LP is a unit that depicts some subset of the system that is simulated separately to the extent possible. At decision points, rather than cloning the physical LPs and simulating them separately for each scenario, multiple virtual LPs, which map to the same physical LP, are created. These virtual LPs will serve like physical LPs but will utilize the calculations of one physical LP in different clones. Thus, a physical LP is only cloned when the results for that physical LP differ among sample paths. In this way, the computation of the LPs without difference between scenarios will be shared among the clones.

For example, when simulating the air traffic system, physical LPs for each airport, say Houston and Atlanta, will be generated. When a storm takes place in Houston, then this will have direct effect on Houston but no direct effect on Atlanta for the time being. Hence two virtual LPs for Atlanta (considering sample paths with or without the storm in Houston) can share the computation of one physical LP for Atlanta. When the storm starts affecting Atlanta, a new physical LP for Atlanta will be generated to incorporate the difference. This technique is shown to significantly reduce the time required to compute multiple alternate futures [60].

In a simulation context, cloning resembles the variance reduction approach known as splitting. Splitting is closely related to the work of Kahn and Harris [62]. In a particle transmission setting, they explained splitting as follows. "Whenever a particle passes from a less important region, it is split in two. Each of the resulting particles is given one-half of the weight of the original particle and is treated independently from then on." Splitting was then widely studied as a variance reduction technique, especially to calculate rare-event probabilities. When used as a variance reduction technique, the idea remains the same. When a sample path reaches a threshold, the path is split into several subpaths of lesser weight, so the paths can reach the desired rare-event region more often.

Splitting and cloning are similar in that both approaches share certain computations. Splitting can be considered as cloning all LPs based on the states of sample paths. However, the performance measure of interest in recent research has generally been limited to a rare-event probability. While cloning concentrates on efficiently evaluating each possible scenario by simulating it once, splitting considers the precision of statistical estimates for a single scenario. Standard output analysis for splitting, which will be used in this chapter, can be found in Section V.5 of Asmussen and Glynn [4].

One complication that can occur in simulation cloning used with space-parallel simulation of a system represented by several LPs, is that future events can interfere with already simulated events. As a result, a rollback algorithm is required. In the previous air traffic example, as the physical LP for Atlanta airport proceeds, the delayed flight information from Houston can arrive after the original arrival time. In this case, the Atlanta LP would have simulated the arrival of a plane which did not, in fact, arrive on time. This requires the simulation to roll back to the point where the error occurred. However, this complication does not cause a problem for splitting, because simulation of the paths after the splitting point has no impact on the path before that point.

In our work, we allow simulation cloning to be flexible. In particular, cloning can take place even if no decision is made and no changes of setting or circumstances have occurred. While in classical splitting, splitting only occurs when a path reaches a threshold, which is a stochastic event, we can set the splitting points to practically any times along the simulation. For example, we can set a decision point to be a certain predefined time point, such as 10 minutes in simulation time, or at any physical point in a feedforward system, such as when a job leaves the third station in a network of five tandem stations.

Simulation cloning at a certain time point has two benefits. First, we do not need to worry about the system being feedforward. Thus, rollback will not be required as future, by definition, cannot interfere with the past. Secondly, we can control the occurrence of cloning with ease. While in splitting, the threshold may never be reached in a replication, if we define the splitting point properly, that is, to be shorter that the simulation time, cloning can be guaranteed to occur. Thus, the number of branches or replications can be easily controlled.

Simulation cloning in this chapter is especially efficient when the latter part of the simulation is more important or has larger variability than the front part. For example, when simulating a battlefield, the beginning of the simulation may involve mere searching and marching and the actual battle may take place after a fair amount of simulation time. If the battle has the greatest impact on the outcome, simulation cloning may be effective. Similarly, when forecasting the demand over a long period, then naturally, the variability of the farther future will be larger. Then in the output analysis, the part with larger variability may have greater impact on the performance measure, and thus require more replications to achieve better confidence. Glasserman and Staum [45] also studied the effects of allocating different amounts of simulation effort to different time steps. However, they concentrate on decreasing the effort by deliberately stopping some sample paths early, whereas our approach increases the simulation effort by cloning some simulation paths.

We consider simulation cloning in the context of transient simulation. Transient simulations, in general, are to be replicated independently multiple times, with the independence among replications being used to obtain statistical reliability. Starting from the idea of sharing some computation effort, we study the use of simulation cloning for efficient transient simulation with statistical reliability. In this approach, simulation is run to decision points where the simulation will be branched into a predefined number of clones. However, in simulation cloning, as some parts of the simulation is shared among replications, independence of observations cannot be insured with traditional output analysis. To retain independence, the notion of a replication has to be redefined. Basically, a replication is defined to be the simulations that share any part of the computation.

The outline of this chapter follows. Section 3.2 describes the basic cloning algorithm. We improve the method by inducing negative correlation, as described in Section 3.3. In particular, we pair each replication from basic cloning with one counterpart that incorporates negative correlation. In Section 3.4, cloning with two decision points is introduced and analyzed. This shows that cloning with multiple decision points is feasible, but of course the analysis and optimal implementation become more involved. In Section 3.5, numerical results provide better understanding and show the benefits of cloning. Finally, Section 3.6 concludes the chapter.

3.2 Basic Cloning

Basic cloning involves branching at a predefined point to a given number of branches, and the simulation results before branching are shared among the clones. The performance measure of interest is defined to be $\varphi(X, Y)$, where X is the vector of random quantities observed before the branching point, and Y is the corresponding vector observed after the branching point. Let $\mu = E[\varphi(X, Y)]$. This is illustrated in Figure 1, where X_r is the vector of random variables that is generated from the r-th replication of the shared part of the simulation and Y_{rs}, the s-th clone in the r-th replication, is the vector of random variables that are generated from clone s after the branching point. To prevent further unnecessary positive correlation, X_r for $r = 1, 2, \ldots, R$, and Y_{rs} for $s = 1, 2, \ldots, S, \forall r$, are independently simulated. Define $\rho_{cn} = \text{Cor}(\varphi(X_1, Y_{11}), \varphi(X_1, Y_{12}))$ as the correlation between replications that share some part of the simulation. In the above notation, c stands for "common," and n stands for "not common." The interpretation of this notation is that the first vector X is positively correlated, and that vector Y is simulated independently.

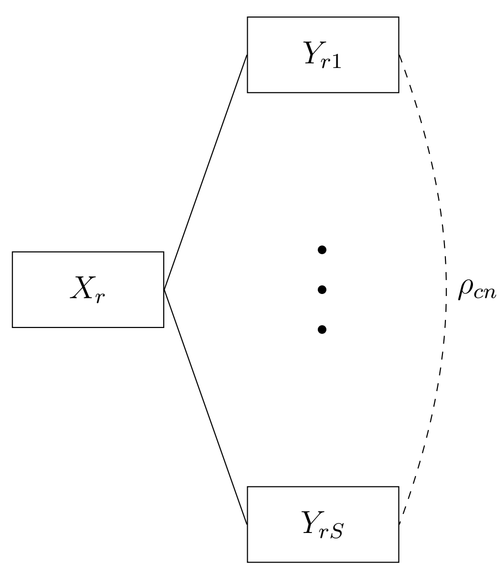

Figure 1: A Single Replication in Basic Cloning

The observations of the performance measure μ is denoted as $V_{rs} = \varphi(X_r, Y_{rs})$.

Therefore, if R is the number of replications and S is the number of clones generated at the branching point, then

$$V_r = \frac{1}{S}\sum_{s=1}^{S}\varphi(X_r, Y_{rs}), \text{for } r = 1, \ldots, R,$$

are independent and identically distributed, and our estimator of the performance measure μ is computed as follows:

$$\hat{\mu} = \frac{1}{RS}\sum_{r=1}^{R}\sum_{s=1}^{S}\varphi(X_r, Y_{rs}).$$

This is as described in Section V.5 of Asmussen and Glynn [4]. Thus a unit depicted in Figure 1 is considered as one replication to ensure independence in basic cloning. In words, replication r involves simulating X_r and Y_{rs} for $s = 1, \ldots, S$, for any r.

To measure the benefit of cloning quantitatively as a function of S, the efficiency of the simulation, denoted by $EF(S)$, can be calculated as follows:

$$EF(S) = \frac{1}{\text{Var}(V_r) \times \text{Effort}(V_r)},$$

where $\text{Var}(V_r)$ denotes the variance associated with one replication of a simulation and $\text{Effort}(V_r)$ is the computational cost involved in simulating the single replication. Using this definition for the efficiency of the simulation is justified in Glynn and Whitt [47].

Let e_X denote $\text{Effort}(X_r)$, e_Y denote $\text{Effort}(Y_{rs})$, and σ^2 denote $\text{Var}(V_{rs})$. We assume that $\text{Effort}(V_{rs}) = e_X + e_Y$ and $\text{Effort}(V_r) = e_X + Se_Y$, ignoring the computational effort associated with branching and averaging the performance measure. Let $f = e_X \;/\; (e_X + e_Y)$ be the fraction of computer effort saved by sharing one observation of X among two observations of Y. We assume that $0 < f < 1$. Thus, $e_Y = e_X[(1-f)/f]$. Then f should be significantly large for a substantial improvement in efficiency. This is because the variance of each replication,

$$\text{Var}(V_r) = \frac{\sigma^2(1 + (S-1)\rho_{cn})}{S}, \tag{1}$$

quickly increases as ρ_{cn} increases. Thus the efficiency of basic cloning is given by:

$$EF(S) = \frac{S}{\sigma^2(1+(S-1)\rho_{cn})(e_X + Se_Y)}. \tag{2}$$

To find the number of branches S^* that maximizes the efficiency, we take the derivative of the reciprocal of the efficiency as follows (ignoring for the moment that S should be integer):

$$\frac{d(\frac{1}{EF(S)})}{dS} = \frac{\sigma^2(S^2\rho_{cn}e_Y - (1-\rho_{cn})e_X)}{S^2},$$

and find the value of S that sets the value equal to 0, if possible. The approach is justified as the second derivative of the function is always non-negative:

$$\frac{d^2(\frac{1}{EF(S)})}{(dS)^2} = \frac{2\sigma^2(1-\rho_{cn})e_X}{S^3} \geq 0.$$

Note that Equation (1) implies that $\rho_{cn} \geq 0$. If $\rho_{cn} > 0$, this yields the optimal S value as follows:

$$S_{opt} = \max\left\{1, \sqrt{\frac{(1-\rho_{cn})f}{\rho_{cn}(1-f)}}\right\}. \tag{3}$$

Then the number of the branches S^* should be either $\lfloor S_{opt} \rfloor$ or $\lceil S_{opt} \rceil$ as the number should be integer. It is sufficient to consider adjacent integers because the function is convex. If $\rho_{cn} = 0$, then it is obvious that S should be as large as possible, as the derivative will be negative.

The ratio of the efficiency of the basic cloning approach with S clones to the crude Monte Carlo approach is calculated as $R(S) = EF(S)/EF(1)$. We have

$$R(S) = \frac{S(e_X + e_Y)}{(1+(S-1)\rho_{cn})(e_X + Se_Y)} = \frac{S}{(1+(S-1)\rho_{cn})(f + S(1-f))}.$$

When cloning is performed, that is, if $S \geq 2$, the ratio increases as ρ_{cn} decreases, and as f increases.

The notion of $R^* = R(S^*)$ is introduced to define the ratio of the efficiency derived from using the optimal value of S^*. Note that when $f \geq \rho_{cn}$, then $S_{opt} = \sqrt{\frac{(1-\rho_{cn})f}{\rho_{cn}(1-f)}}$

and

$$
\begin{aligned}
R(S_{opt}) &= \frac{\sqrt{\frac{(1-\rho_{cn})f}{\rho_{cn}(1-f)}}}{(1-\rho_{cn})f + \sqrt{\frac{(1-\rho_{cn})^3 f(1-f)}{\rho_{cn}}} + \sqrt{\frac{\rho_{cn}(1-\rho_{cn})f^3}{(1-f)}} + f(1-\rho_{cn})} \\
&= \frac{1}{1 - f - \rho_{cn} + 2\rho_{cn}f + 2\sqrt{\rho_{cn}(1-\rho_{cn})f(1-f)}} \\
&= \frac{1}{(\sqrt{\rho_{cn}f} + \sqrt{(1-\rho_{cn})(1-f)})^2}. \qquad (4)
\end{aligned}
$$

It is now clear that $R(S_{opt})$ increases as f increases for any given ρ_{cn}, provided that $f \geq \rho_{cn}$. Similarly, $R(S_{opt})$ increases as ρ_{cn} decreases for any $f \geq \rho_{cn}$. If this tendency is pursued to the extreme, so that ρ_{cn} approaches 0 and f approaches 1, then the ratio $R(S_{opt})$ goes to infinity.

To illustrate, we present a simple numerical analysis. If $\rho_{cn} = 0.5$ and $f = 0.5$, then S^* and R^* will be 1. To get significant improvements, ρ_{cn} must be small and f must be large. For example, if $\rho_{cn} = 0.1$ and $f = 0.9$, then $R^* \simeq 2.78$ with $S^* = 9$, and if $\rho_{cn} = 0.01$ and $f = 0.99$, then $R^* \simeq 25.3$ with $S^* = 99$. In realistic settings, these figures are difficult to achieve. This is because to have f large generally means that we share significant amounts of calculation, and this increases ρ_{cn}. For example, when the split point is approximately halfway, that is, f is approximately 0.5, the correlation ρ_{cn} can easily be greater than 0.5 as half of the simulation shares the common results.

As seen in the previous example, when cloning is the only technique that is used and no special structure is exploited, the increase in variance can outgrow the benefits of the savings from cloning, or the benefit can be small. This is because of the positive correlation, ρ_{cn}, between V_{rj} and V_{rk} for different j and k that is due to sharing the common part X_r of replication r between Y_{rj} and Y_{rk}. To mitigate the effect of undesirable positive correlation, we introduce simulation cloning using induced negative correlation in the next section.

3.3 Cloning Using Induced Negative Correlation

There are different ways to apply cloning and induced negative correlation. In the following subsections, we consider two ways of employing the techniques. In Section 3.3.1, we introduce negative correlation before and after the decision point, whereas in Section 3.3.2, we only introduce negative correlation after the decision point.

3.3.1 Negative correlation before and after the decision point

To offset the increase of the variance from positive correlation that can be seen in Equation (1), we pair two replications from basic cloning and induce negative correlation between the replications. This is depicted in Figure 2, where X'_r and Y'_{rs} denote the vectors of random variables paired with X_r and Y_{rs} with induced negative correlation. As in previous section, X_r, for $r = 1, 2, \ldots, R$, and Y_{rs} for $s = 1, 2, \ldots, S$, are simulated independently to avoid introducing unnecessary positive correlation. Define

$$\rho_{ll} = \mathrm{Cor}(\varphi(X_1, Y_{11}), \varphi(X'_1, Y'_{11})),$$

$$\rho_{ln} = \mathrm{Cor}(\varphi(X_1, Y_{11}), \varphi(X'_1, Y'_{12})).$$

In the above notation, l stands for "linked" via any method for inducing negative correlation, and n now stands for "not common nor linked." Then ρ_{ll} is the correlation between the branches using random variables with induced negative correlation for both the cloned and replicated parts and ρ_{ln} is the correlation between the branches using negative correlation for only the cloned part. Therefore, it is natural to expect that $0 > \rho_{ln} > \rho_{ll}$. The best known method for inducing negative correlation involves simulating the second replication using the antithetic variables of the variables in the first replication. Details can be found in [85]. There are other ways to induce negative correlation, see, e.g., Henderson, Chiera, and Cooke [53], and any methodology that fits the purpose can be used. In the following, we provide a more detailed analysis of cloning with induced negative correlation.

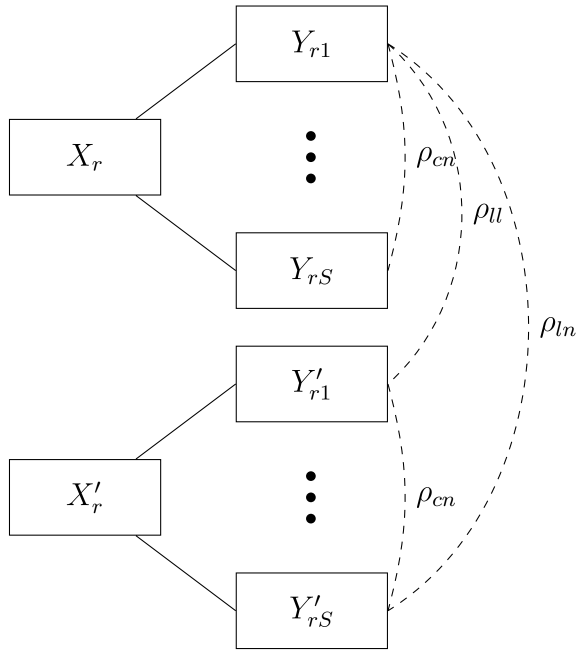

Figure 2: A Single Replication in Cloning Using Induced Negative Correlation Before and After the Decision Point

Before further analysis, additional notation is needed. In particular, the observations of the performance measure μ are denoted as $V_{rs} = \varphi(X_r, Y_{rs})$ and $V'_{rs} = \varphi(X'_r, Y'_{rs})$. Therefore, a replication is now defined as a unit including X_r, X'_r, Y_{rs}, and Y'_{rs} for $s = 1, \ldots, S$. Therefore, an observation from one replication is as follows:

$$V'_r = \frac{1}{2S} \sum_{s=1}^{S} (\varphi(X_r, Y_{rs}) + \varphi(X'_r, Y'_{rs})), \text{for } r = 1, \ldots, R.$$

Then, the estimator of the performance measure should be calculated as follows:

$$\hat{\mu}' = \frac{1}{2RS} \sum_{r=1}^{R} \sum_{s=1}^{S} (\varphi(X_r, Y_{rs}) + \varphi(X'_r, Y'_{rs})).$$

Let $e_X = \text{Effort}(X_r) = \text{Effort}(X'_r)$, $e_Y = \text{Effort}(Y_{rs}) = \text{Effort}(Y'_{rs})$, and $\sigma^2 = \text{Var}(V_{rs}) = \text{Var}(V'_{rs})$. Then, the variance and effort of one replication are as follows:

$$\text{Var}(V'_r) = \frac{\sigma^2}{2S}[1 + \rho_{ll} + (S-1)(\rho_{cn} + \rho_{ln})], \tag{5}$$

$$\text{Effort}(V'_r) = 2(e_X + Se_Y).$$

The efficiency of the cloning is again the multiplication of the reciprocal effort and variance, which turns out to be

$$EF'(S) = \frac{S}{\sigma^2[1+\rho_{ll}+(S-1)(\rho_{cn}+\rho_{ln})](e_X+Se_Y)}. \tag{6}$$

To find the value of S that maximizes the efficiency, we first compute the first two derivatives of the reciprocal of efficiency,

$$\frac{d(\frac{1}{EF'(S)})}{dS} = \sigma^2 \frac{S^2(\rho_{cn}+\rho_{ln})e_Y - (1+\rho_{ll}-\rho_{cn}-\rho_{ln})e_X}{S^2},$$

$$\frac{d^2(\frac{1}{EF'(S)})}{(dS)^2} = \frac{2\sigma^2(1+\rho_{ll}-\rho_{cn}-\rho_{ln})e_X}{S^3}.$$

Note that Equation (5) implies that $\rho_{cn}+\rho_{ln} \geq 0$. If $1+\rho_{ll}-\rho_{cn}-\rho_{ln} \leq 0$, then the problem is trivial as $EF'(S)$ is non-increasing in S, and hence it is always better not to clone at all. Also, if $1+\rho_{ll}-\rho_{cn}-\rho_{ln} > 0$ and $\rho_{cn}+\rho_{ln} = 0$, then the problem is again trivial as $EF'(S)$ is non-decreasing in S. Thus, we only consider the case when $1+\rho_{ll}-\rho_{cn}-\rho_{ln} > 0$ and $\rho_{cn}+\rho_{ln} > 0$ for further analysis. By setting the first derivative to zero, we can obtain the S'_{opt} that minimizes the reciprocal of efficiency, and hence maximizes the efficiency of cloning, as follows:

$$S'_{opt} = \max\left\{1, \sqrt{\frac{(1+\rho_{ll}-\rho_{cn}-\rho_{ln})f}{(\rho_{cn}+\rho_{ln})(1-f)}}\right\}. \tag{7}$$

To find the true optimum with the restriction of S being integer-valued, it is sufficient to consider $\lfloor S'_{opt} \rfloor$ and $\lceil S'_{opt} \rceil$ as the reciprocal of efficiency is a convex function of S when $1+\rho_{ll}-\rho_{cn}-\rho_{ln} \geq 0$. Note, however, that the optimal number of clones S^* depends on f, ρ_{cn}, ρ_{ll}, and ρ_{ln}. Obtaining the exact values of correlations or computational effort may be difficult, and, hence, the number of branches can be picked arbitrarily if desired. The cloning approach with induced negative correlation is summarized in Algorithm 1.

To briefly show the benefit of inducing negative correlation, we again define the efficiency ratio to be the efficiency of cloning with negative correlation divided by that

Algorithm 1 (Splitting with induced negative correlation)

```
Predefine the splitting point and desired number of replications
if the number of clones should be optimal then
    Retrieve required information on variances, covariances, and effort by running
    a short simulation or using some known information
    Compute S'_opt using Equation (7)
    Compare the efficiency EF'(⌊S'_opt⌋) and EF'(⌈S'_opt⌉) to decide the optimal num-
    ber S*
else
    Specify the number of clones S*
end if
while Desired number of replications is not reached do
    Proceed with the simulation until the decision point is reached
    Generate the predefined number S* of clones
    Finish the simulation of the S* clones
    Start the simulation using random variables with induced negative correlation
    against the run just finished. This also concludes with simulation of S* clones
end while
Perform the output analysis
```

of crude Monte Carlo simulation. Then the efficiency ratio with the optimal number of clones S'_{opt} is given by

$$\begin{aligned} R'(S'_{opt}) &= \frac{EF'(S'_{opt})}{EF(1)} \\ &= \frac{S'_{opt}(e_X + e_Y)}{[1 + \rho_{ll} + (S'_{opt} - 1)(\rho_{cn} + \rho_{ln})](e_X + S'_{opt} e_Y)} \\ &= \frac{S'_{opt}}{[1 + \rho_{ll} + (S'_{opt} - 1)(\rho_{cn} + \rho_{ln})](f + S'_{opt}(1 - f))}. \end{aligned}$$

When $f \geq \frac{\rho_{cn}+\rho_{ln}}{1+\rho_{ll}}$ holds, then $S'_{opt} = \sqrt{\frac{(1+\rho_{ll}-\rho_{cn}-\rho_{ln})f}{(\rho_{cn}+\rho_{ln})(1-f)}}$. In this case, $R'(S'_{opt})$ can be expressed as follows:

$$\begin{aligned} R'(S'_{opt}) &= \frac{\sqrt{\frac{(1+\rho_{ll}-\rho_{cn}-\rho_{ln})f}{(\rho_{cn}+\rho_{ln})(1-f)}}}{2((1 + \rho_{ll} - \rho_{cn} - \rho_{ln})f + \sqrt{\frac{(1+\rho_{ll}-\rho_{cn}-\rho_{ln})(\rho_{cn}+\rho_{ln})f^3}{1-f}} + \sqrt{\frac{(1+\rho_{ll}-\rho_{cn}-\rho_{ln})^3 f(1-f)}{\rho_{cn}+\rho_{ln}}})} \\ &= \frac{1}{(\sqrt{(\rho_{cn} + \rho_{ln})f} + \sqrt{(1 + \rho_{ll} - \rho_{cn} - \rho_{ln})(1 - f)})^2}. \end{aligned} \quad (8)$$

Equation (8) shows two limiting behaviors. If f goes to 1 and $\rho_{cn} + \rho_{ln}$ goes to 0, then S'_{opt} and the ratio $R'(S'_{opt})$ will go to infinity regardless of the behavior of ρ_{ll}.

Also, if ρ_{ll} approaches -1, and $\rho_{cn}+\rho_{ln}$ approaches 0, the ratio $R'(S'_{opt})$ goes to infinity regardless of the behavior of $f \geq \frac{\rho_{cn}+\rho_{ln}}{1+\rho_{ll}}$. When this inequality does not hold, then $S'_{opt} = 1$, and thus the ratio $R'(S'_{opt})$ goes to infinity simply if ρ_{ll} approaches -1.

Note that the structure of Equation (8) resembles that of Equation (4). It can be seen that $\rho_{cn} + \rho_{ln}$ is used in lieu of ρ_{cn}, and $1 + \rho_{ll} - \rho_{cn} - \rho_{ln}$ appears in place of $1-\rho_{cn}$. Therefore, $R'(S'_{opt})$ in Equation (8) is always greater than $R(S_{opt})$ in Equation (4), when $0 > \rho_{ln} > \rho_{ll}$ holds. Earlier in this section, we discussed the validity of this inequality. It is also noteworthy that the ratio of Equation (8) to Equation (4), that is, $\frac{R'(S'_{opt})}{R(S_{opt})}$, approaches infinity when one of the two aforementioned conditions for $R'(S'_{opt})$ to go infinity holds, with the additional condition that ρ_{cn} does not converge to zero.

We provide a simple numerical example. Throughout the example, we let $\rho_{cn} = -2\rho_{ln}$. This choice is motivated by the fact that as ρ_{cn} is the correlation from sharing computation and ρ_{ln} is the correlation from inducing negative correlation, it is natural to expect the former to be positive, and the latter to be negative. Also, previously in this section, we showed that $\rho_{cn} + \rho_{ln} \geq 0$. For comparison with examples from the previous section, we let ρ_{cn} and f be the same, and let $\rho_{ll} = -(1-\rho_{cn})$ to have similar limiting behavior as in the previous section. First, let $\rho_{cn} = 0.5$ and $f = 0.5$. Then the best possible ratio is $R^* = 2$ with $S^* = 1$. This shows the extreme case where negative correlation plays the key role. As we proceed with the same correlation scheme, let $\rho_{cn} = 0.1$ and $f = 0.9$, then the ratio increases to $R^* \approx 12.5$ with $S^* = 3$. Finally, when we have $\rho_{cn} = 0.01$ and $f = 0.99$, then we get $R^* \approx 166.81$ with $S^* = 10$. In the previous section without induced negative correlation, the corresponding ratios were 1, 2.78, and 25.3, respectively. Although these are extreme cases, they illustrate the potential benefits of inducing negative correlation.

3.3.2 Negative correlation after the decision point

In this subsection, we will introduce an alternative approach of employing cloning with negative correlation. Unless stated otherwise, all notation is inherited from the previous section. Instead of having a negatively correlated pair of cloned parts X_r and X'_r as in the previous subsection, a replication is a unit consisting of X_r, Y_{rs}, and Y'_{rs} for $s = 1, \ldots, S$. This is depicted in Figure 3, where the new notation $\rho_{cl} = \text{Cor}(\varphi(X_1, Y_{11}), \varphi(X_1, Y'_{11}))$ is introduced to denote the new negative correlation induced by the cloned part.

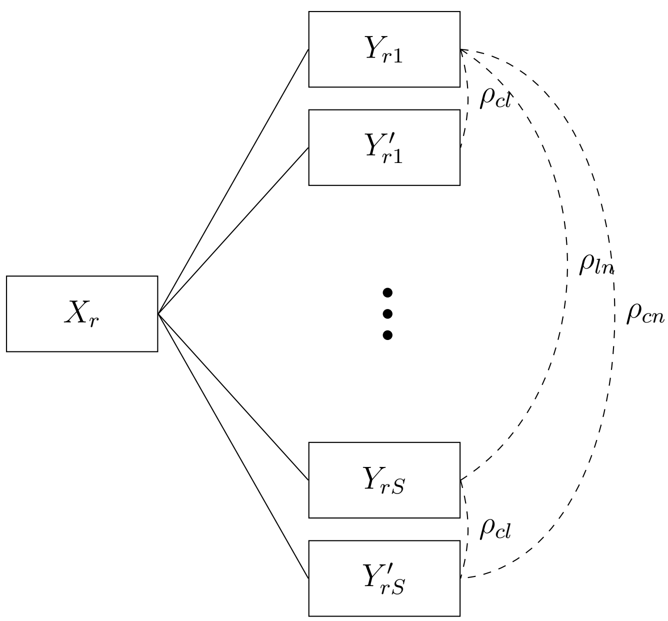

Figure 3: A Single Replication in Cloning Using Induced Negative Correlation Only After the Decision Point

In this approach, the observations of the performance measure μ are denoted as $V_{rs} = \varphi(X_r, Y_{rs})$ and $V'_{rs} = \varphi(X_r, Y'_{rs})$. An observation from one replication is as follows:

$$\bar{V}'_r = \frac{1}{2S} \sum_{s=1}^{S} (\varphi(X_r, Y_{rs}) + \varphi(X_r, Y'_{rs})), \text{for } r = 1, \ldots, R.$$

Then, the estimator of the performance measure should be calculated as follows:

$$\hat{\bar{\mu}}' = \frac{1}{2RS} \sum_{r=1}^{R} \sum_{s=1}^{S} (\varphi(X_r, Y_{rs}) + \varphi(X_r, Y'_{rs})).$$

Moreover, the variance and effort of one replication are as follows:

$$\mathrm{Var}(\bar{V}_r') = \frac{\sigma^2}{2S}[1 + \rho_{cl} + 2(S-1)\rho_{cn}],$$

$$\mathrm{Effort}(\bar{V}_r') = e_X + 2Se_Y.$$

Therefore the efficiency can be calculated as follows:

$$\bar{EF}'(S) = \frac{2S}{\sigma^2[1 + \rho_{cl} + 2(S-1)\rho_{cn}](e_X + 2Se_Y)}.$$

We again look into the reciprocal of this efficiency for simplicity. The first derivative of the reciprocal is as follows:

$$\frac{d(\frac{1}{\bar{EF}'(S)})}{dS} = \frac{\sigma^2[4\rho_{cn}e_Y S^2 - (1 + \rho_{cl} - 2\rho_{cn})e_X]}{2S^2}.$$

This problem is trivial when $1 + \rho_{cl} - 2\rho_{cn} \leq 0$, as $\bar{EF}'(S)$ is non-increasing in S in such a case. Thus, if $1 + \rho_{cl} - 2\rho_{cn} \leq 0$, then $S^* = 1$. On the other hand, if $1 + \rho_{cl} - 2\rho_{cn} > 0$, then the second derivative:

$$\frac{d^2(\frac{1}{\bar{EF}'(S)})}{(dS)^2} = \frac{\sigma^2(1 + \rho_{cl} - 2\rho_{cn})e_X}{S^3} > 0,$$

so that the function is convex. Thus, by setting the first derivative equal to zero, we obtain

$$\bar{S}_{opt}' = \max\left\{1, \sqrt{\frac{(1 + \rho_{cl} - 2\rho_{cn})f}{4\rho_{cn}(1-f)}}\right\}.$$

When $f \geq \frac{4\rho_{cn}}{1+\rho_{cl}+2\rho_{cn}}$ holds, then

$$\bar{S}_{opt}' = \sqrt{\frac{(1 + \rho_{cl} - 2\rho_{cn})f}{4\rho_{cn}(1-f)}}.$$

The ratio of efficiency to crude Monte Carlo, $\bar{R}'(\bar{S}'_{opt})$, is then given as follows:

$$\begin{aligned}
\bar{R}'(\bar{S}'_{opt}) = \frac{E\bar{F}'(\bar{S}'_{opt})}{EF(1)} &= \frac{2\bar{S}'_{opt}(e_X + e_Y)}{[1+\rho_{cl}+2(\bar{S}'_{opt}-1)\rho_{cn}](e_X + 2\bar{S}'_{opt}e_Y)} \\
&= \frac{2\bar{S}'_{opt}}{[1+\rho_{cl}+2(\bar{S}'_{opt}-1)\rho_{cn}](f + 2(1-f)\bar{S}'_{opt})} \\
&= \frac{\sqrt{\frac{(1+\rho_{cl}-2\rho_{cn})f}{\rho_{cn}(1-f)}}}{2(1+\rho_{cl}-2\rho_{cn})f + \sqrt{\frac{(1+\rho_{cl}-2\rho_{cn})^3 f(1-f)}{\rho_{cn}}} + \sqrt{\frac{(1+\rho_{cl}-2\rho_{cn})f^3\rho_{cn}}{1-f}}} \\
&= \frac{1}{(\sqrt{\rho_{cn}f} + \sqrt{(1+\rho_{cl}-2\rho_{cn})(1-f)})^2}. \qquad (9)
\end{aligned}$$

Clearly, the structure of this ratio resembles that of basic cloning in Equation (4). The only difference is that $1-\rho_{cn}$ in Equation (4) is substituted with $1+\rho_{cl}-2\rho_{cn}$. As ρ_{cl} combines the positive correlation of shared part and the negative correlation of cloned part, while ρ_{cn} is the positive correlation only from shared part, $\rho_{cl} \leq \rho_{cn}$ should be natural. Therefore, cloning after the decision point should yield benefits over basic cloning. Similar comparison can be done with Equation (8). In Equation (8), $\rho_{cn}+\rho_{ln}$, and $1+\rho_{ll}-\rho_{cn}-\rho_{ln}$ are used in place of ρ_{cn} and $1+\rho_{cl}-2\rho_{cn}$, respectively. Clearly, $\rho_{cn} \geq \rho_{cn}+\rho_{ln}$ is expected. However, $1+\rho_{cl}-2\rho_{cn}$ and $1+\rho_{ll}-\rho_{cn}-\rho_{ln}$ are expected to be similar as the differences $\rho_{cl}-\rho_{cn}$ and $\rho_{ll}-\rho_{ln}$ both involve comparing systems with and without negative correlation in the cloned part. This suggests that cloning before and after the decision point will generally yield better results than cloning only after the decision point.

Finally, Equation (9) suggests that just as in basic cloning, when ρ_{cn} approaches 0 and f approaches 1, the ratio approaches infinity regardless of ρ_{cl}. Additionally, when ρ_{cn} goes to 0 and ρ_{cl} goes to -1, the ratio goes to infinity regardless of f. The ratio of Equation (9) to Equation (4), that is, $\frac{\bar{R}'(\bar{S}'_{opt})}{R(S_{opt})}$ goes to infinity for the latter scenario if f does not converge to one. Note that ratio of Equation (9) to Equation (8), $\frac{\bar{R}'(\bar{S}'_{opt})}{R'(S'_{opt})}$, goes to zero if $\rho_{cn}+\rho_{ln}$ goes to 0 and f goes to 1 without ρ_{cn} going to zero, or if $\rho_{cn}+\rho_{ln}$ goes to 0 and ρ_{ll} goes to -1 with either ρ_{cn} or $1+\rho_{cl}-2\rho_{cn}$ not going to zero, regardless of the value of $0 < f < 1$.

To apply a similar correlation scheme as in Section 3.3.1, we let $\rho_{cl} = \rho_{ll} + \rho_{cn}$. This scheme is selected with the intuition that ρ_{cl} is greater than ρ_{ll} by the amount of positive correlation incurred by sharing the first part of the simulation instead of inducing negative correlation. Actually, $\rho_{cl} > \rho_{ll} + \rho_{cn}$ is expected, as ρ_{ll} is the negative correlation not only for the cloned part but also the shared part. Therefore, this scheme can be viewed as favorable to the algorithm introduced in this section. If $\rho_{cn} = 1 + \rho_{ll}$ as in Section 3.3.1, then $\rho_{cl} = 2\rho_{cn} - 1$. Let $\rho_{cn} = f = 0.5$, then $S^* = 1$ and $R^* \simeq 1.33$. When $\rho_{cn} = 0.1$ and $f = 0.9$, then $R^* \simeq 9.09$ with $S^* = 1$, and if $\rho_{cn} = 0.01$ and $f = 0.99$, then $R^* \simeq 99.01$ with $S^* = 1$. This is not as efficient as in the previous section where negative correlation is induced before and after the decision point; the corresponding ratios were 2, 12.5, and 166.81. However, compared to basic cloning, where the corresponding values were 1, 2.78, and 25.3, respectively, this is a significant improvement.

The examples suggest that inducing negative correlation only after the decision point is generally more efficient than basic cloning, but less efficient than inducing negative correlation both before and after the decision point. This is intuitive as inducing negative correlation before the decision point impacts all clones, whereas inducing negative correlation after the decision point only impacts pairs of clones. Although the former approach involves more computation per observation, the additional generation of one sample path prior to the decision point is likely to be worthwhile. Therefore, from now on we will concentrate on inducing negative correlation both before and after the decision point, as described in Section 3.3.1.

3.4 Cloning at Two Decision Points

In the previous two sections, we discussed simulation cloning with one decision point. However, there may be cases where it is beneficial to have multiple decision points. In this section, we describe cloning using induced negative correlation with splitting at two decision points. As we already have seen in Section 3.3 that using negative correlation before and after the decision point is desirable, we will do so here. Figure 4 illustrates this case.

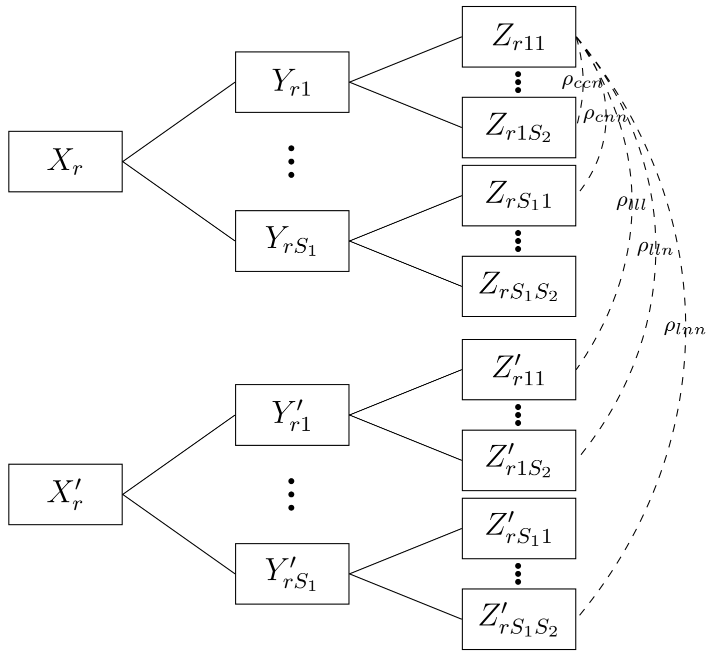

Figure 4: A Single Replication in Cloning with Two Decision Points and Induced Negative Correlation

Let the performance measure of interest be $\varphi(X, Y, Z)$, where Z is added to the previous notation to denote the vector of random quantities observed after the second decision point. Similarly, let the extended notation $Z_{rs_1s_2}$ denote the vector of random variables generated from the s_2-th clone branched from the s_1-th clone of the first decision point in the r-th replication. The observations of the performance measure μ can be denoted as $V_{rs_1s_2} = \varphi(X_r, Y_{rs_1}, Z_{rs_1s_2})$ and $V'_{rs_1s_2} = \varphi(X'_r, Y'_{rs_1}, Z'_{rs_1s_2})$, where X'_r, Y'_{rs_1}, $Z'_{rs_1s_2}$ denote the random variables negatively correlated with X_r, Y_{rs_1}, $Z_{rs_1s_2}$, respectively. As in the previous sections, to avoid introducing positive correlation, X_r, Y_{rs_1}, and $Z_{rs_1s_2}$, $\forall r, s_1, s_2$, are simulated independently. Define correlations as follows:

$$\rho_{ccn} = \mathrm{Cor}(\varphi(X_1, Y_{11}, Z_{111}), \varphi(X_1, Y_{11}, Z_{112})),$$

$$\rho_{cnn} = \mathrm{Cor}(\varphi(X_1, Y_{11}, Z_{111}), \varphi(X_1, Y_{12}, Z_{121})),$$

$$\rho_{lll} = \mathrm{Cor}(\varphi(X_1, Y_{11}, Z_{111}), \varphi(X_1', Y_{11}', Z_{111}')),$$

$$\rho_{lln} = \mathrm{Cor}(\varphi(X_1, Y_{11}, Z_{111}), \varphi(X_1', Y_{11}', Z_{112}')),$$

$$\rho_{lnn} = \mathrm{Cor}(\varphi(X_1, Y_{11}, Z_{111}), \varphi(X_1', Y_{12}', Z_{121}')).$$

Subscripts are used in accordance with the definitions from the previous sections. That is, c, l, and n stand for "common," "linked," and "not common nor linked," respectively. For example, ρ_{lln} is the correlation incurred by inducing negative correlation for first two parts before the second decision point, and having the last part independently simulated. Also, as discussed earlier, if more parts are negatively correlated, then the negative correlation should be larger, and when more parts are shared, then the positive correlation should be larger. Thus, we expect $1 \geq \rho_{ccn} \geq \rho_{cnn} \geq 0$ and $0 \geq \rho_{lnn} \geq \rho_{lln} \geq \rho_{lll} \geq -1$. Note that $\rho_{cnn} + \rho_{lnn} \geq 0$ and $\rho_{ccn} + \rho_{lln} \geq 0$ holds as $\rho_{cn} + \rho_{ln} \geq 0$.

A unit depicted in Figure 4 constitutes one replication when performing the output analysis. It includes $X_r, X_r', Y_{rs_1}, Y_{rs_1}', Z_{rs_1s_2}$, and $Z_{rs_1s_2}'$ for $s_1 = 1, 2, \ldots, S_1$, and $s_2 = 1, 2, \ldots, S_2$, where S_1 and S_2 are the number of clones at the first and second decision points, respectively. Therefore, an observation from one replication is as follows:

$$V_r'' = \frac{1}{2S_1S_2} \sum_{s_1=1}^{S_1} \sum_{s_2=1}^{S_2} (\varphi(X_r, Y_{rs_1}, Z_{rs_1s_2}) + \varphi(X_r', Y_{rs_1}', Z_{rs_1s_2}')), \text{for } r = 1, \ldots, R.$$

Then, the estimator of the performance measure should be calculated as follows:

$$\hat{\mu}'' = \frac{1}{2RS_1S_2} \sum_{r=1}^{R} \sum_{s_1=1}^{S_1} \sum_{s_2=1}^{S_2} (\varphi(X_r, Y_{rs_1}, Z_{rs_1s_2}) + \varphi(X_r', Y_{rs_1}', Z_{rs_1s_2}')).$$

Also, σ^2, which denotes the variance of a base replication before any cloning, now equals $\text{Var}(V_{rs_1s_2}) = \text{Var}(V'_{rs_1s_2})$. Finally, we add the notation for the effort of the last part of the simulation, and let $e_Z = \text{Effort}(Z_{rs_1s_2}) = \text{Effort}(Z'_{rs_1s_2})$.

Let $c_1 = (1 + \rho_{lll} - \rho_{ccn} - \rho_{lln})$, $c_2 = \rho_{cnn} + \rho_{lnn}$, and $c_3 = \rho_{cnn} + \rho_{lnn} - \rho_{ccn} - \rho_{lln}$. Note that $c_2 \geq 0$, and $c_2 \geq c_3$ from the earlier discussion. With the given notation, the variance and effort of one replication are as follows:

$$\begin{aligned}\text{Var}(V''_r) &= \frac{\sigma^2}{2S_1S_2}[1 + (S_2-1)\rho_{ccn} + (S_1-1)S_2\rho_{cnn} + \rho_{lll} + (S_2-1)\rho_{lln} + (S_1-1)S_2\rho_{lnn}] \\ &= \sigma^2\left[\frac{c_1}{2S_1S_2} + \frac{c_2}{2} - \frac{c_3}{2S_1}\right],\end{aligned}$$

$$\text{Effort}(V''_r) = 2(e_X + S_1e_Y + S_1S_2e_Z).$$

We start by discussing the optimal number of replicating branches, denoted S_1^* and S_2^*, for the second and third parts of the simulation. To calculate S_1^* and S_2^*, the efficiency of the simulation should be calculated first. To simplify the calculation, we again find the minimum of the reciprocal of the efficiency (i.e., the product of variance and effort). This is calculated as follows:

$$\frac{1}{EF(S_1,S_2)} = \sigma^2\left[\left(\frac{c_1}{S_1S_2} + c_2 - \frac{c_3}{S_1}\right)e_X + \left(\frac{c_1}{S_2} + S_1c_2 - c_3\right)e_Y + (c_1 + S_1S_2c_2 - S_2c_3)e_Z\right].$$

To find the optimal values of S_1 and S_2, we calculate the partial derivatives with respect to each S_1 and S_2, as follows:

$$\frac{\partial\left(\frac{1}{EF(S_1,S_2)}\right)}{\partial S_1} = \sigma^2\left(-\frac{c_1}{S_1^2S_2}e_X + \frac{c_3}{S_1^2}e_X + c_2e_Y + S_2c_2e_Z\right),$$

$$\frac{\partial\left(\frac{1}{EF(S_1,S_2)}\right)}{\partial S_2} = \sigma^2\left(-\frac{c_1}{S_1S_2^2}e_X - \frac{c_1}{S_2^2}e_Y + S_1c_2e_Z - c_3e_Z\right).$$

Note that if $c_1 \leq c_3$, then the efficiency is nonincreasing in S_1, thus the problem becomes equivalent to a single decision point case. Similarly, if $c_1 \leq 0$, then the efficiency is nonincreasing in S_2 (as $c_2 \geq c_3$), and thus our problem again reduces to the single decision point problem. We can therefore concentrate on the case where $c_1 > 0$ and $c_3 < c_1$. Setting the aforementioned first derivatives equal to zero simplifies as follows:

$$S_1^2 = \frac{c_1 e_X - c_3 e_X S_2}{c_2 e_Y S_2 + S_2^2 c_2 e_Z}, \tag{10}$$

$$S_2^2 = \frac{c_1 e_X + c_1 e_Y S_1}{c_2 e_Z S_1^2 - c_3 e_Z S_1}, \tag{11}$$

when $c_2 > 0$ and $c_2 > c_3$. When the values of the variances and covariances are known, general mathematical software can solve this system of equations. The resulting stationary point is actually the minimum if the determinant of the Hessian matrix is positive. The elements of the Hessian matrix are given as follows:

$$\frac{\partial^2(\frac{1}{EF(S_1,S_2)})}{(\partial S_1)^2} = 2\sigma^2(\frac{c_1}{S_1^3 S_2} e_X - \frac{c_3}{S_1^3} e_X),$$

$$\frac{\partial^2(\frac{1}{EF(S_1,S_2)})}{(\partial S_2)^2} = 2\sigma^2(\frac{c_1}{S_1 S_2^3} e_X + \frac{c_1}{S_2^3} e_Y),$$

$$\frac{\partial^2(\frac{1}{EF(S_1,S_2)})}{\partial S_1 \partial S_2} = \sigma^2(\frac{c_1}{S_1^2 S_2^2} e_X + c_2 e_Z).$$

However, even if the reciprocal of the efficiency is convex, the optimal values obtained from solving the system of equations (10) and (11), will be, in general, non-integer valued. Therefore, a heuristic search around the optimum has to be performed. Heuristic search can be efficient even without solving the system of equations (10)-(11) as the values of S_1 and S_2 are integer and bounded below by 1. The search can be stopped at a local maximum (S_1, S_2) when $EF(S_1, S_2) > EF(S_1 + i, S_2 + j)$ holds $\forall i, j = -1, 0, 1$ with $(i, j) \neq (0, 0)$. However, solving the equations will facilitate initiating the search for the optimal number of clones.

Numerical results can be obtained using a similar scheme as in Section 3.3. The results are shown in Table 1. Let $\rho_{ccn} = \rho_{cn}$ and $f = \frac{e_X + e_Y}{e_X + e_Y + e_Z} = \frac{e_X}{e_X + e_Y}$. This implies that $\frac{e_X}{e_Y} = \frac{f}{1-f}$, $\frac{e_X}{e_Z} = \frac{f^2}{1-f}$, and $\frac{e_Y}{e_Z} = f$. Note that it is only these ratios of efforts that affect equations (10) and (11). Then, $\rho_{lln} = \rho_{ln}$ and $\rho_{lll} = \rho_{ll}$ are natural choices. For the remaining parameter values, we suggest two different schemes. The first three rows of Table 1 show the numerical results with $\rho_{cnn} = (\rho_{ccn})^2$ and $\rho_{lnn} = -(\rho_{lln})^2$.

The next three rows show the results with $2\rho_{cnn} = \rho_{ccn}$ and $2\rho_{lnn} = \rho_{lln}$. The last two columns show the desirable decision point when only one decision point is to be applied, the optimal number of clones, and the efficiency ratio of such case.

Table 1: Numerical Analysis for Two Decision Points Using Scheme from Previous Sections

ρ_{ccn}	ρ_{cnn}	ρ_{lll}	ρ_{lln}	ρ_{lnn}	f	S_1^*	S_2^*	R^*	Single(S^*)	Single R^*
0.5	0.25	-0.5	-0.25	-0.0625	0.5	1	1	2	S_1 or $S_2(1)$	2
0.1	0.01	-0.9	-0.05	-0.0025	0.9	7	1	22.56	$S_1(7)$	22.56
0.01	0.0001	-0.99	-0.005	-0.000025	0.99	81	1	1953.13	$S_1(81)$	1953.13
0.5	0.25	-0.5	-0.25	-0.125	0.5	1	1	2	S_1 or $S_2(1)$	2
0.1	0.05	-0.9	-0.05	-0.025	0.9	4	1	14.56	$S_1(4)$	14.56
0.01	0.005	-0.99	-0.005	-0.0025	0.99	12	1	262.53	$S_1(12)$	262.53

Table 1 shows that the as positive correlations get smaller and the negative correlations and ratio of effort f get larger, the efficiency ratio increases significantly in both scenarios. Also comparing the two scenarios, we can see that the smaller magnitude of ρ_{cnn} and ρ_{lnn} in the first scenario leads to greater efficiency gain with all other values fixed. Note, however, that none of the cases are better off by adding a decision point, as at least one of S_1^* or S_2^* is optimal at one. However, this is not always the case. In Table 2, we consider the same scenarios as in Table 1, but with larger values of the negative correlation ρ_{lll}. As $\rho_{lnn} = -(\rho_{lln})^2$ was used in the first scenario in Table 1, we apply $\rho_{lnn} = (\rho_{lll})^3$ for the first three rows of Table 2. The next three rows of Table 2 show the results with $\rho_{lnn} = 3\rho_{lll}$, which is consistent with $2\rho_{lnn} = \rho_{lln}$ in the second scenario of Table 1.

Table 2: Numerical Analysis for Two Decision Points Using Different Scheme

ρ_{ccn}	ρ_{cnn}	ρ_{lll}	ρ_{lln}	ρ_{lnn}	f	S_1^*	S_2^*	R^*	Single(S^*)	Single R^*
0.5	0.25	-0.3969	-0.25	-0.0625	0.5	1	1	1.66	S_1 or $S_2(1)$	1.66
0.1	0.01	-0.1357	-0.05	-0.0025	0.9	7	4	5.53	$S_1(22)$	4.31
0.01	0.0001	-0.02924	-0.005	-0.000025	0.99	46	19	76.36	$S_1(795)$	45.93
0.5	0.25	-0.375	-0.25	-0.125	0.5	1	1	1.60	S_1 or $S_2(1)$	1.60
0.1	0.05	-0.075	-0.05	-0.025	0.9	3	6	4.24	$S_2(13)$	3.87
0.01	0.005	-0.0075	-0.005	-0.0025	0.99	10	20	42.25	$S_1(140)$	34.71

As in Table 1, Table 2 also shows significantly increasing efficiency ratio as positive correlations get smaller and the negative correlations and ratio of effort f get larger. As ρ_{ll} is relatively smaller in this table, the total benefit decreases compared to Table 1. However, unlike Table 1, we can see that adding an additional decision point increases the efficiency ratio ranging from 9 to 66 percent.

It is noteworthy that the technique of cloning with induced negative correlation can be expanded to multiple decision points. Moreover, as can be seen from the last two rows in both scenarios of Table 2, there are certain special cases when having additional decision point provides benefit. However, we have performed additional experiments that are not presented in the chapter to conserve space, and found that in many cases the computational savings associated with multiple decision points are not substantial. Thus, from now on we concentrate on a single decision point.

3.5 Simulation Results

In this section, we present numerical results for cloning in two settings. In Section 3.5.1, the use of cloning to simulate queueing networks with ten stations connected in tandem is addressed. In Section 3.5.2, cloning results for a profit model with uncertain supply and demand are demonstrated. In both sections, cloning with negative correlation before and after a single decision point is employed, and basic cloning and simply using the negative correlation are also considered for comparison

3.5.1 Ten queueing stations in tandem

In this section, we consider ten G/G/1 queueing stations connected in tandem. The performance measure of interest is the sojourn time of the 100-th customer for the study of transient simulation. We set the decision point to be at the end of a specified station d. Thus, when cloning is applied, the first d stations are shared, and the remaining $10 - d$ stations are cloned. As can be seen from Equation (6), the efficiency is a function of the variance, covariances, and effort. For the experiment, we chose different interarrival and service time distributions to affect the variance and covariances, and chose different decision points to change the effort. Bottlenecks in this experiment are the stations with lower service rate. In some cases, bottleneck stations also have greater service time variability, so that the variance of the performance measure of interest at the station is greater.

In Tables 3 and 4, we provide the optimal number of clones and efficiency ratios for both basic cloning and cloning with negative correlation before and after the decision point for different interarrival and service time distributions, and hence different bottlenecks. Also, to see the effects of using negative correlation, the efficiency ratio of negative correlation without any cloning is provided. Finally, estimated correlation values are presented to provide information about the magnitude of these correlations. The correlation values are estimated through simulation, and the estimates of ρ_{cn}, ρ_{ll}, and ρ_{ln} are denoted as $\hat{\rho}_{cn}$, $\hat{\rho}_{ll}$, and $\hat{\rho}_{ln}$, respectively. The optimal numbers of branches, $\hat{S}^*$ and the efficiency ratios, $\hat{R}(\hat{S}^*)$, are computed using the estimated correlation values and equations (3) and (4) for basic cloning. Also, for ARN and cloning with ARN, the optimal numbers of branches, denoted $\hat{S}^*$ instead of $\hat{S}'^*$ for better readability and the efficiency ratios, $\hat{R}'(\hat{S}^*)$, are computed using the estimated correlation values and equations equations (2) and (6), and equations (7) and (8), respectively. Note that the efficiency ratio of ARN is equal to the efficiency ratio of cloning with ARN when the number of branches equals 1. This is denoted $\hat{R}'(1)$.

If not specified otherwise, the decision point d is set to five, so that the computational effort before and after the decision point is similar. In fact, throughout the analysis, we assume that the computational effort is linear with respect to the number of stations. That is, when the decision point is d, we use $\hat{f} = d/10$ to estimate f. As this ignores the effort associated with generating interarrival times, which form a portion of the shared part of the system, it is likely that $f \geq \hat{f}$. The set of bottleneck stations is indicated in the first row of Tables 3 and 4. Also, if the distributions of the bottleneck change, it is noted in the first row. The interarrival and service time distributions are chosen as follows.

- In Table 3, the interarrival and service times are exponentially distributed. The mean interarrival time is set to 5 and the mean service time at non-bottleneck stations is set to 1. In the first three columns, the mean service time of bottleneck stations equals 4. For the last two columns, we increase the utilization ρ of the bottleneck station 10 to 0.9 and 0.98, by setting the mean service time equal to 4.5 and 4.9, respectively.

- In Table 4, all times are uniformly distributed, denoted U(a, b), where a and b are the lower and upper bounds of the uniform distribution, respectively. This allows us to separate the effects of the mean and variance of the service time distribution. The distribution of the interarrival times is U(4, 6) and non-bottleneck service times are U(0, 2). In the first three columns, the bottleneck service times are U(3, 5), thus the utilization $\rho = 0.8$ is consistent with the first three columns of Table 3. In the last three columns, to assess the effects of the variance, we keep the utilization at 0.8, and use different lower and upper bounds for the service time distribution, as in U(2, 6), U(1, 7), and U(0, 8).

- All the correlation values are estimated with a large enough number of replication that all confidence interval widths are less than 10^{-3}. Also, when the correlation values are smaller than 10^{-3}, more replications were performed to ensure that the values shown in the tables have at least one significant digit (with the exception of Table 5 where the confidence intervals in the first, second and third columns are as wide as the absolute value of the estimated correlation values).

As can be seen from Tables 3 and 4 using basic cloning does not improve the results by a great margin, if any (the improvement ranges from 0 to 50 percent).

Table 3: Tandem Queue with Exponential Interarrival and Service Times

	Bottlenecks	st. 10	st. 9	st. 9, 10	st. 10, $\rho = 0.9$	st. 10, $\rho = 0.98$
Correlations	$\hat{\rho}_{cn}$	0.3452	0.3454	0.4392	0.3622	0.3730
	$\hat{\rho}_{ll}$	-0.3480	-0.3426	-0.3730	-0.3930	-0.4573
	$\hat{\rho}_{ln}$	-0.1987	-0.1993	-0.2612	-0.2130	-0.2298
Basic Cloning	$\hat{S}^*$	1	1	1	1	1
	$\hat{R}(\hat{S}^*)$	1	1	1	1	1
ARN Only	$\hat{R}'(1)$	1.5337	1.5211	1.5948	1.6473	1.8425
Cloning with ARN	$\hat{S}^*$	2	2	2	2	2
	$\hat{R}'(\hat{S}^*)$	1.6698	1.6592	1.6562	1.7631	1.9437

Table 4: Tandem Queue with Uniform Interarrival and Service Times

	Bottlenecks	st. 10	st. 9	st. 9, 10	st. 10 w/ U(2, 6)	st. 10 w/ U(1, 7)	st. 10 w/ U(0, 8)
Correlations	$\hat{\rho}_{cn}$	0.3708	0.3599	0.3613	0.1991	0.0957	0.0485
	$\hat{\rho}_{ll}$	-0.5077	-0.5919	-0.4820	-0.6059	-0.6451	-0.6302
	$\hat{\rho}_{ln}$	-0.2427	-0.2294	-0.2115	-0.1378	-0.0721	-0.0399
Basic Cloning	$\hat{S}^*$	1	1	1	2	3	4
	$\hat{R}(\hat{S}^*)$	1	1	1	1.1120	1.2591	1.3968
ARN Only	$\hat{R}'(1)$	2.0314	2.4504	1.9304	2.5375	2.8176	2.7040
Cloning with ARN	$\hat{S}^*$	2	2	2	2	4	6
	$\hat{R}'(\hat{S}^*)$	2.1491	2.4754	1.9966	2.9282	3.7590	4.1519

Also, as it is widely known from the literature, and also can be seen from our results, the use of antithetic variables improves computational efficiency (in our case, up to approximately 280 percent). However, when cloning is combined with antithetic variables to induce negative correlation, then the computational savings become more significant. Though some of the savings comes from the antithetic variables, combining the two clearly brings better results and the improvement is achieved at a low cost as the approach is easy to implement. In Table 3, we can see that as the utilization gets larger in the bottleneck, the effect of cloning with induced negative correlation increases. Tables 4 shows the effect of the variance of the service time distribution of bottleneck stations. In particular, the greater the variability of the cloned part, the greater the efficiency gain. Finally, comparing the first, second, and third columns of Tables 3 and 4, we can conclude that the position of the bottleneck is does not have a major effect on the results, as long as it is located after the decision point.

From the simulation results, we can see that the traffic intensity and the variability of the bottleneck are key factors that affect the efficiency, with more influence resulting from the variability change. Therefore, the efficiency increases as the cloned part becomes more important (i.e., as the traffic intensity and variability of the bottleneck increases), and exploiting the structure of planned simulation is helpful when using the cloning technique.

Sensitivity analysis with respect to the decision point d is shown in Tables 5 and 6. In these tables, we assume that $f = d/10$. In Table 5, interarrival times are set to be deterministic and service times exponential with $\rho = 0.98$. In this example, as the last two stations form the bottleneck, the best results are obtained when the decision point is after the 8th station. When the decision point is set after the 9th station, although the computing effort decreases, positive correlation greatly increases, and thus cloning becomes less effective. In this case, with the deterministic arrival and large traffic intensity at the bottleneck, we can see significant improvement in efficiency. Finally in

Table 6, interarrival times are set to be deterministic and service times to be uniformly distributed. Nonbottleneck service times follow U(0, 2) and bottleneck service time follow U(30, 50) for higher variability. As the last station is the bottleneck with higher variability, the efficiency ratio increases as the decision point is delayed to the end. Naturally, the optimal decision point is at the 9th station.

Table 5: Different Decision Points for Tandem Queue with Deterministic Arrival and Exponential Service Times and Bottleneck at the 9th and 10th Stations with $\rho = 0.98$

	Decision Point	1	2	5	8	9
Correlations	$\hat{\rho}_{cn}$	0.000095	0.000119	0.000244	0.000389	0.4204
	$\hat{\rho}_{ll}$	-0.3302	-0.3302	-0.3302	-0.3302	-0.3302
	$\hat{\rho}_{ln}$	-0.000019	-0.000004	-0.000038	-0.000058	-0.1220
Basic Cloning	$\hat{S}^*$	34	46	64	101	4
	$\hat{R}(\hat{S}^*)$	1.1040	1.2366	1.9394	4.6293	1.3608
ARN Only	$\hat{R}(1)$	1.4930	1.4930	1.4930	1.4930	1.4930
Cloning with ARN	$\hat{S}^*$	31	38	57	90	3
	$\hat{R}(\hat{S}^*)$	1.6474	1.8423	2.8850	6.8459	1.9740

Table 6: Different Decision Points for Tandem Queue with Deterministic Arrival and Uniform Service Times and Bottleneck at the 10th Station with $\rho = 0.8$ and High Variability

	Decision Point	1	2	5	8	9
Correlations	$\hat{\rho}_{cn}$	0.008326	0.0168	0.0416	0.0668	0.0751
	$\hat{\rho}_{ll}$	-0.9948	-0.9948	-0.9948	-0.9948	-0.9948
	$\hat{\rho}_{ln}$	-0.008324	-0.0168	-0.0415	-0.0663	-0.0746
Basic Cloning	$\hat{S}^*$	4	4	5	7	11
	$\hat{R}(\hat{S}^*)$	1.0547	1.1200	1.4288	2.2719	3.1402
ARN Only	$\hat{R}(1)$	192.2581	192.2581	192.2581	192.2581	192.2581
Cloning with ARN	$\hat{S}^*$	18	9	6	7	9
	$\hat{R}(\hat{S}^*)$	211.0138	228.3665	283.9188	402.7631	521.0764

As can be seen from Tables 5 and 6, it is beneficial to share large parts of the simulation to save computational effort. Nevertheless, the decision point should be chosen carefully so that the key part of the simulation is not shared as in the last column of Tables 5, as the efficiency ratio is not at optimal if the decision point is delayed to the last station. However, it is noteworthy that the last column of Table 5 show fair performance although the choice of decision point is not optimal.

As we mentioned earlier, in Tables 3 through 6, we estimated the correlations ρ_{cn}, ρ_{ll}, and ρ_{ln}, and used these estimated correlations to estimate the optimal number of clones and optimal efficiency ratios. However, in practice, estimated correlation values may be unavailable or imprecise. Therefore, we now perform sensitivity analysis with respect to the number of clones. In Figure 5, parameters from a tandem queue with uniform interarrival and service time distributions are used, corresponding to the second column of Table 4. Although the efficiency ratio is optimal at $S = 2$, cloning is beneficial up to ten clones. In Figure 6, using data from the last column of Table 4, cloning will be beneficial up to 189 clones. In Figure 7, values from the last column of Table 6 is used. In this case, although the optimal number of clones is 9, significant benefit can be achieved for large range of number. Even at $S = 5000$, cloning is beneficial. Figures 5, 6, and 7 suggest that even if the number of clones are not chosen to be optimal, we can achieve efficiency improvement with a reasonable number of clones.

3.5.2 Profit model with uncertain supply and demand

In this section, we consider a simple profit model with uncertain supply and demand. This model is naturally feedforward because the supply is observed prior to the demand, and hence partitioning the system into supply and demand components is equivalent to partitioning it along the time horizon. Thus, the supply is shared, and the demand is cloned. Another aspect of this model is that demand tends to have higher variability compared to supply, as supply is generally easier to control for the party that is interested in this profit model.

We consider a basic, linear profit model. In this model, cost is proportional to the supply and revenue is proportional to the actual sales, which is the minimum of

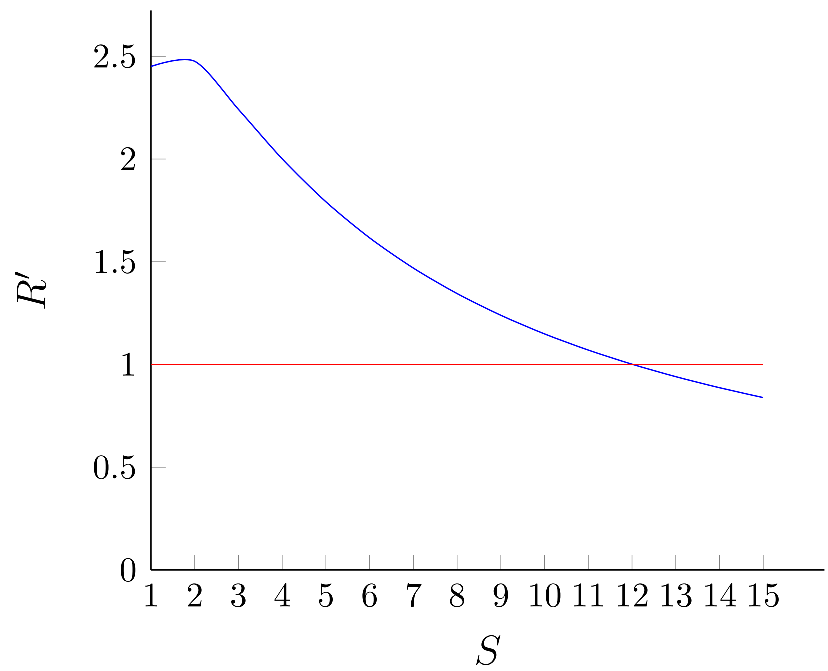

Figure 5: Sensitivity Analysis for Uniform Times with Bottleneck at 9th Station

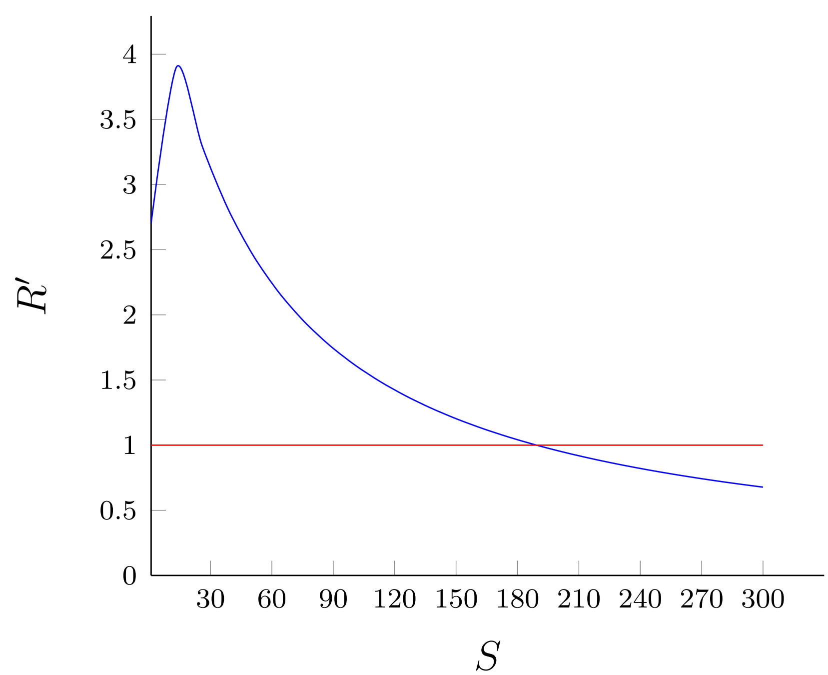

Figure 6: Sensitivity Analysis for Uniform Times with Bottleneck at 10th Station

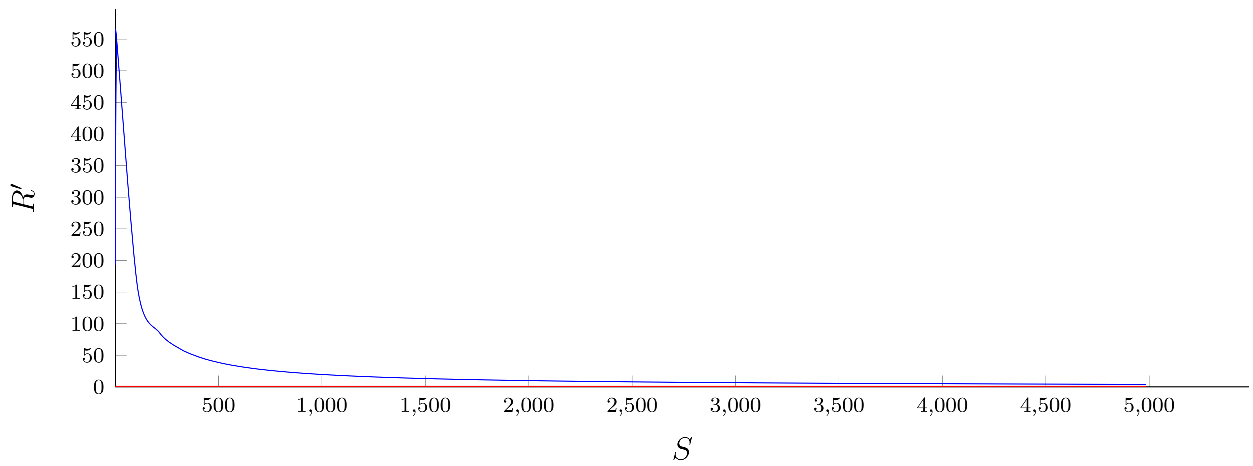

Figure 7: Sensitivity Analysis for Uniform Times with Bottleneck at 10th Station

supply and demand. Then the model is set as follows:

$$\text{Profit} = -\text{Cost} \times \text{Supply} + \text{Price} \times \min(\text{Supply}, \text{Demand}).$$

As a rule of thumb, the selling price is significantly higher than the unit cost. This increases the weight on the latter part of the simulation, which is related to the demand. In this way, the latter part of the simulation, which is being replicated, has both higher variability and significance. More specifically, we set the cost to 1 and the selling price to 10. The effort of generating one demand and one supply are assumed to be equal, and all other computational efforts are ignored. Therefore, f is determined by the effort required for demand and supply.

In the following experiments, demand and supply are assumed to be uniformly distributed with the parameters shown in the first two rows of Tables 7 and 8. In particular, D(a, b) (S(a, b)) denotes that the demand (supply) is uniformly distributed with the lower and upper bounds of a and b. Correlations, and thus optimal number of clones and efficiency, depend on the relative magnitude of supply and demand, overlap of supply and demand, and the variability of supply and demand. Thus the parameter values for the supply and demand distribution are chosen as follows:

- In Table 7, different ranges for demand and supply are considered. The demand

and supply ranges overlap, and the variability of supply and demand is constant. However, as the averages shift, the range of the overlapping area also changes. As demand and supply are both determined by one uniform random variable, $f = 0.5$ is assumed.

- In Table 8, we consider the case where it takes more computational effort to forecast supply. In this table, there is one source of demand, while the number of sources of supply is ten in the first two columns and one hundred in columns 3 and 4. Therefore, the first two columns have $f = 10/11$ and the last two columns have $f = 100/101$. Also, the averages of both demand and total supply remain the same, while the variability of the supply varies. The number in front of "S" denotes the number of suppliers.

- All the experiments are run for 2×10^6 replications and 200 macro replications in Table 7. This guarantees at least one significant digit in all correlation values. (to be edited for Table 8

As in Section 3.5.1, in Tables 7 and 8 we provide the estimated correlation values $\hat{\rho}_{cn}$, $\hat{\rho}_{ll}$, and $\hat{\rho}_{ln}$. The optimal numbers of branches, $\hat{S}^*$, and the efficiency ratios, $\hat{R}(\hat{S}^*)$, are then estimated using equations (3) and (4) for basic cloning, equations (2) and (6) with $S = 1$ for ARN, and equations (7) and (8) for cloning using ARN.

Table 7 shows that the efficiency gain is more significant when the mean demand is smaller and when the overlap is small. When both the mean demand is smaller and the overlap is small, then the sales are more dependent on the demand, and hence the effect of the supply, which is the shared part, becomes smaller. This is consistent with the previous observation that cloning is more efficient when the shared part has less impact on the performance measure of interest. Moreover, the effect of the overlap seems to be greater than the effect of the magnitude of the mean demand. This is because when the overlap is small, due to the structure of the profit function,

Table 7: Profit Model with Different Means

		D(0,10) S(0,10)	D(0,10) S(5,15)	D(0,10) S(9,19)	D(5,15) S(0,10)	D(9,19) S(0,10)
Correlations	$\hat{\rho}_{cn}$	0.3064	0.0056	0.0098	0.9374	0.9999
	$\hat{\rho}_{ll}$	-0.4219	-0.9395	-0.9999	-0.9254	-0.9999
	$\hat{\rho}_{ln}$	-0.2486	0.0040	-0.0098	-0.9254	-0.9999
Basic Cloning	$\hat{S}^*$	2	13	10	1	1
	$\hat{R}(\hat{S}^*)$	1.0206	1.7395	1.6707	1.0000	1.0000
ARN Only	$\hat{R}'(1)$	1.7299	16.5298	10145.5004	13.4103	8136.3442
Cloning with ARN	$\hat{S}^*$	3	2	4	2	4
	$\hat{R}'(\hat{S}^*)$	2.1624	19.0001	13640.1278	15.4164	11225.4346

Table 8: Profit Model with Different Variability

		D(0,10) 10S(0,1)	D(0,10) 10S(0.4,0.6)	D(0,10) 100S(0,0.1)	D(0,10) 100S(0.04,0.06)
Correlations	$\hat{\rho}_{cn}$	0.0481	0.0021	0.0052	0.000205
	$\hat{\rho}_{ll}$	-0.5611	-0.5983	-0.5957	-0.5998
	$\hat{\rho}_{ln}$	-0.0459	-0.0021	-0.0051	-0.000204
Basic Cloning	$\hat{S}^*$	14	69	139	699
	$\hat{R}(\hat{S}^*)$	3.9464	8.4028	34.3150	77.3054
ARN Only	$\hat{R}'(1)$	2.2783	2.4891	2.4733	2.4989
Cloning with ARN	$\hat{S}^*$	44	382	819	5669
	$\hat{R}'(\hat{S}^*)$	16.6739	26.0038	198.4510	243.7143

$\rho_{cn} + \rho_{ln} \simeq 0$ and $\rho_{ll} \simeq -1$. As can be seen in Equation (8), this has large impact on the efficiency ratio.

Table 8 shows the effects of both computational effort and variability. Comparing Table 8 with the first column of Table 7, the impact of f, the ratio of computational effort of shared part and cloned part, is obvious. Also, when the variability decreases in the shared part, then the efficiency ratio increases. However, as the mean supply is constant and equal to the mean demand, the gain is not as dramatic as in the third (fifth) column of Table 7, where supply (demand) has little effect on sales. Again, this supports the claim from the previous section that the variability and significance of the shared part of cloning significantly impact the effectiveness of cloning.

We conclude this section by discussing the estimated correlation values in Tables 3 through 8. In previous sections, we showed that $\rho_{cn} \geq 0$ and $\rho_{cn} \geq -\rho_{ln}$ and our numerical results confirm this. Moreover, we assumed that $0 > \rho_{ln} > \rho_{ll}$. The results in Tables 3 through 8 generally support this assumption. However, in the second column of Table 7 (with D(0,10) and S(5,15)), the negative correlation in the shared random variable (i.e., supply) exceptionally results in a positive correlation ρ_{ln}. This may explain why the efficiency ratio is not as large as might be expected given that sales is more dependent on demand in this case.

3.6 Conclusion

In this chapter, we have discussed the use of cloning in transient simulation. While cloning was originally designed to share some simulation results among sample paths for different scenarios, our approach shares simulation results among different replications of the same system. First, we describe our algorithm and identify the number of clones that optimizes its efficiency. Then, to offset the undesired positive correlation induced by sharing some results, we introduce cloning algorithms with induced negative correlation, and recommend inducing negative correlation both before and after the decision point for best results. Also, we have shown that the algorithms can be extended to the case with multiple decision points. Finally, simulation results are provided to illustrate the performance of our algorithms. Efficiency improvement ranges from 60 percent to approximately 1.36 million percent depending on the structure and performance measure of interest. Our cloning approach is easy to implement and is especially effective when it involves sharing some simulation results that require a substantial amount of computational effort but have little impact on the performance measure among multiple replications.

The cloning technique considered in this chapter is similar to the use of splitting as a rare-event simulation technique. It may be desirable to incorporate our techniques to induce negative correlation into splitting, with the improvement achieved at low cost. In such cases, inducing negative correlation only after the decision point may be worthwhile(as it involves smaller modifications to the original splitting approach), and cloning at multiple decision points corresponds to multilevel splitting.

CHAPTER IV

PARETO SET ESTIMATION USING RANKING AND SELECTION

4.1 Introduction

In this chapter, we consider the problem of selecting the best system when there are multiple objectives, and present an algorithm to estimate a Pareto set using the R&S approach. We provide three different formulations that may serve for different purposes. In each case, we identify what parameter values to choose and prove that the proposed algorithm with the suggested parameters guarantees the desired probability of correct selection. Our work differs from the earlier works in that we can compare multiple objectives, without prioritizing some objectives over others, and provide a guaranteed probability of correct selection.

The outline of this chapter follows. Section 4.2 provides the problem formulations and notation used throughout the chapter. In Section 4.3 we describe our procedure for selecting a Pareto set with a certain probability of correct selection when there are multiple systems and multiple objectives. Section 4.4 considers the three problem formulations that were introduced in Section 4.2 and provides the choices of parameters and validity proofs (that guarantee the desired probability of correct selection). In Section 4.5, we present results from numerical experiments. Finally in Section 4.6, we finish the chapter with a brief summary and conclusion.

4.2 Problem Formulation

Consider the ranking and selection problem with multiple maximization objectives. Instead of choosing a single best system, the goal of the proposed algorithm is to select a set of Pareto optimal systems. A system is Pareto optimal if there exists no other system that can improve upon it in one objective without hurting some other objectives.

Let the set of indices for the systems be $S = \{1, 2, \ldots, k\}$, where $2 \leq k < \infty$. Let X_{imn} be a real-valued observation associated with the m-th objective from replication n of system i. The performance measures are defined as $x_{im} = \mathbb{E}[X_{imn}]$. Also, let $\bar{X}_{im}(n)$ denote the estimate of the mean x_{im} calculated from the average of n independent observations. Let there be $\ell \geq 2$ objectives, and let $H = \{1, 2, \ldots, \ell\}$ be a set of indices of objectives. System i dominates system j, denoted as $j \prec i$, when $\forall m \in H$, $x_{im} \geq x_{jm}$ and $\exists m \in H$, such that $x_{im} > x_{jm}$. Let P denote the Pareto set. System $j \in P$ if $\nexists i$ such that $j \prec i$. Thus $P \neq \emptyset$ is a set of non-dominated systems.

We now define the Pareto set P_{IZ} with indifference zone (IZ). The indifference zone is the smallest actual amount that makes a practical difference to an experimenter. When the difference of two values is less than the indifference zone, then the difference is insignificant. Let $\delta_m > 0$ be the indifference zone for the m-th objective. For objective m, system i is significantly better than j, denoted as $x_{im} >_{IZ} x_{jm}$, if $x_{im} - x_{jm} \geq \delta_m$; systems i and j are indifferent, denoted as $x_{im} =_{IZ} x_{jm}$, if $|x_{im} - x_{jm}| < \delta_m$; and $x_{im} \geq_{IZ} x_{jm}$ holds if $x_{im} - x_{jm} > -\delta_m$. System i dominates system j with IZ, denoted as $j \prec_{IZ} i$, if $x_{im} \geq_{IZ} x_{jm}$, $\forall m$, and $\exists m$ such that $x_{im} >_{IZ} x_{jm}$. System $j \in P_{IZ}$ if $\nexists i$ such that $j \prec_{IZ} i$.

From their definitions, it is clear that P and P_{IZ} can yield completely different sets. Systems that are included in P can be excluded from P_{IZ}. Also, systems that are not included in P are not necessarily excluded from P_{IZ}. More specifically, consider the case with two systems i and j, where $i \neq j$, and let $H_j^i \subset H$ be a subset of objectives such that $\forall m \in H_j^i$, $x_{im} \geq x_{jm}$ holds. Note that $H_j^i \cup H_i^j = H$. If $H_j^i \neq \emptyset$ and $H_i^j \neq \emptyset$, $x_{jm} < x_{im} < x_{jm} + \delta_m$, $\forall m \in H_j^i$, and $x_{im} <_{IZ} x_{jm}$ $\forall m \in H_i^j$, then a system $i \in P$ and $i \notin P_{IZ}$. Conversely, $i \in P_{IZ}$, but $i \notin P$, if $H_i^j = H$, $x_{im} \leq x_{jm} < x_{im} + \delta_m$, $\forall m$, and $x_{im} \neq x_{jm}$ for at least one objective. This is the case when i is dominated by j, but the difference in all performance measures is less than the indifference zone. For example, consider the case with two objectives, three systems, and $\delta_m = 1$ for $m = 1, 2$. Let $x_{11} = 1, x_{12} = 2.2, x_{21} = 2, x_{22} = 2, x_{31} = 1.5$, and $x_{32} = 1.5$. Then, $P = \{1, 2\}$ (because $3 \prec 2$), while $P_{IZ} = \{2, 3\}$ (because $1 \prec_{IZ} 2$). However, note that $P_{IZ} = P$ as δ_m approaches 0 for all m.

Define sets S_D, S_U, and S_A, as the sets of desirable, undesirable, and acceptable systems, respectively. Then

$$S_D = \{j : \forall i \neq j, \exists m \text{ s.t. } x_{jm} >_{IZ} x_{im}\},$$
$$S_U = \{j : \exists i \neq j, \text{ s.t. } \forall m, x_{jm} <_{IZ} x_{im}\},$$
$$S_A = S \setminus (S_D \cup S_U).$$

That is, S_D is the set with systems that are significantly better than other systems in at least one objective, and S_U is the set with systems that are significantly dominated by other systems in all objectives. Therefore, S_A is the set with systems that are in the indifference zone. Note that $S_D \subseteq P \subseteq S_D \cup S_A$ and $S_D \subseteq P_{IZ} \subseteq S_D \cup S_A$. Also, while it is possible that $S_D = \emptyset$ (as in traditional R&S with one performance measure), it is always the case that $S_D \cup S_A \neq \emptyset$. This follows from the fact that $S_U \neq S$, as not all systems can be strictly dominated by other systems in all objectives.

Let $\boldsymbol{X}_{mn} = (X_{1mn}, X_{2mn}, \ldots, X_{kmn})$ be a vector across systems of the n-th observations of the m-th objective. We will need the following assumption throughout the chapter.

Assumption 1. *The random vectors $\boldsymbol{X}_{m1}$, $\boldsymbol{X}_{m2}$, ... are identically distributed multivariate normal with mean vector $(x_{1m}, x_{2m}, \ldots, x_{km})$, variance vector $(\sigma^2_{1m}, \sigma^2_{2m}, \ldots, \sigma^2_{km})$, and positive definite covariance matrix $\sum_m$, where x_{im} and σ^2_{im} are unknown, $\forall i \in S, m \in H$.*

Assumption 1 is similar to normality assumptions that are conventionally used in the ranking and selection literature. It is not very restrictive as the performance measures can be estimated via means of batches of observations, implying that Assumption 1 usually holds in an asymptotic sense.

For all $j \in S_U$, let $i_j \in S_D \cup S_A$ be a system such that $x_{i_j m} >_{IZ} x_{jm}$, $\forall m$. (Note that the implicit assumption that $i_j \in S_D \cup S_A$ holds without loss of generality.) Several of our results will require the following assumption.

Assumption 2. *For all $j \in S_U$, $i_j \in S_D$.*

Without Assumption 2, there may be a case when $i_j \in S_A$ is eliminated before i_j eliminates j. This will leave j without guaranteed elimination. In the ranking and selection literature, it is common to assume that no performance measures fall within the indifference zone. Under such an assumption, clearly $S_A = \emptyset$ holds. If $S_A = \emptyset$, then a system is either in S_U or S_D. If a system j is in S_U, there exists a system that is significantly dominant in all objectives. If such system, say q, is not in S_D, then there exists a system q' that is significantly dominant in all objectives compared to q, and thus to j. If $q' \notin S_D$, the same argument continues, and must end with a system in S_D that is significantly dominant in all objectives to system j, because the number of systems k is finite. Therefore, Assumption 2 is less restrictive than the traditional assumption that $S_A = \emptyset$.

We consider three different formulations for the correct selection of a Pareto set. Let $\hat{P}$ denote the estimate of the Pareto set selected using our procedure. The subscript of the set $\hat{P}$ indicates what formulation is used to select the set. In the first formulation, denoted as CS_{ER}, the Correct Selection is to Eliminate all undesirable systems and Retain all desirable systems. Namely, it is the event, $S_D \subseteq \hat{P}_{ER} \subseteq S_D \cup S_A$. Each of the next two formulations relieves one constraint on $\hat{P}$. First, by removing the constraint of eliminating all undesirable systems, we have the second formulation, denoted as CS_{RD}, in which the Correct Selection is to Retain all the systems that are Desirable. This resembles subset selection in the ranking and selection literature in a sense that the selected set includes a subset whose members are dominant compared to those not included [21]. This event can be expressed as $S_D \subseteq \hat{P}_{RD}$. Second, CS_{EU} is the Correct Selection event where all Undesirable systems are Eliminated. It can be expressed as $\emptyset \neq \hat{P}_{EU} \subseteq S_D \cup S_A$. As the original formulation both retains and eliminates, $CS_{ER} = CS_{EU} \cap CS_{RD}$ holds.

4.3 Pareto Set Selection Procedure (PSSP)

Objective: Identify systems that are Pareto optimal when there are k systems with ℓ objectives, x_{im}, $\forall i \in S, m \in H$. Correct selection can follow any of the three formulations discussed in Section 4.2 that best fits the needs of users.

Setup: Select the overall confidence level $1-\alpha$, indifference zone δ_m, $\forall m$, and first-stage sample size $n_0 \geq 2$. Let 1_A denote the indicator function. That is, if the event A is true, then $1_A = 1$, and else, $1_A = 0$. Pick the constant c and calculate η to be a solution to

$$g(\eta) \equiv \sum_{l=1}^{c}(-1)^{(l+1)}\left[1-\frac{1}{2}1_{\{l=c\}}\right]\left[1+\frac{2\eta(2c-l)l}{c}\right]^{-(n_0-1)/2} = \beta. \tag{12}$$

In general, $c = 1$ is a good choice. (For further reference on selecting the constant c, refer to Section 3.1 of Kim and Nelson [63].) When this is the case, η can be obtained from the closed-form expression $\eta = \frac{1}{2}[(2\beta)^{\frac{2}{1-n_0}} - 1]$. The selection of β depends on the confidence level α and formulation (ER, RD, or EU). Refer to Theorems 4.4.1 through 4.4.3 for further details.

Initialization: Let $M = \{1, 2, \ldots, k\}$ be the set of systems that have not been found to be dominated by other systems and $SS_{im} = \emptyset$ be the set of superior systems to system i in terms of the m-th objective. Let $\hat{P} = \emptyset$ denote the estimate of the Pareto optimal set, $C = \emptyset$ denote a set of systems that have "completed" all needed comparisons, and let $h^2 = 2c\eta \times (n_0 - 1)$. Obtain n_0 observations X_{imn},

$n = 1, 2, \ldots, n_0$, $\forall m \in H$ and $\forall i \in S$. For all $i \neq j$ and $\forall m$, compute

$$S^2_{(ij)^m} = \frac{1}{n_0 - 1} \sum_{n=1}^{n_0} [X_{imn} - X_{jmn} - (\overline{X}_{im}(n_0) - \overline{X}_{jm}(n_0))]^2,$$

the sample variance of the difference between systems i and j in objective m. Set the observation counter $r = n_0$.

Comparison and Stopping: For each pair of systems $i, j \in M \setminus C$ and each $m \in H$ such that $i \neq j$, $j \notin SS_{im}$, $i \notin SS_{jm}$, compare

$$\sum_{n=1}^{r} X_{imn} < \sum_{n=1}^{r} X_{jmn} - R_{(ij)^m}(r), \tag{13}$$

where $R_{(ij)^m}(r) = \max\{0, \frac{1}{2c}(\frac{h^2 S^2_{(ij)^m}}{\delta_m} - \delta_m r)\}$. If (13) holds, add j to SS_{im}. If $j \in SS_{im'}$ for all $m' \neq m$, then eliminate i from all existing sets. If $|M| = 1$, where $|M|$ denotes the number of systems in the set M, then add j to $\hat{P}$, stop, and return $\hat{P}$ as the Pareto optimal set. Otherwise, check if for all $q \in M \setminus \{j\}$, there exists m_q such that $j \in SS_{qm_q}$. If so, then add j to $\hat{P}$. If $\hat{P} = M$, then stop and return $\hat{P}$ as the Pareto optimal set. If $j \in \hat{P}$ and $\hat{P} \neq M$, check if $j \in SS_{qm}$ or $q \in SS_{jm}$ holds, $\forall m \in H$ and $\forall q \in M \setminus \hat{P}$. If so, add j to C. If $|M \setminus C| = 1$, then the remaining system has completed comparisons with all other systems in M, and hence cannot be eliminated by any system. Thus add the remaining system in $M \setminus C$ to $\hat{P}$, then stop and return $\hat{P}$ as the Pareto optimal set. See Figure 8 for the flowchart of this step.

Data Collection: If the procedure was not stopped in the **Comparison and Stopping** step, take one additional observation $X_{im(r+1)}$ for $m \in H$ from the remaining systems in $M \setminus C$, and go to **Comparison and Stopping**.

Note that without introducing the set $C \subseteq \hat{P}$, the PSSP algorithm cannot stop sampling from systems placed in $\hat{P}$ until $\hat{P} = M$, even if comparisons with all systems in $M \setminus \hat{P}$ have been completed. Also, if we let $\lfloor x \rfloor$ denote the largest integer smaller

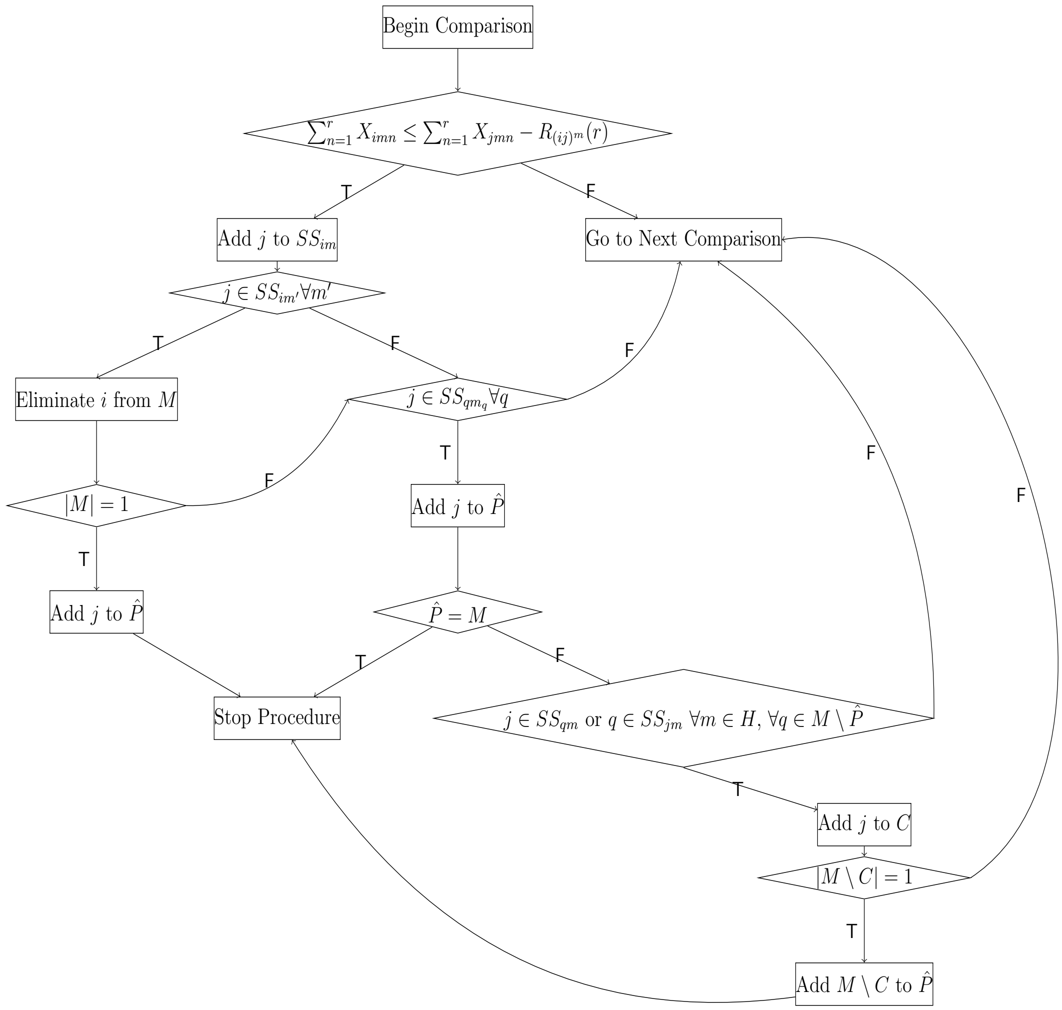

Figure 8: Flowchart for Comparison and Stopping Step in PSSP

than x and

$$N_{(ij)^m} = \left\lfloor \frac{h^2 S^2_{(ij)^m}}{\delta^2_m} \right\rfloor,$$

$$N_i = \max_{j \neq i} \max_{m \in H} N_{(ij)^m},$$

then $N_i + 1$ is the maximum number of observation that is required to make correct decision on system i. Therefore, if $n_0 \geq N_i + 1$, $\forall i$, then the procedure will stop without additional data collection after one execution of the **Comparison and Stopping** step with $\hat{P}$ being the estimated Pareto optimal set.

4.4 Parameter Choices and Validity Proofs

In Section 4.3, we deferred the specification of the parameter β used in Equation (12). To achieve the desired probability of correct selection $1-\alpha$, β must be set differently for each formulation CS_{ER}, CS_{RD}, and CS_{EU}. In the following Sections 4.4.1, 4.4.2, and 4.4.3, we describe the choices of the parameter β and associated validity proofs for the ER, RD, and EU formulations, respectively. Finally in Section 4.4.4 we discuss the expected performance of the formulations.

We begin with stating the following lemma, adapted from Kim and Nelson [63], which is required to prove the validity of PSSP under all formulations. For all systems i, j and objectives m with $x_{im} - x_{jm} \geq \delta_m$, let $CS_{(ij)^m}$ be the event that the m-th objective of system j is found inferior to that of system i. We will need the following lemma:

Lemma 4.4.1. *Under Assumption 1, PSSP satisfies* $\mathbb{P}\{CS_{(ij)^m}\} \geq 1-\beta$, $\forall i, j, m$ *that satisfy* $x_{im} - x_{jm} \geq \delta_m$.

Proof. This follows directly from the proof of Theorem 1 of Kim and Nelson [63], as $P\{ICS\} \leq \beta$, where ICS denotes the probability of incorrect selection made at a given time when comparing two systems in one objective. □

4.4.1 Eliminating the Undesirable, Retaining the Desirable (ER)

In this section, for the correct selection, CS_{ER}, we guarantee that all the systems that are desirable are retained in the final Pareto set, and that all the undesirable systems are eliminated.

Theorem 4.4.1. *Under Assumptions 1 and 2, the PSSP procedure guarantees that*

$$\mathbb{P}\{CS_{ER}\} \geq 1 - \max(k,\ell)(k-1)\beta.$$

Thus to achieve the desired confidence level, β should be set as $\beta = \frac{\alpha}{\max(k,\ell)(k-1)}$.

Proof. Let CS_j be the event that $j \in S_U$ is eliminated and GS_i be the event that $i \in S_D$ is not eliminated. Also, let CS_{ij} be the event that i would eventually eliminate $j \in S_U$. Note that $CS_{i_j j} = \bigcap_{m \in H} CS_{(i_j j)^m}$ as for $j \in S_U$ to be eliminated by i_j, i_j has to be declared better in all objectives. Similarly, for $i \in S_D$ and $j \in S \setminus \{i\}$, let GS_{ij} be the event that j does not eliminate i, thus $GS_i = \bigcap_{j \neq i} GS_{ij}$. That is, for system i to survive, no other system should eliminate i. Then it follows that

$$
\begin{aligned}
\mathbb{P}\{CS_{ER}\} &= \mathbb{P}\left\{ (\bigcap_{j \in S_U} CS_j) \cap (\bigcap_{i \in S_D} GS_i) \right\} \\
&\geq \mathbb{P}\left\{ (\bigcap_{j \in S_U} CS_{i_j j}) \cap (\bigcap_{i \in S_D} \bigcap_{j \neq i} GS_{ij}) \right\} \\
&= \mathbb{P}\left\{ (\bigcap_{j \in S_U} \bigcap_{m \in H} CS_{(i_j j)^m}) \cap [\bigcap_{i \in S_D} (\bigcap_{\jmath \in S_U:\ i_\jmath \neq i} GS_{i\jmath} \cap \bigcap_{\jmath \in S_D \cup S_A \setminus \{i\}} GS_{i\jmath})] \right\} \qquad (14) \\
&\geq \mathbb{P}\left\{ \bigcap_{j \in S_U} \bigcap_{m \in H} CS_{(i_j j)^m} \right\} + \mathbb{P}\left\{ [\bigcap_{i \in S_D} (\bigcap_{\jmath \in S_U:\ i_\jmath \neq i} GS_{i\jmath} \cap \bigcap_{\jmath \in S_D \cup S_A \setminus \{i\}} GS_{i\jmath})] \right\} - 1. \qquad (15)
\end{aligned}
$$

The first equality follows from the definition of the ER formulation. The first inequality holds as $CS_{i_j j}$ is the event that i_j eliminates j and $GS_{i_\jmath}$ is the event that system $\jmath$ does not eliminate i. The second equality follows because for a system to be eliminated, the system has to be inferior in all objectives compared to another system, and system $i_\jmath$ cannot simultaneously eliminate system $\jmath$ and also be eliminated by system $\jmath$ (so that $CS_{i_\jmath \jmath} \subseteq GS_{i_\jmath \jmath}$).

Let $u(d)$ denote the number of systems in $S_U(S_D)$. Note that $d = 0$ implies that $u = 0$ by Assumption 2. In this case, any set returned by the procedure satisfies the condition of correct selection, thus the case is trivial. Otherwise, note that if system i is in S_D, and thus should not be eliminated by any $j \neq i$, then it is guaranteed that i is better than j in at least one objective, say m_{ij}, and it suffices to correctly compare the two systems in that objective. Therefore, from Assumption 1 and Lemma 4.4.1,

$\mathbb{P}\{GS_{ij}\} \geq \mathbb{P}\{CS_{(ij)^{m_{ij}}}\} \geq 1-\beta$. Similarly, $\mathbb{P}\{CS_{(i_j j)^m}\} \geq 1-\beta, \forall m \in H$, whenever $j \in S_U$. Therefore, Equation (15) yields:

$$\begin{aligned}\mathbb{P}\{CS_{ER}\} &\geq (1-\ell u\beta) + [1-(d(k-1)-u)\beta] - 1 \\ &= 1-[u(\ell-1)+d(k-1)]\beta \\ &\geq 1-[d(k-\ell)+k(\ell-1)]\beta. \end{aligned} \tag{16}$$

The first inequality holds by the Bonferroni inequality and Lemma 4.4.1. The second inequality holds as $u \leq k-d$ holds. This implies that the worst-case scenario will be when $u+d=k$.

Finally, if the number of systems exceeds the number of objectives, i.e., $k > \ell$, then the worst case is when all the systems are in the Pareto set, namely $d = k$, and the lower bound (16) is $1-k(k-1)\beta$. (Note that if $k=\ell$, then the worst case does not depend on d or u as long as $d+u=k$.) If the opposite is true, then the worst case happens when $d=1$, so that only one system dominates all other systems, and the lower bound (16) becomes $1-\ell(k-1)\beta$. Therefore, the lower bound (16) is $1-\max(k,\ell)(k-1)\beta$, and the proof is complete. □

Intuitively, eliminating a system in S_U requires ℓ correct comparisons, whereas selecting a system in S_D requires $k-1$ correct comparisons. Therefore, when the number of objectives, ℓ, is small, eliminating a system is easier than confirming one to be Pareto optimal. When the number ℓ is large, a system must be dominated in every objective to be eliminated, and thus it becomes relatively easier to confirm a system to be Pareto optimal.

As discussed in Section 4.2, Assumption 2, is less restrictive than the commonly used assumption that $S_A = \emptyset$. This yields the following corollary. Note that the second inequality in Equation (16) should be an equality in this case, which will not affect the final result.

Corollary 4.4.1. *The result of Theorem 4.4.1 holds under Assumption 1 and the assumption that* $S_A = \emptyset$*.*

4.4.2 Retaining the Desirable (RD)

In this subsection, we analyze the Retain the Desirable formulation for correct selection. This will guarantee that the selected set will retain S_D in the final result with the pre-specified probability of correct selection.

Theorem 4.4.2. *Under Assumption 1, the PSSP procedure guarantees that*

$$\mathbb{P}\{CS_{RD}\} \geq 1 - k(k-1)\beta.$$

Thus to achieve the desired confidence level, β *should be set as* $\beta = \frac{\alpha}{k(k-1)}$.

Proof. Correct selection under the RD formulation implies that any system in S_D is not eliminated. It follows that

$$\begin{aligned}
\mathbb{P}\left\{CS_{RD}\right\} &= \mathbb{P}\left\{\bigcap_{i \in S_D} GS_i\right\} \\
&= \mathbb{P}\left\{\bigcap_{i \in S_D}\bigcap_{j \neq i} GS_{ij}\right\} \\
&\geq 1 - d(k-1)\beta \\
&\geq 1 - k(k-1)\beta,
\end{aligned}$$

where the second equality follows from the definition of GS_i (as in the proof of Theorem 4.4.1), the first inequality follows from the Bonferroni inequality and Lemma 4.4.1, and the last inequality follows from the fact that $d \leq k$. □

Note that under this formulation, we only need Assumption 1. Also, while the lower bound under this formulation is equal to that under the ER formulation when the number of objectives ℓ is smaller than the number of systems k, when $k < \ell$, the RD formulation performs better than the ER formulation, and the improvement equals $(\ell - k)(k - 1)\beta$. As mentioned in Section 4.4.1, selecting a system in S_D requires $k - 1$ correct comparisons, thus the lower bound of the RD formulation does not depend on ℓ. Therefore, with a fixed number of systems, increasing the number of objectives does not affect the performance of the RD formulation. However, in the ER formulation, to correctly eliminate the systems in S_U, ℓ objectives should be considered for the worst case scenario, and that is where the discrepancy comes from. This explains why RD does better than ER when $k < \ell$.

4.4.3 Eliminating the Undesirable (EU)

In this section, we define correct selection to be the event where all the undesirable systems are eliminated with the specified probability of correct selection.

Theorem 4.4.3. *Under Assumptions 1 and 2, PSSP guarantees that*

$$\mathbb{P}\{CS_{EU}\} \geq 1 - \max(k, \ell)\left[k - 1 - \frac{1}{2}(k - \ell)1_{\{k>\ell\}}\right]\beta.$$

Thus, to achieve the desired confidence level, β should be set as $\beta = \frac{\alpha}{\max(k,\ell)\left[k-1-\frac{1}{2}(k-\ell)1_{\{k>\ell\}}\right]}$.

Proof. To guarantee that all the inferior systems $j \in S_U$ are eliminated, we also need to guarantee that the systems $i_j \in S_D$ are not eliminated. Therefore, our bound of the probability of correct selection under the EU formulation is

$$\mathbb{P}\{CS_{EU}\} \geq \mathbb{P}\left\{\bigcap_{j \in S_U} CS_{i_j j} \cap GS_{i_j}\right\}. \tag{17}$$

If $u \geq d$, then the systems in S_D that have to be retained may be, in the worst case scenario, all the systems in S_D. In such case, the probability of correct selection follows:

$$\begin{aligned}\mathbb{P}\{CS_{EU}\} &\geq 1 - [d(k-\ell) + k(\ell - 1)]\beta \\ &\geq 1 - \max(k, \ell)\left[k - 1 - \frac{1}{2}(k - \ell)1_{\{k>\ell\}}\right]\beta,\end{aligned} \tag{18}$$

where the first inequality comes from Equation (16). For the second inequality, we should consider two cases. If $k < \ell$, the same argument as in the proof of Theorem

4.4.1 holds. If $k > \ell$, now that we also have $u \geq d$, the worst case scenario is when $d = \frac{1}{2}k$. Again, when $k = \ell$, the worst case does not depend on the relative size of u and d, but is when $u + d = k$ holds.

When $u < d$, then it follows from Equation (17) that:

$$\begin{aligned}
\mathbb{P}\{CS_{EU}\} &\geq \mathbb{P}\left\{\bigcap_{j\in S_U}(\bigcap_{m\in H} CS_{(i_j j)^m}) \cap (\bigcap_{\jmath\neq j,\ \jmath\neq i_j} GS_{i_j \jmath})\right\} \\
&\geq \mathbb{P}\left\{\bigcap_{j\in S_U}\bigcap_{m\in H} CS_{(i_j j)^m}\right\} + \mathbb{P}\left\{\bigcap_{j\in S_U}\bigcap_{\jmath\neq j,\ \jmath\neq i_j} GS_{i_j \jmath}\right\} - 1 \\
&\geq (1 - \ell u \beta) + [1 - u(k-2)\beta] - 1 \\
&= 1 - (\ell + k - 2)u\beta \\
&> 1 - \frac{k}{2}(\ell + k - 2)\beta, \qquad (19)
\end{aligned}$$

where the third inequality follows from the Bonferroni inequality and Lemma 4.4.1, and the last inequality follows from $u < d$ and $\ell + k > 2$.

Finally, when $k \geq \ell$, then the lower bounds (18) and (19) are equal. When $k < \ell$, comparing Equations (18) and (19) and using $k \geq 2$, the worst case occurs when $u \geq d$. Therefore, the final lower bound for the probability of correct selection is given as in Equation (18). □

It follows that the lower bound of $\mathbb{P}\{CS_{EU}\}$ is greater than that of $\mathbb{P}\{CS_{ER}\}$ when the number of systems k is larger than the number of objectives ℓ. As this formulation relieved the constraint that we should retain all the desirable from the original (ER) formulation, and as intuitively eliminating a system in S_U is easier than retaining a system in S_D when $\ell < k - 1$ (see Section 4.4.1), this is not surprising. As the difference can be expressed as $\frac{k}{2}(k - \ell)\beta$, the improvement gets more significant as the number of systems becomes relatively greater than the number of objectives.

4.4.4 Discussion

As we can see from Sections 4.4.1, 4.4.2, and 4.4.3, the lower bounds on the probability of correct selection in all formulations depend on the relative magnitude of the number of systems k and number of objectives ℓ. In Table 4.4.4, the lower bounds of all formulations when $k \leq \ell$ and $k \geq \ell$ is displayed. Note that the lower bounds are all $1 - k(k-1)$ when $k = \ell$.

Table 9: Lower Bounds on the Probability of Correct Selection

	$k \leq \ell$	$k \geq \ell$
ER	$1 - \ell(k-1)\beta$	$1 - k(k-1)\beta$
RD	$1 - k(k-1)\beta$	$1 - k(k-1)\beta$
EU	$1 - \ell(k-1)\beta$	$1 - k[\frac{1}{2}(k+\ell) - 1]\beta$

If $k < \ell$, the RD formulation has a higher lower bound than both the ER and EU formulations, with the difference being $(\ell - k)(k-1)\beta$. When $k \geq \ell$ holds, the RD formulation has the same lower bound as ER, and the lower bound of the EU formulation is the greatest, with the difference being $\frac{k}{2}(k-\ell)\beta$. These conclusions are reasonable because no matter what the number of systems is, a system $j \in S_U$ is eliminated if it is correctly compared with one system, and that requires ℓ comparisons. By contrast, retaining a system in S_D requires $k-1$ comparisons (as it must be correctly compared to every other system in one objective). Thus eliminating undesirable systems appears to be easier than retaining desirable systems when $\ell < k$, and vice versa. However, if $|\ell - k|$ is kept the same and $k > 2$, the improvement of the RD formulation is greater than that that of the EU formulation. This is due to the fact that to be able to correctly eliminate all systems in S_U, we need to retain some systems in S_D (see Assumption 2). Finally, the savings achieved by relieving the appropriate constraint on the definition of correct selection (either retaining the desirable or eliminating the undesirable depending on the sign of $\ell - k$) grows as $|\ell - k|$ gets larger.

In summary, when the number of objectives ℓ is greater than the number of systems k, then we can expect faster completion of the PSSP procedure by using the RD formulation, and otherwise by using the EU formulation.

4.5 Numerical Results

In this section, we provide the results from numerical experiments under various configurations. Configuration parameters include not only mean, variance, and covariance information, which are used in traditional experiments for ranking and selection, but also the formulations described in Section 4.2, the number of systems k, constraints ℓ, desirable systems d, undesirable systems u, and Pareto optimal systems z. We focus on the case when $k = u + d$, as this matches the worst case scenarios in Theorems 4.4.1, 4.4.2, and 4.4.3. (Andradóttir and Kim [3] also state that the addition of acceptable systems does not significantly impact the performance or validity of their R&S procedures.) Note that this leads to $z = d$. In Section 4.5.1, we briefly describe the experiment configurations that we use. In Section 4.5.2, results that provide evidence of validity are presented. In Section 4.5.3, we provide results that document the performance of the procedure. As the ER formulation involves the most comprehensive estimation of the Pareto set, we concentrate on the results of the ER formulation in Section 4.5.2 and 4.5.3. We will look into other formulations in Section 4.5.4.

4.5.1 Experiment Configurations

In this section, we introduce the configurations that are used to generate the results. We also discuss the choice of parameters. The performance measures of the systems and associated distributions are in accordance with Assumptions 1 and 2. We first specify three mean configurations.

To begin with, we define *One Pareto (OP)* to be the configuration with only one system in the Pareto set that dominates all other systems exactly by δ_m, $\forall m$.

This resembles the Difficult Mean (DM) configuration of the single objective R&S approach as there is a single best system and the difference of the mean of superior and inferior systems equals their indifference zone, as in slippage conditions. Under the OP configuration, $\forall i \in S$ and $\forall m \in H$, consider the following:

$$x_{im} = \begin{cases} \delta_m, & \text{if } i = k, \\ 0, & \text{otherwise.} \end{cases}$$

Under the OP configuration, clearly, $d = z = 1$ and $u = k-1$, where $P = P_{IZ} = \{k\}$.

As the number of systems in the Pareto set also can be a parameter of interest, we next introduce a different configuration called *Many Pareto (MP)*. Under this configuration, we also have control over the number z of Pareto optimal systems, as long as $0 < z \leq \ell$. For all $i \in S$, $m \in H$, consider the following:

$$i \leq z, \quad x_{im} = \begin{cases} \delta_m, & \text{if } m = i, \\ 0, & \text{otherwise;} \end{cases}$$

$$i > z, \quad x_{im} = -\delta_m.$$

In this case, each system i is superior in one objective if $i \leq z$, and all objectives are dominated if $i > z$. As the differences of the performance measures are greater than or equal to the indifference zone, $P = P_{IZ} = \{1, \ldots, z\}$. As $S_A = \emptyset$, $d = k - u = z$ holds. Note that under the RD formulation, $z = \min(k, \ell)$ is the most difficult case (that requires the most correct comparisons), and under the EU formulation, $z = 1$ is such a case.

Finally, we introduce the *Parallel Plane (PP)* configuration. The name comes from the fact that the set of systems S forms parallel planes if plotted using their means. Let K, a, and b be natural numbers and $\delta_m = \delta$, $\forall m \in H$. The parameter a denotes the number of parallel planes, where the plane P_j is defined as a set of systems having $\sum_m x_{im} = [K - (j-1)\ell]\delta$, for $j = 1, 2, \ldots, a$, and b denotes the number of systems on the lowest plane (with $\sum_m x_{im} = [K-(a-1)\ell]\delta$). In addition, each plane P_j has bounds on the values of x_{im}, $\forall i \in S$, $m \in H$. These bounds are set to ensure that the Pareto optimal set is equal to the top plane P_1, $k = u + d$ holds, and the number of systems in each plane $P_1, \ldots, P_{a-1}$ other than the lowest plane P_a is constant. Therefore, the number of systems satisfies $k = (a-1)z + b$, and b should be chosen so that b is not greater than the number of systems z in each plane $P_1, \ldots, P_{a-1}$.

More precisely, let $\mathbb{Z}^+$ denote the set of non-negative integers. Then $S = P_1 \cup \cdots \cup P_a$, where each plane P_j, $j = 1, 2, \ldots, a-1$, satisfies:

$$P_j = \{(x_{i1}, x_{i2}, \ldots, x_{i\ell}) \mid \sum_m x_{im} = [K-(j-1)\ell]\delta,$$
$$(a-j)\delta \le x_{im} \le [K-(a-1)\ell + (a-j)]\delta, x_{im}/\delta \in \mathbb{Z}^+\},$$

and

$$P_a \subseteq P'_a = \{(x_{i1}, x_{i2}, \ldots, x_{i\ell}) | \sum_m x_{im} = [K-(a-1)\ell]\delta, 0 \le x_{im} \le [K-(a-1)\ell]\delta, x_{im}/\delta \in \mathbb{Z}^+\},$$

such that $|P_a| = b$, where we choose the systems in P_a in ascending order of $(x_{i1}, x_{i2}, \ldots, x_{i\ell})$ (i.e., $(x_{i_1 1}, \ldots, x_{i_1 \ell}) > (x_{i_2 1}, \ldots, x_{i_2 \ell})$ if $\exists m \in H$ such that $x_{i_1 m'} = x_{i_2 m'}$ for $m' < m$ and $x_{i_1 m} > x_{i_2 m}$). To understand the bounds on x_{im}, first consider P'_a. As we do not want the performance measures to be negative, the lower bound is 0. Then, as the performance measures of a system in P_a should satisfy $\sum_m x_{im} = [K-(a-1)\ell]\delta$, the means cannot exceed $[K-(a-1)\ell]\delta$. Then, as a system in P_j that dominates a system in P_{j+1} has the mean performances in each objective greater by δ under this configuration, the lower and upper bounds of the upper planes, $P_{a-1}, \ldots, P_1$ should increase by $\delta, \ldots, (a-1)\delta$, respectively.

With all the constraints and bounds on the mean performance measures, the Pareto optimal set P_1 contains the same number of systems as P'_a. Moreover, the number of systems in P'_a is equal to the number of possible combinations of ℓ non-negative integers that sum up to $K-(a-1)\ell$. This yields

$$z = |P_1| = |P'_a| = \binom{K-(a-1)\ell+(\ell-1)}{\ell-1} = \frac{[K-(a-1)\ell+(\ell-1)]!}{[K-(a-1)\ell]!(\ell-1)!}. \quad (20)$$

For ease of intuitive understanding, consider the case when $a=\ell=2$. Then, $z=d=K-1$, and $u=b$ can be chosen such that $0 \leq u \leq K-1$. For a realization of this configuration with $K=6$, $b=5$, and $k=10$, see Figure 9. In this example, $S=P_1 \cup P_2$, where

$$P_1 = \{(x_{i1}, x_{i2}) \mid \sum_m x_{im} = 6, 1 \leq x_{im} \leq 5, x_{im} \in \mathbb{Z}^+\}$$
$$= \{(1,5),(2,4),(3,3),(4,2),(5,1)\},$$

$$P_2 = P'_2 = \{(x_{i1}, x_{i2}) \mid \sum_m x_{im} = 4, 0 \leq x_{im} \leq 4, x_{im} \in \mathbb{Z}^+\}$$
$$= \{(0,4),(1,3),(2,2),(3,1),(4,0)\}.$$

We can eliminate $z-b$ systems from the right-most point in the lower plane to obtain a subset P_2 of size b. Finally, we can see that $z=d=5$, as can be calculated in Equation (20).

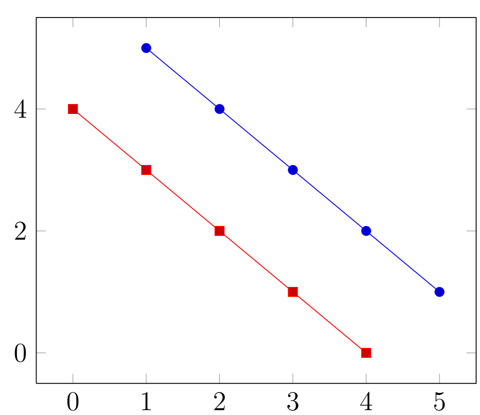

Figure 9: Example PP Configuration with $K=6$, $k=10$, $\ell=2$, $\delta=1$

Under the PP configuration, we can control the number of systems k, while keeping the number of objectives ℓ and the number z of systems in the Pareto set constant.

The OP and MP configurations are motivated by the DM configuration, as mentioned earlier. Thus, we will focus on these configurations when addressing the validity of the PSSP procedure. On the other hand, the PP configuration resembles the Monotone Increasing Mean (MIM) configuration in the R&S literature, and thus shows the efficiency of the procedure in realistic settings.

In addition to the aforementioned mean configurations, we can also change the value of the standard deviation. In most of our experiments, the standard deviation of all objectives equals 1. However, to observe the effects of high variation in some performance measures, we both consider the case where a number of systems have higher variability, and also the case where a number of objectives have higher variability compared to others. In both cases, higher variability corresponds to a standard deviation of 5.

In general, we expect that the performance measures for each system i are correlated. However, as was noted in Healey et al. [52] and references therein, studies have shown that the correlation between the performance measures does not significantly impact the results. Therefore, we will generate each performance measure in a system independently. For now, we also simulate each system independently. However, using Common Random Numbers (CRN) may have an impact on the performance. We again direct the interested reader's attention to [52] where the impact of CRN's among systems is measured.

Finally, there are a few parameters that are common in all configurations. For the constant c, we use $c = 1$, and for the initial number of observations, we use $n_0 = 20$. The confidence level α in all following experiments is set to be 5%, and the indifference zone $\delta_m = 1/\sqrt{n_0}$, $\forall m \in H$. Finally, the number of macro replications is set to ensure that all the half-widths of the probability of correct selection be less than 0.001 and that the half-widths of the total number of replications be less than 10. As the number of systems k and objectives ℓ differs in the experiments, those are presented with the results in following tables and graphs.

4.5.2 Validity

In this section, we present numerical results obtained using PSSP in the ER formulation, mainly focusing on providing supporting results for validity. In the following, PCS denotes the probability of correct selection, REP stands for the average number of observations per system before stopping the procedure, and COMP denotes the number of correct comparisons required to achieve correct selection, as derived in the proofs of Theorems 4.4.1, 4.4.2, and 4.4.3. In standard R&S, COMP is generally $k-1$, as it suffices to compare the best system k with all other systems correctly. However, in the OP configuration (see Table 10), for example, when $k = 2$ and $\ell = 3$, then for correct selection, comparison in each objective has to conclude that system 2 is superior, thus requiring 3 correct comparison. On the other hand, if $k = 3$ and $\ell = 2$, then system 3 has to be dominant in all objectives against systems 1 and 2, thus requiring 4 correct comparisons. Assuming $S_A = \emptyset$ (as in the mean configurations we are using), we can see that the number of correct comparison required for the ER formulation satisfies:

$$\text{COMP} = z(k-1) + (k-z)(\ell-1) = k(z-1) + (k-z)\ell. \tag{21}$$

This implies that the PSSP is likely to be more conservative than standard R&S for the same number of systems. While standard R&S is conservative only in the sense that the best system k is guaranteed to eliminate all other systems in the slippage condition, the fact that we do not know z before the simulation implies that we must also consider the worst case value of z, leading to additional discrepancy between what we can guarantee (in calculating the lower bounds) and what actually is required. Also, COMP increases a lot faster with the number of systems k than it does in standard R&S, and thus our experiments focus on smaller numbers of systems when compared to recent works on R&S.

In Table 10, simulation results under the OP configuration are shown. All PCSs are greater than 0.95, which is the confidence level, thus supporting the validity of the procedure. As we can see from the PCS in the upper right triangle, when $k \le \ell$, then the PCS does not depend significantly on the values of k and ℓ. This is due to the relationship between COMP and the actual lower bound we use to calculate the parameter β. In particular, when $k \le \ell$, then the lower bound in Theorem 4.4.1 is $1 - \ell(k-1)\beta$, meaning the error probabilities are allocated for $\ell(k-1)$ (= COMP) comparisons. On the other hand, when $k > \ell$, we can observe that the procedure gets less conservative as the number of objectives ℓ increases for a fixed number of systems k. This is because COMP increases as ℓ increases, but the lower bound, which is $1 - k(k-1)\beta$ does not change with ℓ. Finally, for a fixed ℓ, then the PCS increases as k increases. This is not surprising as when $k > \ell$, the lower bound for the ER formulation is $1 - k(k-1)\beta$, which increases faster with k than COMP, which is linear in k.

Table 10: Simulation Results under OP Configuration

		ℓ			
k		2	3	5	10
2	PCS	0.955	0.956	0.956	0.955
	REP	122.0	157.0	207.5	289.5
	COMP	2	3	5	10
3	PCS	0.971	0.957	0.956	0.956
	REP	181.3	204.0	263.7	352.0
	COMP	4	6	10	20
5	PCS	0.983	0.975	0.960	0.960
	REP	253.6	283.8	319.6	414.4
	COMP	8	12	20	40
10	PCS	0.992	0.987	0.979	0.959
	REP	341.1	378.3	422.1	478.1
	COMP	18	27	45	90

In addition to k and ℓ, the impact of the number of systems z in the Pareto set is also of interest. In Table 11, simulation results under the MP configuration are shown to illustrate the impact of z. Although the total number of correct comparison is the same as in Equation (21), under the MP configuration, we should consider two different comparisons, as there are some mean performance measures that are different by 2δ. In COMP of Table 11, we let f denote a "difficult" comparison (where the difference is δ), and e denote an "easy" comparison (where the difference is 2δ). For example, if $k = \ell = 3$ and $z = 2$, to retain a Pareto optimal system correctly, we need one easy comparison (against the system not in the Pareto optimal set) and one difficult comparison to be done correctly. Then, the system not in P requires two additional difficult comparisons. That leads to a total of $2(e + f) + 2f$ comparisons from the two systems in P and one system not in P. As we considered the dependency of the PCS on k and ℓ in Table 10, we fix $k = \ell$ in the MP configuration to concentrate on the parameter z that we have not yet examined. In Table 11, we have $k = \ell$ provided in the first row, and show results from different values of $1 \leq z \leq \ell$ along the column.

Table 11 also supports the validity of PSSP, as all the PCS's are greater than 0.95. In this table, we can discover two major causes that impact the PCS. Note that under the MP configuration, there are some mean performance measures that are different by 2δ. Therefore, comparing the diagonal with $k = \ell$ in Table 10 and the first row of Table 11 (where $z = 1$ as in the OP configuration), we can see that the MP configuration yields higher PCS. This implies that the PCS is larger when some differences of performance measures are greater than the indifference zone, as expected. Secondly, for a fixed k and ℓ, the PCS tends to be larger when the number of the systems in Pareto optimal z is not extreme (i.e., $1 < z < k = \ell$), especially when z is not extremely small. This is very obvious when we compare the cases when $z = 1$ and $z = 3$ when $k = \ell = 4$, where the COMPs are the same, but the PCS's are significantly different. Considering the probability of incorrect selection helps with understanding this behavior. When $z = 1$, there is only one system in the Pareto set, and hence, just one wrong comparison with the only system in P can result in an incorrect selection of an undesirable system. However, when $z = 2$, it can be either system 1 or 2 that eliminates the undesirable systems. Similarly, when $k = z = \ell$, the only way to achieve correct selection is not to eliminate any system, and hence one wrong comparison (in objective i for any system i) will likely lead to an incorrect selection. Thus, the PCS eventually starts decreasing when z increases to k.

Table 11: Simulation Results under MP Configuration

z	$k=\ell$	2	3	4	5
1	PCS	0.976	0.972	0.967	0.966
	REP	102	187.3	249.5	299.2
	COMP	$e+f$	$2e+4f$	$3e+9f$	$4e+16f$
2	PCS	0.956	0.992	0.997	0.998
	REP	122	174.3	224.25	268.4
	COMP	$2f$	$2e+4f$	$4e+8f$	$6e+14f$
3	PCS		0.978	0.994	0.998
	REP		212	236.75	267.8
	COMP		$6f$	$3e+9f$	$6e+14f$
4	PCS			0.989	0.997
	REP			276.25	288.4
	COMP			$12f$	$4e+16f$
5	PCS				0.994
	REP				325.4
	COMP				$20f$

4.5.3 Performance

In this section, we provide results that show the impact on the performance of PSSP of the number of systems k, number of objectives ℓ, number of the Pareto optimal systems z, number of systems with higher variability, and number of objectives with higher variability.

In Table 10 from the previous section, it is easy to see that the average number of observations REP increases both in k and ℓ, with more significant impact from k. Also, Table 11 shows that the REP increases as k and ℓ increase. However, the impact of z is not very obvious. Comparing the two rows of the first column, REP increases with the difficulty of COMP. However, comparing the case when $k = \ell = 4$, $z = 1$ with $k = \ell = 4$, $z = 3$, where the COMPs are equal, suggests that the number of possible paths that leads to correct selection also plays a role. That is, for $z = 1$, the only possible (correct) elimination is when all the other systems are compared correctly with system k. However, for $z = 3$, an undesirable system can be eliminated by any of the three systems in the Pareto set.

In the following, we isolate the influence of the number of systems k from the number of objectives ℓ, and see the change in performance with respect to k and ℓ separately. To begin with, we use the PP configuration, where we can fix ℓ and z while controlling the number of systems k. As in the example in Figure 9, let $z = 5$, $\ell = 2$, and k varies from 5 to 10. Under this setting, the average number of replications REP is illustrated in Figure 10. As shown in the graph, the REP increases as k grows, as long as the other parameters are fixed, but the rate of the growth decreases. This is intuitive as the additional decision to be made when the number of systems increases from $k - 1$ to k, is the elimination of one system, and it may be the case that the number of observations collected for $k - 1$ systems are already large enough for the additional decision.

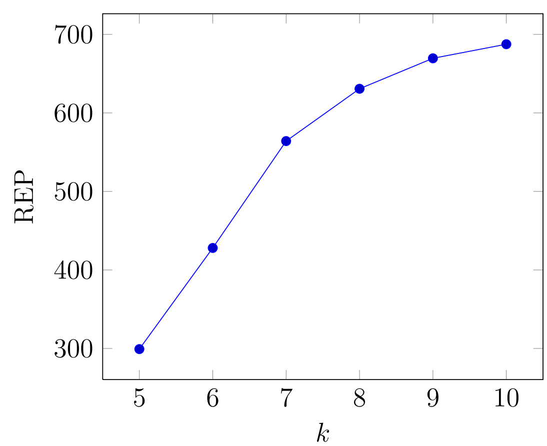

Figure 10: Effects of the Number of Systems

In Figure 11, we look at the impacts of the number of objectives ℓ on performance. To keep the number of systems k, and the number of systems z in the Pareto optimal set constant, we use the OP configuration with $k = 2$ and ℓ varying from 5 to 10. In this case, REP increases almost linearly as ℓ grows, with other parameters fixed. This is reasonable, as the lower bound that decides β decreases linearly with ℓ, as can be seen in Theorem 4.4.1, and $k \leq \ell$ is constant. However, the rate of increase in Figure 11 is smaller than that of Figure 10. That is, the impact of the number of systems is greater than that of the number of objectives. Again, considering COMP in Table 10, this is not surprising as the decrease of the lower bound for fixed $\ell \leq k$ is super-linear when k increases.

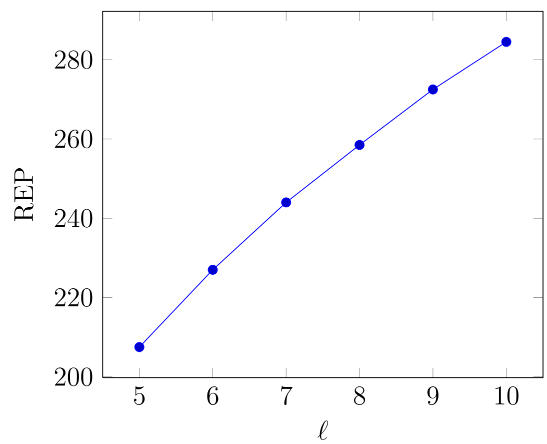

Figure 11: Effects of the Number of Objectives

Finally, we provide the results with different variances of performance measures. Consider the OP configuration with $k = \ell = 5$. We assume there are low variances which are set to be 1, and high variances which are set to be 25. In Figure 12, we consider the case when the number of systems with high variances changes. All objectives in a system have the same variance, and system k will be the first to have the higher variance. That is, when number of systems with high variance is 1, then that system is system k (which is the Pareto optimal system). In this case, we observe that REP increases significantly as soon as a system with high variances is introduced. The rate of increase in REP decreases as the number of systems increases. This is because, for correct selection, all the systems should correctly compare with system k and be eliminated under this configuration. Therefore, introducing the higher variances to system $k \in P$ causes the acute increase in the number of required observations.

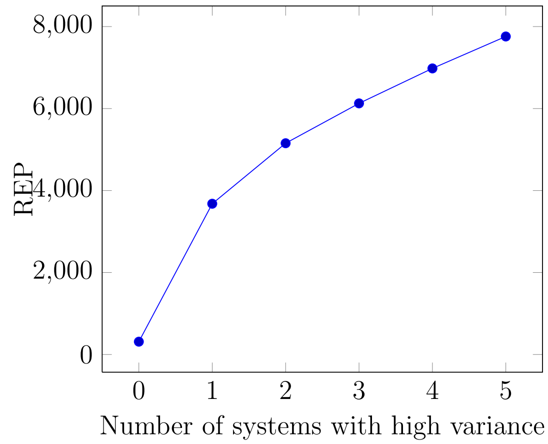

Figure 12: Effects of Number of Systems with High Variances, Starting from System k

Thus, the next experiment uses the same configuration with the same parameters as in Figure 12, namely the OP configuration with $k = \ell = 5$, but we now increase the variances of the performance measures of system $k \in P$ last. That is, only when all five systems have high variance, will system k have the higher variance. Figure 13 shows the change in REP relative to the number of systems with high variance. Now, we can see that the rate of increase is the most significant from 4 to 5, when the higher variance is introduced to the system that dominates all other systems, which is consistent with Figure 12.

In Figure 14, we consider the case where the number of objectives with higher variances in each system differs under the OP configuration with $k = \ell = 5$. In this graph, we can see the increase in REP as the number of objectives with higher variances increases. Similarly with Figure 12, REP increases as the number of objectives with higher variance increases. The rate of increase decreases, however, as the additional observations required to correctly compare the first objectives with higher variances may already have been sufficient to correctly compare an additional objective with higher variances. This also explains why the rate of increase decreases in Figure 13 until the system in the Pareto set has higher variances.

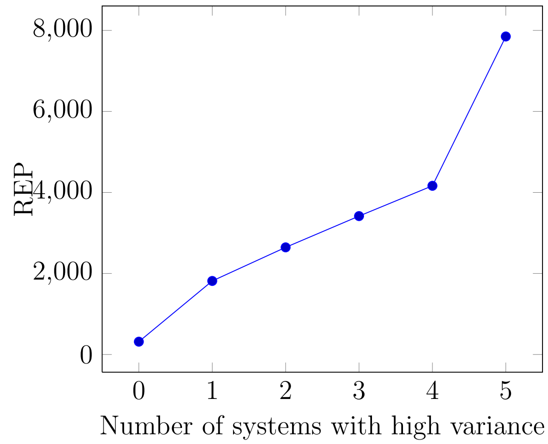

Figure 13: Effects of Number of Systems with High Variances, Starting from System 1

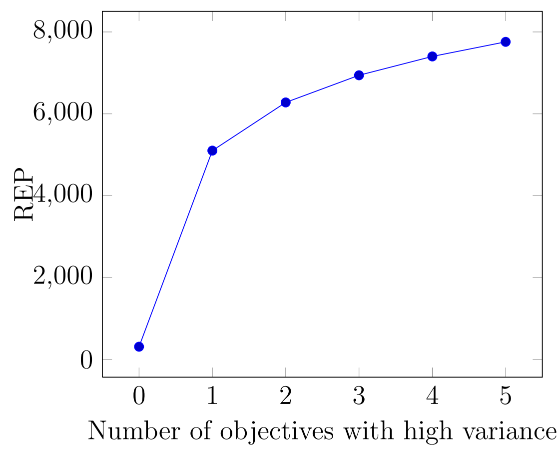

Figure 14: Effects of Number of Objectives with High Variances

4.5.4 Comparing Different Formulations

In Sections 4.5.2 and 4.5.3, we focused on the ER formulation. In this section, we would like to see the impact of different formulations, ER, RD, and EU. We introduce REP_k, which is the total number of observations over all the systems. That is $REP_k = k \times \text{REP}$. This will give a sense of how the number of total observations increases as k increases, as in the standard R&S literature, while we used the average number in the previous sections to be consistent in the comparison for different k and ℓ. (For example, obtaining one observation of $\ell = 1$ performance measure for k systems increases REP_k by k, whereas observing ℓ performance measures of $k = 1$ system increases REP_k by 1; in both cases REP increases by 1.)

First, we consider the same configuration and settings as in Table 10. That is, under the OP configuration, we change the number of systems k and number of objectives ℓ and look into the result of each formulation. The results are shown in Table 12. Clearly, validity of the procedure is maintained under all formulations as all the PCS's are greater than 0.95.

Note that while the PCS of the ER and EU formulations are similar, that of the RD formulation is considerably more conservative. This is due to the fact that this is a relatively "easy" configuration for RD, as there is only one system to retain and that system strictly dominates all other systems in all objectives. This results in the high PCS, but not necessarily lower REP_k for RD. Moreover, the performance of each formulation is as expected from their lower bounds in the validity proofs. That is, when $k > \ell$, then EU outperforms other two formulations with less conservative PCS and smaller REP_k. Similarly, when $k < \ell$, then RD performs better than other two. As expected from the discussion in Section 4.4.4, the relative savings in REP_k is greater for RD than for EU.

Table 12: OP Configuration under Different Formulations

k		ℓ = 2			3			5			10		
		ER	RD	EU	ER	RD	EU	ER	RD	EU	ER	RD	EU
2	PCS	0.955	0.999	0.956	0.956	1.000	0.955	0.956	1.000	0.955	0.955	1.000	0.956
	REP_k	244	244	244	314	272	314	415	303	415	569	331	570
3	PCS	0.971	1.000	0.963	0.957	1.000	0.958	0.956	1.000	0.956	0.956	1.000	0.957
	REP_k	554	555	509	618	618	617	791	695	791	1056	794	1056
5	PCS	0.983	1.000	0.974	0.975	1.000	0.968	0.960	1.000	0.960	0.960	1.000	0.960
	REP_k	1268	1268	1137	1419	1419	1336	1598	1598	1598	2072	1809	2072
10	PCS	0.992	1.000	0.986	0.987	1.000	0.980	0.979	1.000	0.971	0.959	1.000	0.957
	REP_k	3411	3416	3032	3783	3787	3442	4221	4224	3976	4781	4779	4782

We also consider the MP configuration, where the number of Pareto optimal systems z varies, as in Table 11. In Table 13, REP_k is almost identical over the formulations when $k = \ell$, as the lower bounds in Theorems 4.4.1, 4.4.2, and 4.4.3 are all identical in such case. However, under the MP configuration, we can see that, when z increases, the procedure becomes less conservative for the RD formulation compared to Table 12 as now multiple systems need to be retained to make a correct selection for the RD formulation. By contrast, as z increases, the EU formulation becomes more conservative. Note that when $k = z$, any outcome does not violate the constraints for CS_{EU}, as there are no system to eliminate.

Table 13: MP Configuration under Different Formulations

z	$k = \ell$	2			3			4			5		
		ER	RD	EU	ER	RD	EU	ER	RD	EU	ER	RD	EU
1	PCS	0.976	1.000	0.977	0.972	1.000	0.970	0.967	1.000	0.968	0.966	1.000	0.96512
	REP_k	204	204	204	562	561	560	998	998	998	1496	1496	1497
2	PCS	0.956	0.955	1.000	0.992	0.993	0.999	0.997	0.998	0.999	0.998	0.999	0.99898
	REP_k	244	245	245	523	522	524	897	896	896	1342	1342	1342
3	PCS				0.978	0.978	1.000	0.994	0.994	1.000	0.998	0.998	0.99995
	REP_k				636	637	636	947	947	947	1339	1340	1339
4	PCS							0.989	0.989	1.000	0.997	0.996	1.000
	REP_k							1105	1105	1104	1444	1445	1444
5	PCS										0.994	0.994	1.000
	REP_k										1626	1627	1627

From Tables 12 and 13, we can see that our procedure always yields the probability of correct selection to be greater than or equal to $1 - \alpha$ (set to be 0.95 throughout the experiments), regardless of the problem formulation. We can see a clear tendency of number of required replications increasing when the number of systems k, or the number of objectives ℓ, increases. The effect is more pronounced when the number of systems increases. Finally, as expected from the lower bound of the probability of correct selection in the validity proofs of each formulation, RD performs best when $k < \ell$, and EU is best when $k > \ell$.

4.6 Conclusion

In this chapter, we provided the PSSP procedure that estimates a Pareto set using Ranking and Selection when there are multiple objectives. We introduced three different formulations, namely Eliminate the undesirable and Retain the desirable (ER), Retain the Desirable (RD), and Eliminate the Undesirable (EU) depending on the desired properties of the final set produced by the procedure, and proved validity of our procedure for each formulation. Finally, we introduce three mean configurations that are motivated by the Difficult Mean and Monotone Increasing Mean configurations that are widely used in the Ranking and Selection literature. The experimental results under all formulations and configurations show that the procedure estimates the Pareto set at the desired level of confidence, supporting validity of the procedure. The probability of correct selection shows that the selection gets more conservative, as the discrepancy between the actual and worst-case numbers of required correct comparisons becomes larger. Also, the results display the impact of several factors, such as the number of systems k, number of objectives ℓ, variability of performance measures, and problem formulation on the performance of PSSP. In particular, the performance is better for smaller k and ℓ, but is more impacted by the number of systems k. Similarly, the performance is better when the performance measures do not have high variance. Finally, the Retain the Desirable (RD) formulation performs better than other two formulations when $k < \ell$, and the Eliminate Undesirable (EU) formulation improves performance when $k > \ell$.

CHAPTER V

DIFFICULT INPUT ANALYSIS PROBLEMS: EXPERIMENTS USING THE M/G/1 QUEUE

5.1 Introduction

In this chapter, we consider the simulation input analysis problem of fitting a distribution to data. We assume that the problem is difficult in that the underlying distribution is not a known theoretical distribution. In such cases, an empirical distribution may be a good choice. However, one obvious restriction of an empirical distribution is that in its nature, the support is bounded, while that of many existing distributions are not. Thus, an empirical distribution can be a poor estimate when the underlying distribution is unbounded.

For a remedy in such case, Bratley, Fox, and Schrage [15] suggest to use a quasi-empirical distribution. In particular, when there is a reason to believe that the support of an underlying distribution is non-negative (for example, waiting time in system, failure rate, positive observations, etc.), they suggest to fit an empirical distribution to most of the given data, and to fit a shifted exponential for a few extreme data on the right tail. They are aware of the fact that this is not a rigorous approach, but claim that it is an easy-to-use and often rational choice. The use of the exponential tail is supported by the work of Weissman [95], stating that the successive spacings between the k largest observations become asymptotically exponential for a wide class of distributions, such as the exponential, Weibull, normal, lognormal, and logistic distributions.

However, the importance of heavy-tails in real-life data has emerged and the motivation of our work starts from the fact that the Pareto distribution has gained popularity upon increasing necessity of incorporating heavy-tail distributions, but that it is not one of those distributions that can be approximated by an exponential tail as in Weissman [95]. Therefore, our interest lies in identifying an approach that can be applied in more generality, with reasonable ease, so that it can be used for simulation in practice. For this purpose, we examine the impact of a few different approaches on the M/G/1 simulation. We choose this model as it is a simple model that is representative of the models used in the discrete-event simulation. More specifically, we are interested in the total waiting time in the queue in steady-state, W, which has the known mean from the Pollaczek-Khinchine formula:

$$\mathbb{E}[W] = \frac{\rho + \lambda\mu\mathrm{Var}(S)}{2(\mu - \lambda)}, \tag{22}$$

where S is the service time random variable, λ is the reciprocal of the mean of the interarrival time, μ is the service rate, and $\rho = \frac{\lambda}{\mu}$ is the system utilization. We assume that we do not know the service time distribution but have a sample data set. We are particularly interested in tail probabilities for W, as Sigman [87] stated that when the service time distribution satisfies a certain set of assumptions, then the tail probabilities of W can be approximated by:

$$\mathbb{P}\{W > x\} \sim \frac{\rho}{1-\rho}\mathbb{P}\{S_e > x\},$$

where S_e is the integrated tail of the service time random variable. The Cumulative Distribution Function (CDF) of the integrated tail distribution, $F_e(x)$, of a random variable X with CDF $F(x)$ for $x > 0$ is:

$$F_e(x) = \frac{1}{\mathbb{E}[X]} \int_0^x \bar{F}(y)dy,$$

where $F(0) = 0$, and $\bar{F}(y) = 1 - F(y)$, $\forall y$. Therefore, we believe that properly estimating the tail distribution of the service time distribution is crucial when simulating to estimate the tail probabilities of the waiting time. Therefore, to correctly simulate the mean and tail probabilities of the waiting time, the first and second moments and the tail probabilities of the service time distribution should be estimated correctly. Note that most of the aforementioned studies in M/G/1 with a heavy-tail distribution concentrate on the case when the tails are extremely heavy, so that the second moment of the distribution is infinite. Therefore, we would like to see how approaches that take the Pareto distribution into consideration performs, when we know the system is stable and the waiting time is finite.

In Section 5.2, we introduce some of the previously studied approaches that we adopt, along with the modifications we introduce. In Section 5.3, we show the effects of these choices on a simulation through the numerical experiments. Finally we present our suggested approach and conclusions in Section 5.4. Additional numerical results for this chapter are provided in Appendix A.

5.2 Candidate Approaches

Our purpose in this chapter is to investigate different approaches to estimate an unknown input distribution and suggest one to use when a certain amount of sample data is given, but either we have little knowledge on its distribution or the distribution does not fit known theoretical distributions. When we know the form of the underlying distribution *a priori*, then the input analysis problem boils down to estimating the parameters of the distribution. Also, when a theoretical distribution under consideration is a good fit for the data, then after estimating the required parameters for the set of competing theoretical distributions, each distribution can be evaluated and ranked using some measures of goodness-of-fit. Standard simulation textbooks such as Law and Kelton [67] are good sources of information on this subject. By contrast, we look into cases that are more difficult because theoretical distributions do not provide a good fit to the data.

Note that we consider non-negative observations throughout the chapter. This will not be limiting the use of the suggested approach, as if the distribution is unbounded in both directions, then we can separate the data that are negative and positive. We can take absolute values of negative observations and use the same approach. We consider three different categories of distributions, namely, bounded, light-tailed, and heavy-tailed. When the original distribution is bounded, then as mentioned earlier in Section 5.1, an empirical distribution may be a good fit in most cases. However, when the original distribution is unlikely to be bounded, then there may be cases where we should consider other possibilities that incorporate the fact that the underlying distribution may have an unbounded right tail.

In the following sections, we consider multiple candidate approaches that may be used in this case, and provide some technical details. To begin with, in Section 5.2.1, we introduce the mixture of empirical and exponential distributions adapted from Bratley et al. [15]. In Section 5.2.2, a mixture of empirical and Pareto distributions is introduced, and in Section 5.2.3, we modify the algorithm in Section 5.2.1 by using a different approach to decide the cut-off point between the empirical and exponential distributions. In Section 5.2.4, we briefly state an algorithm that resembles that of Section 5.2.1, but with the Pareto tail instead of an exponential tail. In Section 5.2.5, a hypothesis test for rejecting the Pareto tail when applicable is stated, and in Section 5.2.6, an approach to compare between two tail distributions, the exponential and Pareto, is described. Finally, in Section 5.2.7, we suggest several heuristic approaches.

5.2.1 Classic Quasi-empirical Distribution with Exponential Tails

When we have no insight on distribution, and an empirical distribution is suspected to be a poor fit due to the presence of a right tail, Bratley, Fox, and Schrage [15] suggest to use a quasi-empirical distribution, which is a mixed empirical and exponential distribution. Let $X_1, X_2, \ldots, X_n$ be the independent observations from the distribution to estimate, and let $X_{[i]}$, $i = 1, 2, \ldots, n$, denote the ordered statistics of

X_i, $i = 1, 2, \ldots, n$, in non-decreasing order. Bratley et al. [15] suggest to pick a value $k \in \{1, 2, 3, 4, 5\}$ that best matches the variance of the sample data, and to fit an interpolated piece-wise linear CDF to the first $n-k$ data points $X_{[i]}$, $i = 1, 2, \ldots, n-k$, and a shifted exponential to the right of $X_{[n-k]}$. The idea is that when the sample size n goes to infinity, there exists a wide range of distributions that have the successive spacings between the k largest observations become asymptotically exponential, as stated in Weissman [95]. Although Bratley et al. [15] chose to use a piece-wise linear CDF for the empirical part of the data in their original approach, for simplicity, we used a piece-wise constant CDF throughout. Also, in the experiments, instead of choosing $k \in \{1, 2, 3, 4, 5\}$ by matching the second moment, we use $k = 5$. Note that Bratley et al. [15] appear to be considering the case when the sample size n is relatively small (for example, 25). In our experiments, the number of observations is significantly larger than 25, and hence even the largest suggested value, $k = 5$ is significantly small compared to the number of observations. Therefore, in this approach, we will focus on experimenting with the effectiveness of using a small number of data points to fit a tail.

Assume that $F(0) = 0$ and define $X_{[0]} = 0$. Then the CDF for this quasi-empirical distribution follows:

$$F(x) = \begin{cases} \dfrac{i}{n}, & X_{[i]} \leq x < X_{[i+1]}, i = 0, 1, \ldots, n-k-1, \\ 1 - \dfrac{k}{n} \exp(-\lambda(x - X_{[n-k]})), & x \geq X_{[n-k]}, \end{cases} \tag{23}$$

where

$$\lambda = \frac{k}{\sum_{i=n-k+1}^{n} (X_{[i]} - X_{[n-k]})}. \tag{24}$$

We can see that the expected value of the random variable with this CDF matches the average estimated from the samples $X_1, X_2, \ldots, X_n$.

Let $\lceil u \rceil$ denote the smallest integer larger than u. To generate independent and identically distributed random variables $Y_1, Y_2, \ldots$ with this distribution using inverse transform when we have the sample data $X_1, X_2, \ldots, X_n$, follow these steps:

1. Generate independent pseudo-random numbers $U_1, U_2, \ldots$.

2. If $U_i \leq \frac{n-k}{n}$, then $Y_i \leftarrow X_{[\lceil nU_i \rceil]}$,
   else $Y_i \leftarrow X_{[n-k]} - \frac{\ln[n(1-U_i)/k]}{\lambda}$.

5.2.2 New Quasi-empirical Distribution with Pareto Tails

While an exponential tail may be a good resort for a light-tailed distribution, it is hard to believe that a heavy-tailed distribution also can be approximated well with an exponential tail, as mentioned in Section 5.1. With the growing interest in and documented presence of heavy-tailed distributions, we would like to explore the effects of using Pareto distributions (or power-law distributions) for the right tail.

There are two parameters that define the Pareto distribution, namely the lower bound, x_{min}, from which the Pareto distribution can be fit and the tail index α. The following is a widely accepted representation of the CDF of the Pareto distribution:

$$F(x) = 1 - \left(\frac{x}{x_{min}}\right)^{1-\alpha}, \tag{25}$$

for all $x \geq x_{min}$, $\alpha > 1$. To fit this distribution with given data, we adopt an algorithm from Clauset et al. [29] as it is a recent and widely accepted framework that is generally applicable when sample data is given. However, as can be seen from Equation (22), the mean waiting time in an M/G/1 queue is infinite when the service time distribution has an infinite variance. Therefore, when we consider the Pareto distribution, we concentrate on cases where $\alpha > 3$, so that $\mathbb{E}[W]$ is well defined (see Equation (22)), while the focus in [29] is on $2 < \alpha < 3$. For analysis and simulation studies considering cases where the variances are not finite, see the references [1, 5, 14, 36] introduced in Section 5.1.

We now describe the approach in Clauset et al. [29]. First, their tail index α uses a Maximum Likelihood Estimator (MLE) that is equivalent to the traditional Hill's estimator, which was introduced in [55]. Assuming that the data is drawn from a Pareto distribution for $x \geq x_{min}$, the MLE for this tail index follows:

$$\hat{\alpha} = 1 + m \left[\sum_{i=1}^{m} \ln \left(\frac{X_i'}{x_{min}} \right) \right]^{-1}, \tag{26}$$

where X_i', $i = 1, 2, \ldots, m$, are the observed values that are greater than or equal to the lower bound x_{min}. In practice, it is very rare to have the entire data obey a power law, and hence the choice of x_{min} is also essential. Therefore, before calculating the estimate of the tail index, we should first estimate x_{min}. One can visually pick x_{min} by observing a few different plots, but we would like to consider a more objective method that is presented in Clauset et al. [29] and implemented in Gillespie [42]. The idea of this approach is to choose the value of $\hat{x}_{min}$ among all observations $X_1, \ldots, X_n$ to minimize the discrepancy of the fitted tail distribution and the empirical distribution of the data above $\hat{x}_{min}$. If $\hat{x}_{min}$ is estimated to be smaller than the actual cutoff, then the estimator of the tail index α will also include the data that are not actually heavy-tailed. On the other hand, if $\hat{x}_{min}$ is larger than the actual x_{min}, then a number of observations will be lost.

Among measures that quantify the difference between two distributions, one of the most commonly used ones for non-normal data is the Kolmogorov-Smirnov (KS) statistic. This is the maximum distance between the CDF of the data and the fitted model, which Gillespie [42] estimates as follows:

$$D(x_{min}) = \max_{X_i \geq x_{min}} |S(X_i) - P_{x_{min}}(X_i)|, \tag{27}$$

where $S(x)$ is the piecewise-constant CDF of the empirical sample data to fit (i.e., the data points X_i with $X_i \geq x_{min}$), and $P_{x_{min}}(x)$ is that of the fitted Pareto distribution as in Equation (25) with α as in Equation (26). Then the estimate for x_{min} is a data point from a sample data set that satisfies:

$$\hat{x}_{min} \in \mathop{\arg\min}_{x_{min} \in \{X_1, \ldots, X_n\}} D(x_{min}), \tag{28}$$

see Gillespie [42]. Once x_{min} is properly chosen, we can use the estimate of tail index obtained from Equation (26), and we have a fitted Pareto distribution.

Let $k \in \{0, \ldots, n-1\}$ be such that $\hat{x}_{min} = X_{[n-k]}$. Then $m = k+1$ and α can be estimated as in Equation (26) with $\{X'_1, \ldots, X'_m\} = \{X_{[n-k]}, \ldots, X_{[n]}\}$. Then the CDF of the random variable with this mixture distribution follows:

$$F(x) = \begin{cases} \dfrac{i}{n}, & X_{[i]} \leq x < X_{[i+1]}, i = 0, 1, \ldots, n-k-1, \\ 1 - \dfrac{k}{n}\left(\dfrac{x}{X_{[n-k]}}\right)^{1-\hat{\alpha}}, & x \geq X_{[n-k]}. \end{cases} \tag{29}$$

Then, the generation of independent and identically distributed random variables $Y_1, Y_2, \ldots$ with this distribution from the sample data $X_1, X_2, \ldots, X_n$, using the inverse-transform method can be done as follows:

1. Generate independent pseudo-random numbers $U_1, U_2, \ldots$.
2. If $U_i \leq \frac{n-k}{n}$, then $Y_i \leftarrow X_{[\lceil nU_i \rceil]}$,
   else $Y_i \leftarrow X_{[n-k]}[\frac{n(1-U_i)}{k}]^{\frac{1}{1-\hat{\alpha}}}$.

Other than the most obvious deviation from the approach of Bratley et al. [15] that we consider the Pareto distribution, not the exponential distribution, for the right tail, the estimation of the cutoff point differs. In the suggested approach in this section (from [29, 42]), x_{min} can be any data point X_i, $i = 1, \ldots, n$, and hence k can be any integer from 0 to $n-1$, whereas the approach in [15] chooses $k \in \{1, 2, 3, 4, 5\}$. This may be due to the fact that Bratley et al. [15] seem to focus on the case when n is small, but it is interesting that such a small number as 5 can be used to fit a tail distribution.

5.2.3 New Quasi-empirical Distribution with Exponential Tail

As stated in Sections 5.2.1 and 5.2.2, Bratley et al. [15] restrict the number of extreme values k to be within a small range of numbers. With possibilities that we may often be in a situation with more abundant data than was the case back then, we experiment with the estimation of x_{min} without such restriction, when the underlying distribution may be light-tailed. This approach uses a mix of empirical and exponential distributions, as in Section 5.2.1, but uses the KS statistic to choose the cutoff point as in Section 5.2.2. More specifically, we choose $\hat{x}_{min}$ as in Equation (28), where $P_{x_{min}}(x)$ in Equation (27) is now a CDF of a fitted shifted exponential distribution. Estimation of the parameter λ of this exponential distribution stays as in Equation (24), except that we now use $x_{min} = X_{[n-k]}$, where $k \in \{0, 1, \ldots, n-1\}$. The CDF of the resulting (mixed) random variable is the same as in Equation (23) with the aforementioned modifications in k. Generation of the random variables can also be done similarly as in Section 5.2.1.

5.2.4 Quasi-empirical Distribution with Pareto Tails and $k = 5$

In this section, we consider fitting the Pareto distribution to a fixed number of extreme values (without estimating x_{min}). Again, as in Section 5.2.1, we would like to see the impact of estimating the tail using a small number k of extreme data points, and hence used $k = 5$. In that case, the $n - k$ smallest data points will be used to fit the empirical distribution, and a Pareto distribution will be fitted to the right of $X_{[n-k]}$. Let $x_{min} = X_{[n-k]}$; then the tail index α can be obtained as in Equation (26) where $m = k+1$ and $X'_1, \ldots, X'_m$ are substituted with $X_{[n-k]}, \ldots, X_{[n]}$. Also, the CDF of this quasi-empirical distribution can be written as in Equation (29), and the generation of the random variables done as in Section 5.2.2.

5.2.5 Hypothesis testing for Heavy-tailed Distribution

Now that we have approaches to fit both light-tailed and heavy-tailed distributions to a data set, we need to choose which provides a better fit. Again, a graph can be used for visual decision, but we will concentrate on more quantitative approaches. In this section, we state one of the approaches in Clauset et al. [29] in testing whether the data is actually heavy-tailed (the other one is provided in Section 5.2.6), which is to perform a hypothesis test on whether the underlying distribution is actually heavy-tailed. That is, given a data set, $X_1, \ldots, X_n$, and the estimated quasi-empirical distribution with Pareto tail, we test the hypothesis that the estimated distribution is the actual underlying distribution. This can be done with a goodness-of-fit test using the KS statistic. This test first generates a large number of synthetic data sets of size n drawn from the combined Pareto distribution with the estimated parameters and the empirical distribution for values $x < x_{min}$. Then, for each synthetic data set, we estimate the parameters α and $\hat{x}_{min}$ for the Pareto distribution using Equations (26) and (28), along with the KS statistic as in Equation (27). Then the *p-value* generated from this test is the probability that the synthetic data sets have larger value of the KS statistic compared to the original sample. Roughly speaking, it is the probability that if we apply the same distribution-fitting process to a synthetic data set of the same size from the estimated distribution, the resulting value of the KS statistic will be larger than that observed for the original data set. Thus, a significantly small number indicates that the data appears to be far from the combined empirical and Pareto distribution.

Clauset et al. [29] state that the accuracy of the *p-value* is known to be approximately $(4s)^{-0.5}$, where s is the number of synthetic sample sets. That is, for accuracy of the *p-value* to the second digit, at least 2500 sample sets should be generated, which yields an accuracy of 0.01. In the experiment, we set $s = 1000$, resulting in an accuracy of 0.015. The suggested threshold of *p-value* in [29] is 0.1, which can be interpreted as the probability of rejecting the null hypothesis when the null hypothesis is true. However, we would also like to consider being conservative in selecting the heavy Pareto tail as the true distribution. Thus in Section 5.3, we also experiment with *p-values* of 0.5 and 0.9, and observe the effects of the different thresholds. Obviously, as the *p-value* gets larger, the rejection of the Pareto tail becomes more frequent.

5.2.6 Comparison between Tail Distributions

Another approach from [29] in deciding whether the tail is heavy-tailed is to directly compare fitted light-tailed and heavy-tailed distributions. In doing so, Clauset et al. [29] use the method proposed by Vuong [94]. This approach compares the likelihood ratio of two competing distributions, and suggests which distribution is a better fit to a given data set, if applicable.

More specifically, the likelihood of a set of independent observations $X'_1, \ldots, X'_m$, of extreme values that are used to fit a tail distribution, can be written as

$$L_j = \prod_{i=1}^{m} p_j(X'_i),$$

where $p_j(x)$ is the Probability Density Function (PDF) of distribution $j = 1, 2$. Then, the log likelihood ratio R of the two distributions follows:

$$R = \ln \left(\frac{L_1}{L_2} \right).$$

Therefore, $R > 0$ implies a possibility that distribution 1 is a better fit, and $R < 0$ supports better goodness of fit of distribution 2. As R is essentially a sum of independent terms, from the Central Limit Theorem, R can be approximated by a normal distribution. Then, the *p-value* of this test is the probability that the absolute value of this normal random variable is larger than the observed $|R|$. Therefore, if the *p-value* is small, then it is likely that one distribution is a superior fit compared to the other, but when not, it is difficult to conclude which distribution is a better fit.

In comparing the two tail distributions in our approach, we would compare an exponential tail against a Pareto tail. However, this approach assumes that both distributions have the same number of data points, thus requiring x_{min} to be equal in both distributions. In the numerical experiments in Section 5.3, we will see the impact of this test applied with x_{min} estimated for both the exponential and Pareto distributions.

5.2.7 Other Approaches

In addition to the approaches considered in the previous sections, there are several heuristic approaches to consider. For example, the KS statistics for both the fitted Pareto and exponential tails are known, as these values are calculated to determine the best x_{min} in Sections 5.2.2 and 5.2.3, respectively. Therefore, we can directly compare the magnitude of the two. That is, if the KS statistic for the exponential distribution is larger, then we choose the Pareto distribution for the right tail, and vice versa. The approach in Section 5.2.5 also utilizes the KS statistic of the fitted Pareto distribution to test if we can reject the Pareto distribution. The difference in this approach is that we now compare the KS statistic of the Pareto distribution with that of the exponential distribution, without generating extra sample sets or calculating extra statistics, while in the previous approach the comparison was against the synthetic data sets.

There may be other heuristic approaches that can be used depending on the nature of the simulation of interest. In this chapter, as our focus is on simulating an M/G/1 queueing system, we may have some additional knowledge, such as the stability of the system, which will restrict the values of the mean interarrival and service times. This implies that we can impose some bounds on the parameters that we are estimating to ensure such characteristics. The same is true when it is known that certain moments of the performance measures of interest are finite, as we can see from the fact that $\mathbb{E}[W]$ is finite only when $\text{Var}\{S\}$ is finite (see Equation (22)). Finally, in the experiments in Section 5.3, we will explore the possibility of combining the aforementioned approaches in choosing one distribution over the other.

5.3 Numerical Results

We use simulation results for an M/G/1 queueing system to present the effects of the algorithms described in Section 5.2 in estimating the service time distribution. In this model, the interarrival time distribution is exponential, and we assume that it is known. In that way, we can concentrate on the impact of our approach on estimating the service time distribution.

In the following experiments, we first generate a sample data set of size $n \in \{100, 500, 1000\}$ from a certain distribution, and apply the approaches considered in Section 5.2 to fit a distribution to the data. Let A denote an interarrival time random variable and S a service time random variable. Then, we run an M/G/1 simulation with the known interarrival distribution, which is the exponential with parameters that sets the system utilization, $\rho = \mathbb{E}[S]/\mathbb{E}[A]$, as specified in each experiment, and the fitted and chosen service distribution. The underlying distributions used to generate the service time data are the exponential, Pareto, uniform, mixture of uniform and exponential, and mixture of uniform and Pareto distributions. As we are interested in steady-state performance, we use the batch means approach to estimate $\mathbb{E}[W]$ and $\mathbb{P}\{W > t_i\}$, for $i = 1, 2, 3$, where W is the mean waiting time in queue in steady-state. The choice of the thresholds t_1, t_2, t_3 is different in each experiment. The goal is to study the impact of the input distribution on the tail behavior, but not as far out in the tail as in rare-event simulation. The values of t_1, t_2, t_3 will be noted in each experiment. Then, we repeat this process with a new sample set of n data points N times, where N is the number of macro replications. In this way, we can obtain statistical validity of each steady-state performance measure, and not be biased by a single sample set generated for the input distribution.

We simulate the waiting times of the system using Lindley's equation [73] to generate the waiting time of the i-th customer, W_i, for $i = 1, 2, \ldots$. More specifically, when S_i is the service time of the i-th customer and A_i is the time between the arrival of the i-th and $(i+1)$-th customers, then W_i, for $i = 1, 2, \ldots$ can be calculated through:

$$W_{i+1} = \max(0, W_i + S_i - A_i),$$

where $W_1 = 0$. Then, we truncate the first 1000 observations to remove the initial bias. The number of batches is 30, where each batch contains 200,000 observations.

Finally, as a measure of accuracy, we use the Mean of Relative Error (MRE) of the observations from the N macro-replications. Let Z_i, $i = 1, \ldots, N$, be the observations of a performance measure of interest from any approach with a fitted service time distribution. Similarly, let Y_i, $i = 1, \ldots, N$, be the observations simulated with the true service time distribution. Let $\widehat{\mathbb{E}[Y]} = \frac{1}{N}\sum_i Y_i$. We use the estimated mean of each performance measure, as the true values of the tail probabilities for some service time distributions are not known in closed form. Then the accuracy measure, MRE, is estimated as follows:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{|Z_i - \widehat{\mathbb{E}[Y]}|}{\widehat{\mathbb{E}[Y]}}. \tag{30}$$

In each experiment, the approaches we discussed in Section 5.2 are applied and the results are compared. We now list all the approaches we consider with the notations of the approaches provided in parentheses:

1. Fitting an exponential distribution to the entire sample set (Exp);

2. Fitting a Pareto distribution to the entire sample set (Par);

3. Fitting an empirical distribution to the entire sample set (Emp);

4. Fitting a "New" quasi-empirical distribution with an exponential tail, as in Section 5.2.3 (NQE_{exp});

5. Fitting a "New" quasi-empirical distribution with a Pareto tail, as in Section 5.2.2 (NQE_{Par});

6. Fitting a "Classic" quasi-empirical distribution with an exponential tail, as in Section 5.2.1 using $k = 5$ (CQE_{exp});

7. Fitting a "Classic" quasi-empirical distribution with a Pareto tail, as in Section 5.2.4 using $k = 5$ (CQE_{Par});

8. Applying hypothesis testing with different *p-values* (0.1, 0.5, 0.9), as in Section 5.2.5 and using the empirical distribution or "New" quasi-empirical distribution with the exponential tail when the Pareto tail distribution is rejected ($HT^{emp}_{p\text{-}value}$ or $HT^{NQE}_{p\text{-}value}$);

9. Comparing the exponential and Pareto tails for the "New" quasi-empirical distribution through likelihood ratio using $\hat{x}_{min}$ from the exponential or Pareto distribution or both (to be conservative in choosing the Pareto), as in Section 5.2.6 (LR_{exp}, LR_{Par}, LR_{cons} – LR_{cons} only chooses the Pareto distribution when both LR_{exp} and LR_{Par} do);

10. Comparing the KS statistics calculated in Sections 5.2.2 and 5.2.3 (KS);

11. Bounding α below by a bound, e.g., to make a system stable and the second moment finite. When the bound is violated, an alternative distributions is used, namely the "New" quasi-empirical distribution Approach 4 or empirical distribution (B^{NQE}_{bound}, B^{emp}_{bound});

12. Combining the HT Approach 8 and bounded α Approach 11 ($HT + B$);

13. Combining the LR Approach 9 and bounded α Approach 11 ($LR + B$);

14. Combining the KS Approach 10 and bounded α Approach 11 ($KS + B$).

Before getting into the experiments, we elaborate on Approach 11. The motivation of imposing the bounds on the tail index is that we know the mean waiting time is infinite when the tail index is less than or equal to 3 (see Equation (22)). Thus bounding the tail index below by 3 helps improve performance. However, when the tail index is larger but extremely close to 3, the resulting M/G/1 queue may be very volatile. Therefore, when we know that the queueing system we are simulating is stable and the mean of the waiting time is finite, we consider being more conservative in choosing the Pareto tail. Therefore, we implement Approach 11 with both 3 and 4 as the bound.

We now provide additional information about the underlying distributions of the service times. As we are mostly interested in the tail behavior, we categorize the distributions into bounded, light-tailed, and heavy-tailed distributions. For the category of bounded distributions, we use the uniform distribution with range $(0, 1)$, and the service rate is 1. For this distribution, the traffic intensity $\rho = 0.25$. For light-tailed distributions, we use both the exponential and mixture of uniform and exponential distributions. For the exponential distribution, the traffic intensity $\rho \in \{0.5, 0.75\}$. When using the mixture of the uniform and exponential, for a certain probability q, a uniform random variable is generated, and with probability $1 - q$, a shifted exponential random variable is generated. We shift the exponential so that the random variables from the uniform and shifted exponential do not overlap, but the support is continuous. We look at $q = 0.25, 0.75$, and this leads to $\rho = 0.475$ and $\rho = 0.375$. In the following tables, we denote such distribution as q Unif(0,10) + $(1 - q)$ Exp(1). Finally, for heavy-tailed distributions, we use the Pareto distribution and mixture of uniform and Pareto distributions. For the Pareto distribution, we consider Par(10, 3.5), Par(10, 4), Par(10, 5), and Par(10, 10), where the first parameter is x_{min} and the second, α. The traffic intensities of these cases are, 0.75, 0.75, 0.667, and 0.5625, respectively. The mixture of the uniform with range $(0, 10)$ and Pareto with $x_{min} = 10$ and $\alpha = 4$ is constructed similarly as the mixture of the uniform and exponential. For $q = 0.25, 0.75$, the traffic intensities are 0.625, 0.375, respectively. The values of t_1, t_2, t_3 are provided in the tables where the results from each distribution are displayed.

Finally, each steady-state simulation generates an estimate of the mean $\mathbb{E}[W]$ with half-width less than 1% of the estimated $\mathbb{E}[W]$ for most choices of the underlying and fitted distributions other than the Pareto distribution. However, there are two main exceptions. The first is when the estimated mean service time is significantly larger than the actual value, leading to a traffic intensity close to 1. The second involves the Pareto distribution when the actual or estimated tail index α is small. This is because the variance of the Pareto distribution increases sharply as α decreases, impacting the variance of the waiting time. All the half-widths of the tail probabilities are less than 10^{-4}. These estimates are then used to compute the MRE as in Equation (30). The number of macro replications $N = 108$ for the tables provided in Appendix A, and $N = 216$ for those provided in Section 5.3 (Tables 14, 15, 16, 18, and 19).

In the following sections, we compare the approaches described in Section 5.2, as a function of the service time distribution and number of sample service times. In Section 5.3.1, we examine the mean relative error of the different approaches when the number of observations from the real service time distribution is $n = 100$, the smallest in this experiment. In Section 5.3.2, we consider the cases when we have more observations of the service time, namely $n \in \{500, 1000\}$. Finally, in Section 5.3.3, we provide a brief summary of the numerical experiments, and compare the approaches.

5.3.1 One Hundred Observed Service Times

In this section, we consider the case when we have 100 data points observed from the true service time distribution. We examine the results obtained from the different approaches and compare the effectiveness of each approach. In Tables 14 and 15, the results when the underlying service time distribution is the exponential are shown. The traffic intensities are 0.5 and 0.75, respectively, in Tables 14 and 15, while the service time distribution are kept the same. We can see that using a heavy right tail when the underlying distribution is light-tailed can be significantly off, from the Fitted Par, NQE_{Par}, and CQE_{Par} rows of Table 15. Also, the results from the HT and KS approaches show that they do not do a sufficiently good job in rejecting the Pareto distribution, even when the underlying distribution is the exponential. On the other hand, the likelihood ratio testing performs well in choosing the right distribution for the tail. We can also see that the bounds on α improve the performance of the algorithm, especially when the bound is 4. Finally, we can see that the overall performance is better when the traffic intensity is low as in Table 14, while the relative performance of the approaches is indifferent to ρ.

Table 14: Mean Relative Error: Exp(1), $\rho = 0.5$, $n = 100$

	Mean Relative Error	$\mathbb{E}[W]$ $= 1$	$\mathbb{P}\{W \geq 5\}$ ≈ 0.0410	$\mathbb{P}\{W \geq 8\}$ ≈ 0.0092	$\mathbb{P}\{W \geq 10\}$ ≈ 0.0034
Theoretical	Fitted Exp	0.232	0.457	0.711	0.912
	Fitted Par	5.31E+75	23.4	108	296
	Empirical	0.251	0.528	0.809	1.02
Quasi-Empirical	NQE_{exp}	0.241	0.481	0.745	0.951
	NQE_{Par}	39302	2.46	9.97	25.2
	CQE_{exp}	0.253	0.523	0.848	1.12
	CQE_{Par}	99.0	0.540	1.06	2.42
Hypothesis Test	$HT_{0.1}^{emp}$	39302	2.46	9.97	25.2
	$HT_{0.1}^{NQE}$	39302	2.46	9.97	25.2
	$HT_{0.5}^{emp}$	12143	2.26	9.05	22.7
	$HT_{0.5}^{NQE}$	12143	2.26	9.03	22.7
	$HT_{0.9}^{emp}$	863	1.41	5.10	12.2
	$HT_{0.9}^{NQE}$	863	1.39	5.07	12.2
Likelihood Ratio	LR_{exp}	0.241	0.481	0.745	0.951
	LR_{Par}	0.241	0.480	0.746	0.956
	LR_{cons}	0.241	0.481	0.745	0.951
Heuristics	KS	21.6	0.660	1.59	3.16
	B_3^{emp}	0.499	0.646	1.67	3.25
	B_3^{NQE}	0.491	0.616	1.61	3.17
	B_4^{emp}	0.254	0.524	0.842	1.140
	B_4^{NQE}	0.244	0.484	0.780	1.066
	$HT_{0.1}^{emp} + B_3^{emp}$	0.499	0.646	1.67	3.25
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.491	0.616	1.61	3.17
	$HT_{0.5}^{emp} + B_3^{emp}$	0.497	0.645	1.66	3.23
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.489	0.613	1.60	3.15
	$HT_{0.9}^{emp} + B_3^{emp}$	0.461	0.629	1.54	2.91
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.452	0.591	1.47	2.82
	$LR_{exp} + B_3^{NQE}$	0.241	0.481	0.745	0.951
	$LR_{Par} + B_3^{NQE}$	0.241	0.480	0.746	0.956
	$LR_{cons} + B_3^{NQE}$	0.241	0.481	0.745	0.951
	$KS + B_3^{NQE}$	0.250	0.479	0.791	1.09

On the contrary, Table 16 shows that fitting the exponential to the Pareto distribution is not a very good idea. However, it is noteworthy that fitting a light tail

Table 15: Mean Relative Error: Exp(1), $\rho = 0.75$, $n = 100$

	MRE	$\mathbb{E}[W]$ $= 3$	$\mathbb{P}\{W \geq 10\}$ ≈ 0.062	$\mathbb{P}\{W \geq 15\}$ ≈ 0.018	$\mathbb{P}\{W \geq 20\}$ ≈ 0.005
Theoretical	Fitted Exp	0.505	1.01	1.97	3.97
	Fitted Par	6.76E+71	15.2	55.6	197
	Empirical	0.501	1.01	1.95	3.89
Quasi-Empirical	NQE_{exp}	0.505	1.01	1.97	3.98
	NQE_{Par}	6.86E+04	3.72	11.91	37.9
	CQE_{exp}	0.502	1.01	1.97	3.95
	CQE_{Par}	174	0.886	1.85	5.07
Hypothesis Test	$HT_{0.1}^{emp}$	27781	3.65	11.66	37.0
	$HT_{0.1}^{NQE}$	27781	3.65	11.66	37.0
	$HT_{0.5}^{emp}$	12314	3.41	10.75	33.8
	$HT_{0.5}^{NQE}$	12314	3.41	10.75	33.8
	$HT_{0.9}^{emp}$	1169	2.43	7.09	21.1
	$HT_{0.9}^{NQE}$	1169	2.44	7.13	21.2
Likelihood Ratio	LR_{exp}	0.505	1.01	1.97	3.98
	LR_{Par}	0.505	1.00	1.97	3.99
	LR_{cons}	0.505	1.01	1.97	3.98
Heuristics	KS	388	1.24	2.85	7.01
	B_3^{emp}	1.64	1.50	3.79	9.84
	B_3^{NQE}	1.64	1.49	3.76	9.80
	B_4^{emp}	0.513	1.04	2.03	4.10
	B_4^{NQE}	0.517	1.03	2.02	4.13
	$HT_{0.1}^{emp} + B_3^{emp}$	1.64	1.50	3.79	9.84
	$HT_{0.1}^{NQE} + B_3^{NQE}$	1.64	1.49	3.76	9.80
	$HT_{0.5}^{emp} + B_3^{emp}$	1.64	1.50	3.76	9.77
	$HT_{0.5}^{NQE} + B_3^{NQE}$	1.63	1.49	3.74	9.74
	$HT_{0.9}^{emp} + B_3^{emp}$	1.55	1.41	3.42	8.69
	$HT_{0.9}^{NQE} + B_3^{NQE}$	1.55	1.41	3.43	8.76
	$LR_{exp} + B_3^{NQE}$	0.505	1.01	1.97	3.98
	$LR_{Par} + B_3^{NQE}$	0.505	1.00	1.97	3.99
	$LR_{cons} + B_3^{NQE}$	0.505	1.01	1.97	3.98
	$KS + B_3^{NQE}$	0.510	1.02	2.00	4.05

to an underlying Pareto distribution does not work as badly as the opposite case. Hypothesis testing with a higher threshold seems to be rejecting the correct Pareto distribution too often, while the likelihood ratio test does not perform well when the x_{min} is estimated using the exponential tail. One approach that still works well is to use the NQE approach with the exponential tail when a bound on the tail index is violated, especially with the bound set at 4 (i.e., B_4^{NQE}). Also, we can see in Table 22 in Appendix A, that bounding the tail index at four also works when the actual tail index is smaller than four.

For the interest of the length of the chapter, the results under other underlying distributions are attached in Tables 22 to 29 in Appendix A for reference. A few notable results from those distributions are that, although Clauset et al. [29] concentrated on the case $2 < \alpha < 3$ in estimating the tail index, the performance of the NQE_{Par} approach actually improves as the tail index of the true service time distribution increases. With the larger tail index, the relative error of the overall approaches are smaller. The MRE also tends to be smaller for the mixture distributions, except when the underlying distribution is 0.75 Unif(0, 10) + 0.25 Par(10, 4), which may be due to the fact that the number of data points to estimate the Pareto tail, which is only 25, is not large enough.

Although B_4^{NQE} does not perform the best in each underlying distribution, it is usually highly ranked, with no significant difference with the best approach. It is noteworthy that even the uniform distribution, which is bounded, can be approximated well with an unbounded distribution without significant loss of accuracy compared to the empirical distribution. Therefore, these numerical results suggest that the best overall approach is to first estimate the tail index for the Pareto distribution using the approach in Section 5.2.2, but restrict the tail index α to be greater than 4. When the estimated tail index is less than 4, we will use the approach in Section 5.2.3 instead, which is to use the quasi-empirical distribution with the exponential tail and cutoff

Table 16: Mean Relative Error: Pareto (10, 4), $\rho = 0.75$, $n = 100$

	MRE	$\mathbb{E}[W]$ $= 30$	$\mathbb{P}\{W \geq 100\}$ ≈ 0.060	$\mathbb{P}\{W \geq 120\}$ ≈ 0.039	$\mathbb{P}\{W \geq 140\}$ ≈ 0.025
Theoretical	Fitted Exp	0.524	1.35	1.65	1.08
	Fitted Par	0.306	0.609	0.736	0.655
	Empirical	0.315	0.697	0.842	0.845
Quasi-Empirical	NQE_{exp}	0.290	0.655	0.776	0.769
	NQE_{Par}	0.412	0.694	0.858	0.776
	CQE_{exp}	0.303	0.675	0.803	0.799
	CQE_{Par}	0.418	0.796	0.840	0.905
Hypothesis Test	$HT_{0.1}^{emp}$	0.421	0.716	0.887	0.800
	$HT_{0.1}^{NQE}$	0.415	0.702	0.869	0.790
	$HT_{0.5}^{emp}$	0.448	0.784	0.983	0.901
	$HT_{0.5}^{NQE}$	0.429	0.747	0.928	0.853
	$HT_{0.9}^{emp}$	0.398	0.756	0.938	0.910
	$HT_{0.9}^{NQE}$	0.373	0.714	0.874	0.843
Likelihood Ratio	LR_{exp}	0.291	0.659	0.779	0.767
	LR_{Par}	0.282	0.627	0.734	0.716
	LR_{cons}	0.291	0.659	0.779	0.767
Heuristics	KS	0.376	0.683	0.835	0.762
	B_3^{emp}	0.379	0.676	0.830	0.747
	B_3^{NQE}	0.379	0.677	0.830	0.747
	B_4^{emp}	0.283	0.606	0.723	0.751
	B_4^{NQE}	0.268	0.575	0.682	0.734
	$HT_{0.1}^{emp} + B_3^{emp}$	0.388	0.698	0.859	0.771
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.381	0.684	0.841	0.762
	$HT_{0.5}^{emp} + B_3^{emp}$	0.414	0.766	0.955	0.872
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.395	0.729	0.900	0.824
	$HT_{0.9}^{emp} + B_3^{emp}$	0.398	0.756	0.938	0.910
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.373	0.714	0.874	0.843
	$LR_{exp} + B_3^{NQE}$	0.291	0.659	0.779	0.767
	$LR_{Par} + B_3^{NQE}$	0.282	0.627	0.734	0.716
	$LR_{cons} + B_3^{NQE}$	0.291	0.659	0.779	0.767
	$KS + B_3^{NQE}$	0.376	0.683	0.835	0.762

estimated using the KS statistic. Other competitive approaches for $n = 100$ are using empirical distributions, CQE_{exp}, or B_4^{emp}.

In Table 17, we show the difference in mean relative error of B_4^{NQE} and the best approach for all four performance measures and each distribution, if there is an approach better than B_4^{NQE}. Let $p_i = \mathbb{P}\{W \geq t_i\}$, for $i = 1, 2, 3$. We can see that there is not a single approach that is always the best approach, and also that the difference between the MRE of the B_4^{NQE} and the best approach is not very significant in most cases. Also, it may be worth mentioning that although CQE_{Par} appears the most often as the best approach, we do not recommend using it, as there are cases when its performance is unacceptably poor. For example, when the true service time distribution is an exponential and $\rho = 0.5$, the MRE of CQE_{Par} is 99.0, while the best approach (Exp) has an MRE of 0.232, and B_4^{NQE} has an MRE of 0.244. In general, most approaches (other than B_4^{NQE}) have worst-case performances that are unacceptably poor.

Table 17: Comparison Between B_4^{NQE} and the Best Approach for Each Distribution, $n = 100$

MRE Difference	Exp(1) $\rho = 0.5$	Exp(1) $\rho = 0.75$	Par(10,3.5)	Par(10, 4)
$\mathbf{E}[W]$	0.012 (Exp)	0.016 (Emp)	0.266 (Exp)	0.014 (LR_{Par} + B_3^{NQE})
p_1	0.027 (Exp)	0.144 (CQE_{Par})	0.192 (LR_{Par} + B_3^{NQE})	0
p_2	0.069 (Exp)	0.17 (CQE_{Par})	0.325 (LR_{Par} + B_3^{NQE})	0
p_3	0.054 (Exp)	0.24 (Emp)	0.578 (CQE_{Par})	0.079 (Par)
	Par(10, 5)	Par(10, 10)	Unif(0, 1)	0.25 Unif(0, 10) + 0.75 Exp(1)
$\mathbf{E}[W]$	0.004 (CQE_{Par})	0	0.001 ($HT_{0.5}^{emp}$)	0.049 (CQE_{Par})
p_1	0.001 (Par)	0	0	0.129 (CQE_{Par})
p_2	0.004 (Par)	0.002 (Par)	0	0.212 (CQE_{Par})
p_3	0.032 (Emp)	0.003 (Par)	0.038 (CQE_{Par})	0.312 (CQE_{Par})
	0.75 Unif(0, 10) + 0.25 Exp(1)	0.25 Unif(0, 10) + 0.75 Par(10, 4)	0.75 Unif(0, 10) + 0.25 Par(10, 4)	
$\mathbf{E}[W]$	0.039 (CQE_{Par})	0	0.001 (NQE_{exp})	
p_1	0.028 (CQE_{Par})	0.012 (LR_{Par})	0	
p_2	0.096 (CQE_{Par})	0.032 (LR_{Par})	0	
p_3	0.248 (CQE_{Par})	0.177 (CQE_{Par})	0	

5.3.2 More Observed Service Times

In this section, we look into cases where we have more observations of the service times than in Section 5.3.1, namely $n = 500$ or $n = 1000$. For comparison, Tables 18 and 19 show the MRE for the exponential distribution with $\rho = 0.5$ when the sample sizes are 500 and 1000, respectively. As there are more data points to fit and evaluate the distributions, the performances of all the approaches under all the distributions generally improve as the sample sizes increase. Also, the differences of relative errors among approaches become less significant. Still, the general conclusions that there is no single approach that works the best in all situations, and that B_4^{NQE} generally performs well, remain valid. It is worth noting that the worst-case results for B_4^{NQE} occur for the Par(10, 3.5) distribution when $n = 1000$. This is not surprising as B_4^{NQE} is designed to avoid underestimating the tail index, but in this case it leads to rejecting the Pareto distribution when there is a significant amount of input data and the tail index is estimated correctly. With more data points, underestimation of α is less likely and there are other approaches that become competitive, including NQE_{Par}, KS, and $KS + B_3^{NQE}$. In Appendix A, we attach the results from all the other distributions when the sample size is 500 (Tables 30 to 39) or 1000 (Tables 40 to 49).

As in Section 5.3.1, we compare the MREs of the recommended method, B_4^{NQE} and the best approach, if it performs better than the B_4^{NQE}, for $n = 500$, 1000 in Tables 20 and 21, respectively. Clearly, the discrepancy between the MRE of the best approach and B_4^{NQE} decreases as n increases, in general. Also, as expected, when the true service time distribution is one of the theoretical distributions, the best approach is the actual distribution, especially when n is large. However, even in such cases, other than for the smallest tail probabilities, B_4^{NQE} shows comparable performance.

Table 18: Mean Relative Error: Exp(1), $\rho = 0.5$, $n = 500$

	Mean Relative Error	$\mathbb{E}[W] = 1$	$\mathbb{P}\{W \geq 5\} \approx 0.0410$	$\mathbb{P}\{W \geq 8\} \approx 0.0092$	$\mathbb{P}\{W \geq 10\} \approx 0.0034$
Theoretical	Fitted Exp	0.113	0.223	0.335	0.411
	Fitted Par	6.63E+84	23.4	108	297
	Empirical	0.118	0.249	0.401	0.511
Quasi-Empirical	NQE_{exp}	0.114	0.232	0.352	0.435
	NQE_{Par}	16.2	0.756	2.65	6.00
	CQE_{exp}	0.118	0.248	0.409	0.541
	CQE_{Par}	0.196	0.332	0.424	0.459
Hypothesis Test	$HT_{0.1}^{emp}$	16.2	0.756	2.65	6.00
	$HT_{0.1}^{NQE}$	16.2	0.756	2.65	6.00
	$HT_{0.5}^{emp}$	16.2	0.756	2.65	6.00
	$HT_{0.5}^{NQE}$	16.2	0.756	2.65	6.00
	$HT_{0.9}^{emp}$	15.4	0.710	2.46	5.53
	$HT_{0.9}^{NQE}$	15.4	0.710	2.46	5.53
Likelihood Ratio	LR_{exp}	0.114	0.232	0.352	0.435
	LR_{Par}	0.114	0.232	0.352	0.435
	LR_{cons}	0.114	0.232	0.352	0.435
Heuristics	KS	0.114	0.232	0.352	0.435
	B_3^{emp}	0.418	0.491	1.54	3.22
	B_3^{NQE}	0.417	0.489	1.54	3.22
	B_4^{emp}	0.127	0.260	0.511	0.823
	B_4^{NQE}	0.127	0.257	0.497	0.798
	$HT_{0.1}^{emp} + B_3^{emp}$	0.418	0.491	1.54	3.22
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.417	0.489	1.54	3.22
	$HT_{0.5}^{emp} + B_3^{emp}$	0.418	0.491	1.54	3.22
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.417	0.489	1.54	3.22
	$HT_{0.9}^{emp} + B_3^{emp}$	0.411	0.483	1.51	3.15
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.411	0.481	1.51	3.16
	$LR_{exp} + B_3^{NQE}$	0.114	0.232	0.352	0.435
	$LR_{Par} + B_3^{NQE}$	0.114	0.232	0.352	0.435
	$LR_{cons} + B_3^{NQE}$	0.114	0.232	0.352	0.435
	$KS + B_3^{NQE}$	0.114	0.232	0.352	0.435

Table 19: Mean Relative Error: Exp(1), $\rho = 0.5$, $n = 1000$

	Mean Relative Error	$\mathbb{E}[W]$ $= 1$	$\mathbb{P}\{W \geq 5\}$ ≈ 0.0410	$\mathbb{P}\{W \geq 8\}$ ≈ 0.0092	$\mathbb{P}\{W \geq 10\}$ ≈ 0.0034
Theoretical	Fitted Exp	0.078	0.16	0.24	0.29
	Fitted Par	1.38E+79	23.4	108	297
	Empirical	0.083	0.18	0.30	0.38
Quasi-Empirical	NQE_{exp}	0.080	0.16	0.25	0.31
	NQE_{Par}	0.349	0.37	1.18	2.44
	CQE_{exp}	0.083	0.183	0.298	0.390
	CQE_{Par}	0.102	0.248	0.355	0.398
Hypothesis Test	$HT_{0.1}^{emp}$	0.349	0.374	1.18	2.44
	$HT_{0.1}^{NQE}$	0.349	0.374	1.18	2.44
	$HT_{0.5}^{emp}$	0.349	0.374	1.18	2.44
	$HT_{0.5}^{NQE}$	0.349	0.374	1.18	2.44
	$HT_{0.9}^{emp}$	0.349	0.374	1.18	2.44
	$HT_{0.9}^{NQE}$	0.349	0.374	1.18	2.44
Likelihood Ratio	LR_{exp}	0.080	0.164	0.252	0.313
	LR_{Par}	0.080	0.164	0.252	0.312
	LR_{cons}	0.080	0.164	0.252	0.313
Heuristics	KS	0.080	0.164	0.252	0.313
	B_3^{emp}	0.317	0.361	1.13	2.31
	B_3^{NQE}	0.316	0.360	1.13	2.31
	B_4^{emp}	0.102	0.210	0.507	0.908
	B_4^{NQE}	0.101	0.204	0.496	0.892
	$HT_{0.1}^{emp} + B_3^{emp}$	0.317	0.361	1.13	2.31
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.316	0.360	1.13	2.31
	$HT_{0.5}^{emp} + B_3^{emp}$	0.317	0.361	1.13	2.31
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.316	0.360	1.13	2.31
	$HT_{0.9}^{emp} + B_3^{emp}$	0.317	0.361	1.13	2.31
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.317	0.360	1.13	2.31
	$LR_{exp} + B_3^{NQE}$	0.080	0.164	0.252	0.313
	$LR_{Par} + B_3^{NQE}$	0.080	0.164	0.252	0.312
	$LR_{cons} + B_3^{NQE}$	0.080	0.164	0.252	0.313
	$KS + B_3^{NQE}$	0.080	0.164	0.252	0.313

Table 20: Comparison Between B_4^{NQE} and the Best Approach for Each Distribution, $n = 500$

MRE Difference	Exp(1) $\rho = 0.5$	Exp(1) $\rho = 0.75$	Par(10,3.5)	Par(10, 4)
$\mathbf{E}[W]$	0.014 (Exp)	0.016 (Exp)	0.111 (Exp)	0.011 (Par)
p_1	0.034 (Exp)	0.038 (Exp)	0.278 (Par)	0.067 (Par)
p_2	0.162 (Exp)	0.165 (CQE_{Par})	0.354 (Par)	0.095 (Par)
p_3	0.387 (Exp)	0.494 (Exp)	0.419 (Exp)	0.124 (Par)
	Par(10, 5)	Par(10, 10)	Unif(0, 1)	0.25 Unif(0, 10) + 0.75 Exp(1)
$\mathbf{E}[W]$	0	0	0.002 (Emp)	0.019 (CQE_{Par})
p_1	0	0	0	0.052 (CQE_{Par})
p_2	0.002 (Par)	0	0	0.088 (CQE_{Par})
p_3	0.007 (Par)	0	0.013 (Emp)	0.134 (CQE_{Par})
	0.75 Unif(0, 10) + 0.25 Exp(1)	0.25 Unif(0, 10) + 0.75 Par(10, 4)	0.75 Unif(0, 10) + 0.25 Par(10, 4)	
$\mathbf{E}[W]$	0.022 (CQE_{Par})	0.036 (CQE_{Par})	0.057 (CQE_{Par})	
p_1	0.014 (CQE_{Par})	0	0	
p_2	0.050 (CQE_{Par})	0	0	
p_3	0.157 (CQE_{Par})	0	0	

Table 21: Comparison Between B_4^{NQE} and the Best Approach for Each Distribution, $n = 1000$

MRE Difference	Exp(1) $\rho = 0.5$	Exp(1) $\rho = 0.75$	Par(10,3.5)	Par(10, 4)
$\mathbf{E}[W]$	0.023 (Exp)	0.010 (Exp)	0.642 (Par)	0.012 (Par)
p_1	0.049 (Exp)	0.027 (Exp)	0.369 (Par)	0.051 (Par)
p_2	0.259 (Exp)	0.117 (Exp)	0.579 (Par)	0.070 (Par)
p_3	0.600 (Exp)	0.313 (Exp)	0.807 (Par)	0.093 (Par)
	Par(10, 5)	Par(10, 10)	Unif(0, 1)	0.25 Unif(0, 10) + 0.75 Exp(1)
$\mathbf{E}[W]$	0.002 (Par)	0	0	0.012 (CQE_{Par})
p_1	0.005 (Par)	0	0	0.033 (CQE_{Par})
p_2	0.009 (Par)	0	0.001 (KS)	0.058 (CQE_{Par})
p_3	0.018 (Par)	0	0.004 (Emp)	0.088 (CQE_{Par})
	0.75 Unif(0, 10) + 0.25 Exp(1)	0.25 Unif(0, 10) + 0.75 Par(10, 4)	0.75 Unif(0, 10) + 0.25 Par(10, 4)	
$\mathbf{E}[W]$	0.014 (CQE_{Par})	0.030 (CQE_{Par})	0.079 (CQE_{Par})	
p_1	0.009 (CQE_{Par})	0.003 (NQE_{exp})	0.038 (CQE_{Par})	
p_2	0.031 (CQE_{Par})	0	0	
p_3	0.109 (CQE_{Par})	0	0	

5.3.3 Summary and Recommendation

In Sections 5.3.1 and 5.3.2, we presented simulation results for the steady-state waiting time of the M/G/1 queue, when the interarrival distribution is exponentially distributed with a known parameter, and the service time distribution is not known, but observed data is available. We applied all the approaches we discussed in Section 5.2 to estimate the mean and tail probabilities of the waiting time W, for eleven different distributions. In Tables 17, 20, and 21, we can see the best approaches under all the experiments considered.

Clearly, no approach prevails as the best in all cases, but we conclude that B_4^{NQE} is a reasonable approach to be used in most situations, especially if only limited amounts of input data are available, as the discrepancies are fairly small in all the cases, and its MREs are always among the smallest ones, if not the smallest. Although CQE_{Par} appears the most often as the best approach, this approach has shown poor performance under certain experiment settings, and thus cannot be recommended. The same is true of the other approaches in that their worst-case performances are generally not comparable to the best approach or B_4^{NQE}. However, if there is ample input data, it is reasonable to have more confidence in the accuracy of small tail indices, and in such cases, there are other competitive approaches including NQE_{Par}, KS, and $KS + B_3^{NQE}$.

5.4 Conclusion

In this chapter, we discussed different approaches to fit a distribution to sample data, when theoretical distributions may not provide good fit, and applied these approaches to simulate the waiting time process of the M/G/1 queueing system. The main approaches that we consider are quasi-empirical distributions, which are the mixtures of empirical and right-tail distributions. We discussed quasi-empirical distributions, both with exponential and Pareto tails, and with or without estimating the cutoff point between the empirical and tail distributions. Then, we presented some approaches to decide on the form of the tail distribution or to compare the goodness-of-fit of different fitted tails. Using each approach, we estimated the mean and tail probabilities of the waiting time process, focusing on performance measures that are not "too rare."

As seen from the results, unfortunately there is no magic solution that works the best in all situations. However, the best overall results are achieved by B_4^{NQE}, which is the approach that uses a quasi-empirical distribution with a Pareto tail if the estimated tail index is greater than four, and otherwise uses a quasi-empirical distribution with an exponential tail. The motivation to use this approach is that while we recognize the need to use non-exponentially decaying distributions in certain situations, we do not allow the tail to be extremely heavy, such that the second moment is infinite or significantly large, unless large amounts of input data justify such a choice. This is appropriate in the domain we consider, as we know that the mean waiting time of an M/G/1 system depends heavily on the second moment. Although we have not experimented with other problem domains, the volatility of the waiting time process for small tail indices arises due to the heavy tail of the service time distribution itself, and hence we expect that underestimation of the tail index will also be detrimental in other fields. Finally, our recommendation is valid for different number of observations of the underlying distribution, but the mean relative error decreases overall as the number of samples increases, and certain other approaches that do not exclude small tail indices are also worth considering for large amounts of input data.

This study can be fortified by examining more extensive cases, such as the number of samples n, underlying distributions, traffic intensities, performance measures, number of macro replications, etc. We chose an M/G/1 queueing system as it is a simple system that is representative of other systems studied via discrete-event simulation. However, it would be desirable to apply the approach to other domains in future research. Moreover, as we are recommending to impose a bound on the tail index, the development of a scientific approach for selecting the value of this lower bound is crucial (e.g., smaller lower bounds may be appropriate when there is ample input data). Finally, the incorporation of the approaches we discussed in widely-used distribution-fitting packages would be valuable.

CHAPTER VI

CONCLUSION

This thesis advances the simulation field by contributing three valid and efficient simulation algorithms. In this chapter, we briefly state the main contributions of each algorithm in Section 6.1, and discuss future work that enhances the proposed methods in Section 6.2.

6.1 Contribution

In Chapter 3, we proposed an algorithm that improves the efficiency of transient simulation by using cloning. While cloning was originally designed to share some simulation results among sample paths for different scenarios, our approach shares simulation results among different replications of the same system. We presented our algorithm and identified the number of clones that optimizes its efficiency. Then, to improve performance, we introduced cloning algorithms with induced negative correlation. Finally, we supplied numerical results that support the efficiency of the algorithm and provided insights about its sensitivity to the choices of the number of clones and position of the splitting point.

In Chapter 4, we proposed a procedure that considers optimization problems with multiple objectives and estimates the Pareto set using Ranking and Selection (a Pareto set is a set of systems that are not dominated in all objectives). Previously, Ranking and Selection was geared toward the single objective case, possibly with constraints. Our procedure is designed for cases with multiple objectives when we do not have prior information on the importance of the objectives. We proved the validity of the procedure under three formulations. In addition, we proposed configurations to test the validity and performance of our procedure for multiple objectives, and provided numerical results that confirm the validity of our approach and address its efficiency.

In Chapter 5, we discussed different approaches to fit a distribution to sample data when theoretical distributions may not provide good fit, and applied these approaches to simulate the waiting time process of an M/G/1 queueing system. Our focus is on quasi-empirical distributions, which are mixtures of empirical and right-tail distributions. We also discuss the choice of both the right-tail distribution and the cutoff between the empirical and right-tail distributions, and finally provide a comparison among the approaches in simulation experiments. Numerical results show that while no approach performs the best in all cases, bounding the tail index below seems to achieve the best overall performance.

6.2 Future Research

In this section, we discuss how the subjects presented in this thesis can be enhanced with further research on the following topics:

1. In Chapter 3, we can improve the algorithm by finding the optimal number and position of decision points and by devising a better way of estimating the number of clones to use in the simulation.

2. In Chapter 4, the lower bounds for each formulation can be improved when independence between systems and/or objectives is assumed.

3. In Chapter 5, the development of a scientific approach for selecting the optimal value of the lower bound on the estimated tail index of a Pareto distribution can improve the approach.

APPENDIX A

ADDITIONAL NUMERICAL RESULTS FOR CHAPTER 5

For completeness, we provide the mean relative error for all the approaches, distributions, and performance measures. Sections A.1, A.2, and A.3 provide the results for the different sample sizes, $n = 100, 500$, and 1000, respectively.

A.1 Numerical Results: One Hundred Observed Service Times

In this section, we provide the results for $n = 100$ that are not presented in Section 5.3.1. See Tables 22 to 29.

A.2 Numerical Results: Five Hundred Observed Service Times

In this section, we provide detailed results when $n = 500$, see Tables 30 to 39.

Table 22: Mean Relative Error: Par(10,3.5), $\rho = 0.75$, $n = 100$

	Mean Relative Error	$\mathbb{E}[W]$ $= 45$	$\mathbb{P}\{W \geq 150\}$ ≈ 0.056	$\mathbb{P}\{W \geq 200\}$ ≈ 0.032	$\mathbb{P}\{W \geq 250\}$ ≈ 0.020
Theoretical	Fitted Exp	0.366	0.871	1.01	1.13
	Fitted Par	0.996	0.882	1.15	1.41
	Empirical	0.698	1.03	1.31	1.58
Quasi-Empirical	NQE_{exp}	0.632	0.957	1.24	1.51
	NQE_{Par}	0.717	0.920	1.21	1.49
	CQE_{exp}	0.566	0.963	1.22	1.46
	CQE_{Par}	0.578	0.904	0.919	0.922
Hypothesis Test	$HT_{0.1}^{emp}$	0.918	1.03	1.41	1.80
	$HT_{0.1}^{NQE}$	0.884	1.03	1.40	1.78
	$HT_{0.5}^{emp}$	0.981	1.12	1.54	1.97
	$HT_{0.5}^{NQE}$	0.912	1.08	1.48	1.89
	$HT_{0.9}^{emp}$	0.885	1.05	1.40	1.77
	$HT_{0.9}^{NQE}$	0.811	0.988	1.34	1.69
Likelihood Ratio	LR_{exp}	0.633	0.960	1.24	1.50
	LR_{Par}	0.458	0.783	0.950	1.07
	LR_{cons}	0.633	0.960	1.24	1.50
Heuristics	KS	0.734	0.972	1.27	1.54
	B_3^{emp}	0.605	0.866	1.12	1.34
	B_3^{NQE}	0.605	0.865	1.12	1.34
	B_4^{emp}	0.695	1.02	1.30	1.58
	B_4^{NQE}	0.632	0.955	1.24	1.50
	$HT_{0.1}^{emp} + B_3^{emp}$	0.807	0.979	1.31	1.65
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.772	0.970	1.30	1.63
	$HT_{0.5}^{emp} + B_3^{emp}$	0.869	1.07	1.44	1.82
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.800	1.02	1.38	1.74
	$HT_{0.9}^{emp} + B_3^{emp}$	0.773	1.00	1.30	1.62
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.699	0.933	1.24	1.54
	$LR_{exp} + B_3^{NQE}$	0.633	0.960	1.24	1.50
	$LR_{Par} + B_3^{NQE}$	0.430	0.763	0.915	1.02
	$LR_{cons} + B_3^{NQE}$	0.633	0.960	1.24	1.50
	$KS + B_3^{NQE}$	0.622	0.917	1.17	1.40

Table 23: Mean Relative Error: Pareto (10, 5), $\rho \simeq 0.67$, $n = 100$

	Mean Relative Error	$\mathbb{E}[W]$ $= 15$	$\mathbb{P}\{W \geq 50\}$ ≈ 0.0656	$\mathbb{P}\{W \geq 70\}$ ≈ 0.0262	$\mathbb{P}\{W \geq 100\}$ ≈ 0.0070
Theoretical	Fitted Exp	0.780	1.91	3.42	6.92
	Fitted Par	0.127	0.278	0.404	0.624
	Empirical	0.124	0.284	0.413	0.604
Quasi-Empirical	NQE_{exp}	0.229	0.548	0.891	1.59
	NQE_{Par}	0.134	0.289	0.432	0.705
	CQE_{exp}	0.123	0.284	0.407	0.593
	CQE_{Par}	0.230	0.499	0.623	0.752
Hypothesis Test	$HT_{0.1}^{emp}$	0.140	0.303	0.457	0.753
	$HT_{0.1}^{NQE}$	0.147	0.319	0.486	0.816
	$HT_{0.5}^{emp}$	0.142	0.313	0.472	0.764
	$HT_{0.5}^{NQE}$	0.175	0.394	0.620	1.08
	$HT_{0.9}^{emp}$	0.137	0.304	0.457	0.728
	$HT_{0.9}^{NQE}$	0.212	0.492	0.800	1.44
Likelihood Ratio	LR_{exp}	0.220	0.529	0.856	1.52
	LR_{Par}	0.217	0.521	0.842	1.48
	LR_{cons}	0.220	0.529	0.856	1.52
Heuristics	KS	0.146	0.328	0.492	0.793
	B_3^{emp}	0.134	0.289	0.432	0.705
	B_3^{NQE}	0.134	0.289	0.432	0.705
	B_4^{emp}	0.127	0.281	0.412	0.643
	B_4^{NQE}	0.127	0.279	0.408	0.636
	$HT_{0.1}^{emp} + B_3^{emp}$	0.140	0.303	0.457	0.753
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.147	0.319	0.486	0.816
	$HT_{0.5}^{emp} + B_3^{emp}$	0.142	0.313	0.472	0.764
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.175	0.394	0.620	1.08
	$HT_{0.9}^{emp} + B_3^{emp}$	0.137	0.304	0.457	0.728
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.212	0.492	0.800	1.44
	$LR_{exp} + B_3^{NQE}$	0.220	0.529	0.856	1.52
	$LR_{Par} + B_3^{NQE}$	0.217	0.521	0.842	1.48
	$LR_{cons} + B_3^{NQE}$	0.220	0.529	0.856	1.515
	$KS + B_3^{NQE}$	0.146	0.328	0.492	0.793

Table 24: Mean Relative Error: Pareto (10, 10), $\rho = 0.56$, $n = 100$

	MRE	$\mathbb{E}[W]$ $= 7.35$	$\mathbb{P}\{W \geq 10\}$ ≈ 0.2787	$\mathbb{P}\{W \geq 30\}$ ≈ 0.0444	$\mathbb{P}\{W \geq 50\}$ ≈ 0.0071
Theoretical	Fitted Exp	0.963	0.365	2.93	10.3
	Fitted Par	0.033	0.031	0.083	0.138
	Empirical	0.034	0.031	0.084	0.139
Quasi-Empirical	NQE_{exp}	0.245	0.107	0.734	2.46
	NQE_{Par}	0.034	0.031	0.085	0.141
	CQE_{exp}	0.034	0.031	0.085	0.140
	CQE_{Par}	0.066	0.061	0.161	0.249
Hypothesis Test	$HT_{0.1}^{emp}$	0.035	0.032	0.087	0.146
	$HT_{0.1}^{NQE}$	0.062	0.041	0.168	0.439
	$HT_{0.5}^{emp}$	0.034	0.031	0.086	0.144
	$HT_{0.5}^{NQE}$	0.145	0.072	0.427	1.36
	$HT_{0.9}^{emp}$	0.034	0.031	0.087	0.144
	$HT_{0.9}^{NQE}$	0.197	0.090	0.585	1.93
Likelihood Ratio	LR_{exp}	0.236	0.103	0.706	2.36
	LR_{Par}	0.236	0.103	0.706	2.36
	LR_{cons}	0.236	0.103	0.706	2.36
Heuristics	KS	0.117	0.061	0.342	1.06
	B_3^{emp}	0.034	0.031	0.085	0.141
	B_3^{NQE}	0.034	0.031	0.085	0.141
	B_4^{emp}	0.034	0.031	0.085	0.141
	B_4^{NQE}	0.034	0.031	0.085	0.141
	$HT_{0.1}^{emp} + B_3^{emp}$	0.035	0.032	0.087	0.146
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.062	0.041	0.168	0.439
	$HT_{0.5}^{emp} + B_3^{emp}$	0.034	0.031	0.086	0.144
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.145	0.072	0.427	1.36
	$HT_{0.9}^{emp} + B_3^{emp}$	0.034	0.031	0.087	0.144
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.197	0.090	0.585	1.93
	$LR_{exp} + B_3^{NQE}$	0.236	0.103	0.706	2.36
	$LR_{Par} + B_3^{NQE}$	0.236	0.103	0.706	2.36
	$LR_{cons} + B_3^{NQE}$	0.236	0.103	0.706	2.36
	$KS + B_3^{NQE}$	0.117	0.061	0.342	1.06

Table 25: Mean Relative Error: Uniform (0,1), $\rho = 0.25$, $n = 100$

	MRE	$\mathbb{E}[W]$ $= 0.11$	$\mathbb{P}\{W \geq 0.05\}$ ≈ 0.2135	$\mathbb{P}\{W \geq 0.1\}$ ≈ 0.0919	$\mathbb{P}\{W \geq 0.5\}$ ≈ 0.0165
Theoretical	Fitted Exp	0.505	0.054	0.287	2.394
	Fitted Par	4.13E+73	3.68	9.88	59.7
	Empirical	0.082	0.053	0.096	0.167
Quasi-Empirical	NQE_{exp}	0.112	0.053	0.103	0.468
	NQE_{Par}	0.131	0.053	0.098	0.262
	CQE_{exp}	0.082	0.053	0.096	0.166
	CQE_{Par}	0.085	0.055	0.105	0.172
Hypothesis Test	$HT_{0.1}^{emp}$	0.131	0.053	0.098	0.264
	$HT_{0.1}^{NQE}$	0.139	0.053	0.101	0.333
	$HT_{0.5}^{emp}$	0.081	0.053	0.095	0.170
	$HT_{0.5}^{NQE}$	0.094	0.053	0.095	0.368
	$HT_{0.9}^{emp}$	0.082	0.053	0.096	0.167
	$HT_{0.9}^{NQE}$	0.110	0.053	0.102	0.448
Likelihood Ratio	LR_{exp}	0.112	0.053	0.103	0.468
	LR_{Par}	0.112	0.053	0.103	0.468
	LR_{cons}	0.112	0.053	0.103	0.468
Heuristics	KS	0.095	0.053	0.096	0.354
	B_3^{emp}	0.131	0.053	0.098	0.262
	B_3^{NQE}	0.131	0.053	0.098	0.262
	B_4^{emp}	0.082	0.052	0.094	0.186
	B_4^{NQE}	0.082	0.052	0.093	0.204
	$HT_{0.1}^{emp} + B_3^{emp}$	0.131	0.053	0.098	0.264
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.139	0.053	0.101	0.333
	$HT_{0.5}^{emp} + B_3^{emp}$	0.081	0.053	0.095	0.170
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.094	0.053	0.095	0.368
	$HT_{0.9}^{emp} + B_3^{emp}$	0.082	0.053	0.096	0.167
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.110	0.053	0.102	0.448
	$LR_{exp} + B_3^{NQE}$	0.112	0.053	0.103	0.468
	$LR_{Par} + B_3^{NQE}$	0.112	0.053	0.103	0.468
	$LR_{cons} + B_3^{NQE}$	0.112	0.053	0.103	0.468
	$KS + B_3^{NQE}$	0.095	0.053	0.096	0.354

Table 26: Mean Relative Error: 0.25 Uniform (0,10) + 0.75 Exp(1), $\rho = 0.475$, $n = 100$

	MRE	$\mathbb{E}[W]$ $= 4.0235$	$\mathbb{P}\{W \geq 20\}$ ≈ 0.0377	$\mathbb{P}\{W \geq 30\}$ ≈ 0.0095	$\mathbb{P}\{W \geq 40\}$ ≈ 0.0024
Theoretical	Fitted Exp	1.134	3.17	8.52	20.8
	Fitted Par	2.23E+71	25.5	104	418
	Empirical	0.179	0.351	0.532	0.736
Quasi-Empirical	NQE_{exp}	0.182	0.366	0.558	0.778
	NQE_{Par}	0.181	0.359	0.547	0.761
	CQE_{exp}	0.179	0.352	0.533	0.739
	CQE_{Par}	0.132	0.230	0.335	0.449
Hypothesis Test	$HT_{0.1}^{emp}$	0.181	0.359	0.547	0.761
	$HT_{0.1}^{NQE}$	0.181	0.359	0.547	0.761
	$HT_{0.5}^{emp}$	0.181	0.359	0.547	0.761
	$HT_{0.5}^{NQE}$	0.181	0.359	0.547	0.761
	$HT_{0.9}^{emp}$	0.181	0.357	0.543	0.755
	$HT_{0.9}^{NQE}$	0.181	0.359	0.546	0.760
Likelihood Ratio	LR_{exp}	0.182	0.366	0.558	0.778
	LR_{Par}	0.182	0.366	0.558	0.778
	LR_{cons}	0.182	0.366	0.558	0.778
Heuristics	KS	0.181	0.361	0.550	0.766
	B_3^{emp}	0.181	0.359	0.547	0.761
	B_3^{NQE}	0.181	0.359	0.547	0.761
	B_4^{emp}	0.181	0.359	0.547	0.761
	B_4^{NQE}	0.181	0.359	0.547	0.761
	$HT_{0.1}^{emp} + B_3^{emp}$	0.181	0.359	0.547	0.761
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.181	0.359	0.547	0.761
	$HT_{0.5}^{emp} + B_3^{emp}$	0.181	0.359	0.547	0.761
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.181	0.359	0.547	0.761
	$HT_{0.9}^{emp} + B_3^{emp}$	0.181	0.357	0.543	0.755
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.181	0.359	0.546	0.760
	$LR_{exp} + B_3^{NQE}$	0.182	0.366	0.558	0.778
	$LR_{Par} + B_3^{NQE}$	0.182	0.366	0.558	0.778
	$LR_{cons} + B_3^{NQE}$	0.182	0.366	0.558	0.778
	$KS + B_3^{NQE}$	0.181	0.361	0.550	0.766

Table 27: Mean Relative Error: 0.75 Uniform (0,10) + 0.25 Exp(1), $\rho = 0.325$, $n = 100$

	MRE	$\mathbb{E}[W]$ $= 1.5431$	$\mathbb{P}\{W \geq 5\}$ ≈ 0.1259	$\mathbb{P}\{W \geq 10\}$ ≈ 0.0412	$\mathbb{P}\{W \geq 20\}$ ≈ 0.0035
Theoretical	Fitted Exp	1.025	0.532	1.79	10.5
	Fitted Par	4.54E+75	6.94	23.3	282
	Empirical	0.326	0.345	0.442	1.01
Quasi-Empirical	NQE_{exp}	0.329	0.345	0.446	1.05
	NQE_{Par}	0.330	0.346	0.447	1.06
	CQE_{exp}	0.327	0.345	0.442	1.02
	CQE_{Par}	0.291	0.318	0.351	0.812
Hypothesis Test	$HT_{0.1}^{emp}$	0.330	0.346	0.447	1.06
	$HT_{0.1}^{NQE}$	0.330	0.346	0.447	1.06
	$HT_{0.5}^{emp}$	0.330	0.346	0.446	1.06
	$HT_{0.5}^{NQE}$	0.330	0.346	0.446	1.06
	$HT_{0.9}^{emp}$	0.328	0.345	0.444	1.03
	$HT_{0.9}^{NQE}$	0.329	0.345	0.446	1.05
Likelihood Ratio	LR_{exp}	0.329	0.345	0.446	1.05
	LR_{Par}	0.329	0.345	0.446	1.05
	LR_{cons}	0.329	0.345	0.446	1.05
Heuristics	KS	0.329	0.345	0.445	1.05
	B_3^{emp}	0.330	0.346	0.447	1.06
	B_3^{NQE}	0.330	0.346	0.447	1.06
	B_4^{emp}	0.330	0.346	0.447	1.06
	B_4^{NQE}	0.330	0.346	0.447	1.06
	$HT_{0.1}^{emp} + B_3^{emp}$	0.330	0.346	0.447	1.06
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.330	0.346	0.447	1.06
	$HT_{0.5}^{emp} + B_3^{emp}$	0.330	0.346	0.446	1.06
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.330	0.346	0.446	1.06
	$HT_{0.9}^{emp} + B_3^{emp}$	0.328	0.345	0.444	1.03
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.329	0.345	0.446	1.05
	$LR_{exp} + B_3^{NQE}$	0.329	0.345	0.446	1.05
	$LR_{Par} + B_3^{NQE}$	0.329	0.345	0.446	1.05
	$LR_{cons} + B_3^{NQE}$	0.329	0.345	0.446	1.05
	$KS + B_3^{NQE}$	0.329	0.345	0.445	1.05

Table 28: Mean Relative Error: 0.25 Uniform (0,10) + 0.75 Par(10,4), $\rho = 0.625$, $n = 100$

	MRE	$\mathbb{E}[W]$ $= 13.352$	$\mathbb{P}\{W \geq 50\}$ ≈ 0.0593	$\mathbb{P}\{W \geq 70\}$ ≈ 0.0283	$\mathbb{P}\{W \geq 90\}$ ≈ 0.0149
Theoretical	Fitted Exp	0.593	1.36	1.79	2.05
	Fitted Par	1.03E+76	15.9	34.3	66.1
	Empirical	0.538	0.705	1.084	1.53
Quasi-Empirical	NQE_{exp}	0.318	0.624	0.906	1.19
	NQE_{Par}	0.437	0.567	0.840	1.13
	CQE_{exp}	0.433	0.694	1.062	1.48
	CQE_{Par}	0.311	0.628	0.763	0.843
Hypothesis Test	$HT_{0.1}^{emp}$	0.433	0.565	0.841	1.13
	$HT_{0.1}^{NQE}$	0.433	0.565	0.841	1.13
	$HT_{0.5}^{emp}$	0.469	0.628	0.952	1.30
	$HT_{0.5}^{NQE}$	0.464	0.617	0.936	1.28
	$HT_{0.9}^{emp}$	0.526	0.686	1.054	1.47
	$HT_{0.9}^{NQE}$	0.411	0.652	0.983	1.32
Likelihood Ratio	LR_{exp}	0.321	0.626	0.911	1.20
	LR_{Par}	0.329	0.525	0.737	0.940
	LR_{cons}	0.321	0.626	0.911	1.20
Heuristics	KS	0.434	0.580	0.860	1.15
	B_3^{emp}	0.503	0.603	0.924	1.30
	B_3^{NQE}	0.396	0.593	0.881	1.18
	B_4^{emp}	0.507	0.620	0.959	1.38
	B_4^{NQE}	0.287	0.537	0.769	1.02
	$HT_{0.1}^{emp} + B_3^{emp}$	0.499	0.601	0.925	1.30
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.392	0.591	0.883	1.19
	$HT_{0.5}^{emp} + B_3^{emp}$	0.535	0.664	1.036	1.47
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.423	0.643	0.978	1.34
	$HT_{0.9}^{emp} + B_3^{emp}$	0.526	0.686	1.054	1.47
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.411	0.652	0.983	1.32
	$LR_{exp} + B_3^{NQE}$	0.321	0.626	0.911	1.20
	$LR_{Par} + B_3^{NQE}$	0.288	0.550	0.779	0.998
	$LR_{cons} + B_3^{NQE}$	0.321	0.626	0.911	1.20
	$KS + B_3^{NQE}$	0.393	0.606	0.902	1.20

Table 29: Mean Relative Error: 0.75 Uniform (0,10) + 0.25 Par(10,4), $\rho = 0.375$, $n = 100$

	MRE	$\mathbb{E}[W] = 3.3107$	$\mathbb{P}\{W \geq 20\} \approx 0.0384$	$\mathbb{P}\{W \geq 30\} \approx 0.0173$	$\mathbb{P}\{W \geq 40\} \approx 0.0091$
Theoretical	Fitted Exp	0.370	0.863	0.825	0.559
	Fitted Par	6.94E+68	25.1	57.0	109
	Empirical	0.286	0.485	0.836	1.23
Quasi-Empirical	NQE_{exp}	0.188	0.423	0.522	0.616
	NQE_{Par}	1.02	0.557	0.902	1.26
	CQE_{exp}	0.251	0.482	0.764	1.00
	CQE_{Par}	0.354	0.496	0.685	0.777
Hypothesis Test	$HT_{0.1}^{emp}$	1.02	0.557	0.902	1.26
	$HT_{0.1}^{NQE}$	1.02	0.557	0.902	1.26
	$HT_{0.5}^{emp}$	1.02	0.560	0.908	1.27
	$HT_{0.5}^{NQE}$	1.02	0.559	0.907	1.27
	$HT_{0.9}^{emp}$	0.628	0.518	0.843	1.21
	$HT_{0.9}^{NQE}$	0.588	0.494	0.746	1.00
Likelihood Ratio	LR_{exp}	0.188	0.423	0.522	0.616
	LR_{Par}	0.452	0.426	0.574	0.746
	LR_{cons}	0.188	0.423	0.522	0.616
Heuristics	KS	0.734	0.463	0.683	0.930
	B_3^{emp}	0.683	0.539	0.869	1.22
	B_3^{NQE}	0.652	0.525	0.814	1.10
	B_4^{emp}	0.282	0.467	0.770	1.12
	B_4^{NQE}	0.189	0.406	0.459	0.505
	$HT_{0.1}^{emp} + B_3^{emp}$	0.683	0.539	0.869	1.22
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.652	0.525	0.814	1.10
	$HT_{0.5}^{emp} + B_3^{emp}$	0.683	0.543	0.876	1.23
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.651	0.528	0.819	1.11
	$HT_{0.9}^{emp} + B_3^{emp}$	0.628	0.518	0.843	1.21
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.588	0.494	0.746	1.00
	$LR_{exp} + B_3^{NQE}$	0.188	0.423	0.522	0.616
	$LR_{Par} + B_3^{NQE}$	0.233	0.425	0.558	0.703
	$LR_{cons} + B_3^{NQE}$	0.188	0.423	0.522	0.616
	$KS + B_3^{NQE}$	0.475	0.454	0.645	0.845

Table 30: Mean Relative Error: Exp(1), $\rho = 0.75$, $n = 500$

	MRE	$\mathbb{E}[W]$ $= 3$	$\mathbb{P}\{W \geq 10\}$ ≈ 0.062	$\mathbb{P}\{W \geq 15\}$ ≈ 0.018	$\mathbb{P}\{W \geq 20\}$ ≈ 0.005
Theoretical	Fitted Exp	0.175	0.380	0.582	0.835
	Fitted Par	3.27E+69	15.2	55.7	197
	Empirical	0.178	0.392	0.597	0.849
Quasi-Empirical	NQE_{exp}	0.175	0.382	0.583	0.836
	NQE_{Par}	11.7	1.29	3.69	10.5
	CQE_{exp}	0.178	0.391	0.600	0.860
	CQE_{Par}	0.217	0.431	0.540	0.652
Hypothesis Test	$HT_{0.1}^{emp}$	11.7	1.29	3.69	10.5
	$HT_{0.1}^{NQE}$	11.7	1.29	3.69	10.5
	$HT_{0.5}^{emp}$	11.7	1.29	3.69	10.5
	$HT_{0.5}^{NQE}$	11.7	1.29	3.69	10.5
	$HT_{0.9}^{emp}$	6.09	1.19	3.34	9.35
	$HT_{0.9}^{NQE}$	6.09	1.19	3.34	9.35
Likelihood Ratio	LR_{exp}	0.175	0.382	0.583	0.836
	LR_{Par}	0.176	0.383	0.584	0.838
	LR_{cons}	0.175	0.382	0.583	0.836
Heuristics	KS	0.175	0.382	0.583	0.836
	B_3^{emp}	0.400	0.668	1.54	3.47
	B_3^{NQE}	0.399	0.666	1.54	3.47
	B_4^{emp}	0.193	0.425	0.718	1.17
	B_4^{NQE}	0.191	0.418	0.705	1.15
	$HT_{0.1}^{emp} + B_3^{emp}$	0.400	0.668	1.54	3.47
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.399	0.666	1.54	3.47
	$HT_{0.5}^{emp} + B_3^{emp}$	0.400	0.668	1.54	3.47
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.399	0.666	1.54	3.47
	$HT_{0.9}^{emp} + B_3^{emp}$	0.398	0.663	1.53	3.44
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.398	0.662	1.53	3.44
	$LR_{exp} + B_3^{NQE}$	0.175	0.382	0.583	0.836
	$LR_{Par} + B_3^{NQE}$	0.176	0.383	0.584	0.838
	$LR_{cons} + B_3^{NQE}$	0.175	0.382	0.583	0.836
	$KS + B_3^{NQE}$	0.175	0.382	0.583	0.836

Table 31: Mean Relative Error: Par(10,3.5), $\rho = 0.75$, $n = 500$

	Mean Relative Error	$\mathbb{E}[W]$ $= 45$	$\mathbb{P}\{W \geq 150\}$ ≈ 0.056	$\mathbb{P}\{W \geq 200\}$ ≈ 0.032	$\mathbb{P}\{W \geq 250\}$ ≈ 0.020
Theoretical	Fitted Exp	0.173	0.495	0.433	0.433
	Fitted Par	0.211	0.330	0.397	0.446
	Empirical	0.314	0.628	0.842	1.03
Quasi-Empirical	NQE_{exp}	0.287	0.615	0.758	0.858
	NQE_{Par}	0.212	0.360	0.437	0.495
	CQE_{exp}	0.290	0.614	0.795	0.931
	CQE_{Par}	0.457	0.847	0.931	0.968
Hypothesis Test	$HT_{0.1}^{emp}$	0.214	0.372	0.449	0.503
	$HT_{0.1}^{NQE}$	0.213	0.368	0.446	0.505
	$HT_{0.5}^{emp}$	0.246	0.445	0.559	0.652
	$HT_{0.5}^{NQE}$	0.236	0.434	0.527	0.593
	$HT_{0.9}^{emp}$	0.303	0.553	0.731	0.894
	$HT_{0.9}^{NQE}$	0.273	0.530	0.662	0.759
Likelihood Ratio	LR_{exp}	0.273	0.566	0.706	0.808
	LR_{Par}	0.218	0.376	0.455	0.513
	LR_{cons}	0.273	0.566	0.706	0.808
Heuristics	KS	0.213	0.361	0.438	0.496
	B_3^{emp}	0.212	0.360	0.437	0.495
	B_3^{NQE}	0.212	0.360	0.437	0.495
	B_4^{emp}	0.313	0.626	0.838	1.03
	B_4^{NQE}	0.284	0.608	0.751	0.852
	$HT_{0.1}^{emp} + B_3^{emp}$	0.214	0.372	0.449	0.503
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.213	0.368	0.446	0.505
	$HT_{0.5}^{emp} + B_3^{emp}$	0.246	0.445	0.559	0.652
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.236	0.434	0.527	0.593
	$HT_{0.9}^{emp} + B_3^{emp}$	0.303	0.553	0.731	0.894
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.273	0.530	0.662	0.759
	$LR_{exp} + B_3^{NQE}$	0.273	0.566	0.706	0.808
	$LR_{Par} + B_3^{NQE}$	0.218	0.376	0.455	0.513
	$LR_{cons} + B_3^{NQE}$	0.273	0.566	0.706	0.808
	$KS + B_3^{NQE}$	0.213	0.361	0.438	0.496

Table 32: Mean Relative Error: Par(10,4), $\rho = 0.75$, $n = 500$

	MRE	$\mathbb{E}[W]$ $= 30$	$\mathbb{P}\{W \geq 100\}$ ≈ 0.060	$\mathbb{P}\{W \geq 120\}$ ≈ 0.039	$\mathbb{P}\{W \geq 140\}$ ≈ 0.025
Theoretical	Fitted Exp	0.530	1.41	1.69	1.93
	Fitted Par	0.136	0.281	0.332	0.382
	Empirical	0.200	0.459	0.584	0.719
Quasi-Empirical	NQE_{exp}	0.193	0.478	0.588	0.692
	NQE_{Par}	0.152	0.313	0.377	0.442
	CQE_{exp}	0.196	0.458	0.577	0.701
	CQE_{Par}	0.266	0.599	0.684	0.755
Hypothesis Test	$HT_{0.1}^{emp}$	0.153	0.315	0.380	0.445
	$HT_{0.1}^{NQE}$	0.153	0.316	0.380	0.445
	$HT_{0.5}^{emp}$	0.162	0.345	0.420	0.496
	$HT_{0.5}^{NQE}$	0.158	0.338	0.406	0.473
	$HT_{0.9}^{emp}$	0.167	0.370	0.459	0.553
	$HT_{0.9}^{NQE}$	0.169	0.385	0.470	0.554
Likelihood Ratio	LR_{exp}	0.188	0.445	0.549	0.651
	LR_{Par}	0.164	0.345	0.415	0.483
	LR_{cons}	0.188	0.445	0.549	0.651
Heuristics	KS	0.154	0.319	0.384	0.450
	B_3^{emp}	0.152	0.313	0.377	0.442
	B_3^{NQE}	0.152	0.313	0.377	0.442
	B_4^{emp}	0.170	0.382	0.479	0.581
	B_4^{NQE}	0.147	0.348	0.428	0.505
	$HT_{0.1}^{emp} + B_3^{emp}$	0.153	0.315	0.380	0.445
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.153	0.316	0.380	0.445
	$HT_{0.5}^{emp} + B_3^{emp}$	0.162	0.345	0.420	0.496
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.158	0.338	0.406	0.473
	$HT_{0.9}^{emp} + B_3^{emp}$	0.167	0.370	0.459	0.553
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.169	0.385	0.470	0.554
	$LR_{exp} + B_3^{NQE}$	0.188	0.445	0.549	0.651
	$LR_{Par} + B_3^{NQE}$	0.164	0.345	0.415	0.483
	$LR_{cons} + B_3^{NQE}$	0.188	0.445	0.549	0.651
	$KS + B_3^{NQE}$	0.154	0.319	0.384	0.450

Table 33: Mean Relative Error: Par(10,5), $\rho = 0.667$, $n = 500$

	MRE	$\mathbb{E}[W]$ $= 15$	$\mathbb{P}\{W \geq 50\}$ ≈ 0.656	$\mathbb{P}\{W \geq 70\}$ ≈ 0.026	$\mathbb{P}\{W \geq 100\}$ ≈ 0.007
Theoretical	Fitted Exp	0.774	1.90	3.40	6.809
	Fitted Par	0.061	0.133	0.188	0.276
	Empirical	0.074	0.166	0.262	0.456
Quasi-Empirical	NQE_{exp}	0.101	0.245	0.389	0.674
	NQE_{Par}	0.060	0.133	0.190	0.283
	CQE_{exp}	0.072	0.165	0.260	0.441
	CQE_{Par}	0.126	0.301	0.422	0.583
Hypothesis Test	$HT_{0.1}^{emp}$	0.060	0.133	0.190	0.283
	$HT_{0.1}^{NQE}$	0.060	0.133	0.190	0.283
	$HT_{0.5}^{emp}$	0.062	0.139	0.201	0.299
	$HT_{0.5}^{NQE}$	0.071	0.161	0.237	0.371
	$HT_{0.9}^{emp}$	0.065	0.149	0.220	0.342
	$HT_{0.9}^{NQE}$	0.080	0.187	0.287	0.475
Likelihood Ratio	LR_{exp}	0.073	0.173	0.257	0.404
	LR_{Par}	0.069	0.160	0.235	0.360
	LR_{cons}	0.073	0.174	0.259	0.407
Heuristics	KS	0.060	0.133	0.190	0.283
	B_3^{emp}	0.060	0.133	0.190	0.283
	B_3^{NQE}	0.060	0.133	0.190	0.283
	B_4^{emp}	0.060	0.133	0.190	0.283
	B_4^{NQE}	0.060	0.133	0.190	0.283
	$HT_{0.1}^{emp} + B_3^{emp}$	0.060	0.133	0.190	0.283
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.060	0.133	0.190	0.283
	$HT_{0.5}^{emp} + B_3^{emp}$	0.062	0.139	0.201	0.299
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.071	0.161	0.237	0.371
	$HT_{0.9}^{emp} + B_3^{emp}$	0.065	0.149	0.220	0.342
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.080	0.187	0.287	0.475
	$LR_{exp} + B_3^{NQE}$	0.073	0.173	0.257	0.404
	$LR_{Par} + B_3^{NQE}$	0.069	0.160	0.235	0.360
	$LR_{cons} + B_3^{NQE}$	0.073	0.174	0.259	0.407
	$KS + B_3^{NQE}$	0.060	0.133	0.190	0.283

Table 34: Mean Relative Error: Par(10,10), $\rho = 0.5625$, $n = 500$

	MRE	$\mathbb{E}[W]$ $= 7.35$	$\mathbb{P}\{W \geq 10\}$ ≈ 0.2787	$\mathbb{P}\{W \geq 30\}$ ≈ 0.0444	$\mathbb{P}\{W \geq 50\}$ ≈ 0.0071
Theoretical	Fitted Exp	0.962	0.365	2.93	10.3
	Fitted Par	0.016	0.015	0.038	0.063
	Empirical	0.016	0.015	0.041	0.068
Quasi-Empirical	NQE_{exp}	0.139	0.061	0.416	1.40
	NQE_{Par}	0.015	0.015	0.038	0.062
	CQE_{exp}	0.016	0.015	0.040	0.067
	CQE_{Par}	0.028	0.025	0.073	0.117
Hypothesis Test	$HT_{0.1}^{emp}$	0.015	0.015	0.038	0.062
	$HT_{0.1}^{NQE}$	0.015	0.015	0.038	0.062
	$HT_{0.5}^{emp}$	0.016	0.015	0.039	0.063
	$HT_{0.5}^{NQE}$	0.069	0.034	0.200	0.635
	$HT_{0.9}^{emp}$	0.016	0.015	0.039	0.065
	$HT_{0.9}^{NQE}$	0.087	0.041	0.255	0.828
Likelihood Ratio	LR_{exp}	0.121	0.054	0.361	1.20
	LR_{Par}	0.103	0.047	0.307	1.01
	LR_{cons}	0.121	0.054	0.361	1.20
Heuristics	KS	0.032	0.021	0.090	0.247
	B_3^{emp}	0.015	0.015	0.038	0.062
	B_3^{NQE}	0.015	0.015	0.038	0.062
	B_4^{emp}	0.015	0.015	0.038	0.062
	B_4^{NQE}	0.015	0.015	0.038	0.062
	$HT_{0.1}^{emp} + B_3^{emp}$	0.015	0.015	0.038	0.062
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.015	0.015	0.038	0.062
	$HT_{0.5}^{emp} + B_3^{emp}$	0.016	0.015	0.039	0.063
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.069	0.034	0.200	0.635
	$HT_{0.9}^{emp} + B_3^{emp}$	0.016	0.015	0.039	0.065
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.087	0.041	0.255	0.828
	$LR_{exp} + B_3^{NQE}$	0.121	0.054	0.361	1.20
	$LR_{Par} + B_3^{NQE}$	0.103	0.047	0.307	1.01
	$LR_{cons} + B_3^{NQE}$	0.121	0.054	0.361	1.20
	$KS + B_3^{NQE}$	0.032	0.021	0.090	0.247

Table 35: Mean Relative Error: Unif(0, 1), $\rho = 0.25$, $n = 500$

	MRE	$\mathbb{E}[W]$ ≈ 0.11	$\Pr\{W \geq 0.05\}$ ≈ 0.2135	$\Pr\{W \geq 0.1\}$ ≈ 0.0919	$\Pr\{W \geq 0.5\}$ ≈ 0.0165
Theoretical	Fitted Exp	0.494	0.027	0.280	2.37
	Fitted Par	1.05E+83	3.68	9.89	59.7
	Empirical	0.040	0.026	0.047	0.080
Quasi-Empirical	NQE_{exp}	0.061	0.026	0.050	0.314
	NQE_{Par}	0.042	0.026	0.047	0.093
	CQE_{exp}	0.040	0.026	0.047	0.080
	CQE_{Par}	0.040	0.026	0.047	0.080
Hypothesis Test	$HT_{0.1}^{emp}$	0.042	0.026	0.047	0.092
	$HT_{0.1}^{NQE}$	0.044	0.026	0.048	0.124
	$HT_{0.5}^{emp}$	0.040	0.026	0.047	0.079
	$HT_{0.5}^{NQE}$	0.050	0.026	0.049	0.201
	$HT_{0.9}^{emp}$	0.040	0.026	0.047	0.080
	$HT_{0.9}^{NQE}$	0.057	0.026	0.049	0.284
Likelihood Ratio	LR_{exp}	0.061	0.026	0.050	0.314
	LR_{Par}	0.061	0.026	0.050	0.314
	LR_{cons}	0.061	0.026	0.050	0.314
Heuristics	KS	0.043	0.026	0.047	0.140
	B_3^{emp}	0.042	0.026	0.047	0.093
	B_3^{NQE}	0.042	0.026	0.047	0.093
	B_4^{emp}	0.042	0.026	0.047	0.093
	B_4^{NQE}	0.042	0.026	0.047	0.093
	$HT_{0.1}^{emp} + B_3^{emp}$	0.042	0.026	0.047	0.092
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.044	0.026	0.048	0.124
	$HT_{0.5}^{emp} + B_3^{emp}$	0.040	0.026	0.047	0.079
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.050	0.026	0.049	0.201
	$HT_{0.9}^{emp} + B_3^{emp}$	0.040	0.026	0.047	0.080
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.057	0.026	0.049	0.284
	$LR_{exp} + B_3^{NQE}$	0.061	0.026	0.050	0.314
	$LR_{Par} + B_3^{NQE}$	0.061	0.026	0.050	0.314
	$LR_{cons} + B_3^{NQE}$	0.061	0.026	0.050	0.314
	$KS + B_3^{NQE}$	0.043	0.026	0.047	0.140

Table 36: Mean Relative Error: 0.25 Emp + 0.75 Exp(1), $\rho = 0.475$, $n = 500$

	MRE	$\mathbb{E}[W]$ $= 4.0235$	$\mathbb{P}\{W \geq 20\}$ ≈ 0.0377	$\mathbb{P}\{W \geq 30\}$ ≈ 0.0095	$\mathbb{P}\{W \geq 40\}$ ≈ 0.0024
Theoretical	Fitted Exp	1.14	3.17	8.54	20.9
	Fitted Par	1.31E+56	25.5	104	419
	Empirical	0.182	0.359	0.542	0.751
Quasi-Empirical	NQE_{exp}	0.182	0.360	0.544	0.753
	NQE_{Par}	0.184	0.366	0.555	0.771
	CQE_{exp}	0.182	0.360	0.543	0.752
	CQE_{Par}	0.165	0.314	0.467	0.637
Hypothesis Test	$HT_{0.1}^{emp}$	0.184	0.366	0.555	0.771
	$HT_{0.1}^{NQE}$	0.184	0.366	0.555	0.771
	$HT_{0.5}^{emp}$	0.184	0.366	0.555	0.771
	$HT_{0.5}^{NQE}$	0.184	0.366	0.555	0.771
	$HT_{0.9}^{emp}$	0.184	0.366	0.555	0.771
	$HT_{0.9}^{NQE}$	0.184	0.366	0.555	0.771
Likelihood Ratio	LR_{exp}	0.182	0.360	0.544	0.753
	LR_{Par}	0.182	0.360	0.544	0.753
	LR_{cons}	0.182	0.360	0.544	0.753
Heuristics	KS	0.183	0.362	0.547	0.759
	B_3^{emp}	0.184	0.366	0.555	0.771
	B_3^{NQE}	0.184	0.366	0.555	0.771
	B_4^{emp}	0.184	0.366	0.555	0.771
	B_4^{NQE}	0.184	0.366	0.555	0.771
	$HT_{0.1}^{emp} + B_3^{emp}$	0.184	0.366	0.555	0.771
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.184	0.366	0.555	0.771
	$HT_{0.5}^{emp} + B_3^{emp}$	0.184	0.366	0.555	0.771
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.184	0.366	0.555	0.771
	$HT_{0.9}^{emp} + B_3^{emp}$	0.184	0.366	0.555	0.771
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.184	0.366	0.555	0.771
	$LR_{exp} + B_3^{NQE}$	0.182	0.360	0.544	0.753
	$LR_{Par} + B_3^{NQE}$	0.182	0.360	0.544	0.753
	$LR_{cons} + B_3^{NQE}$	0.182	0.360	0.544	0.753
	$KS + B_3^{NQE}$	0.183	0.362	0.547	0.759

Table 37: Mean Relative Error: 0.75 Unif(0,10) + 0.25 Exp(1), $\rho = 0.325$, $n = 500$

	MRE	$\mathbb{E}[W]$ $= 1.5431$	$\mathbb{P}\{W \geq 5\}$ ≈ 0.1259	$\mathbb{P}\{W \geq 10\}$ ≈ 0.0412	$\mathbb{P}\{W \geq 20\}$ ≈ 0.0035
Theoretical	Fitted Exp	1.03	0.534	1.79	10.5
	Fitted Par	6.04E+64	6.94	23.3	282
	Empirical	0.331	0.351	0.449	1.02
Quasi-Empirical	NQE_{exp}	0.331	0.351	0.449	1.03
	NQE_{Par}	0.333	0.352	0.451	1.05
	CQE_{exp}	0.331	0.351	0.449	1.03
	CQE_{Par}	0.311	0.337	0.401	0.889
Hypothesis Test	$HT_{0.1}^{emp}$	0.333	0.352	0.451	1.05
	$HT_{0.1}^{NQE}$	0.333	0.352	0.451	1.05
	$HT_{0.5}^{emp}$	0.333	0.352	0.451	1.05
	$HT_{0.5}^{NQE}$	0.333	0.352	0.451	1.05
	$HT_{0.9}^{emp}$	0.333	0.352	0.451	1.05
	$HT_{0.9}^{NQE}$	0.333	0.352	0.451	1.05
Likelihood Ratio	LR_{exp}	0.331	0.351	0.449	1.03
	LR_{Par}	0.331	0.351	0.449	1.03
	LR_{cons}	0.331	0.351	0.449	1.03
Heuristics	KS	0.332	0.351	0.449	1.03
	B_3^{emp}	0.333	0.352	0.451	1.05
	B_3^{NQE}	0.333	0.352	0.451	1.05
	B_4^{emp}	0.333	0.352	0.451	1.05
	B_4^{NQE}	0.333	0.352	0.451	1.05
	$HT_{0.1}^{emp} + B_3^{emp}$	0.333	0.352	0.451	1.05
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.333	0.352	0.451	1.05
	$HT_{0.5}^{emp} + B_3^{emp}$	0.333	0.352	0.451	1.05
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.333	0.352	0.451	1.05
	$HT_{0.9}^{emp} + B_3^{emp}$	0.333	0.352	0.451	1.05
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.333	0.352	0.451	1.05
	$LR_{exp} + B_3^{NQE}$	0.331	0.351	0.449	1.03
	$LR_{Par} + B_3^{NQE}$	0.331	0.351	0.449	1.03
	$LR_{cons} + B_3^{NQE}$	0.331	0.351	0.449	1.03
	$KS + B_3^{NQE}$	0.332	0.351	0.449	1.03

Table 38: Mean Relative Error: 0.25 Unif(0,10) + 0.75 Par(10,4), $\rho = 0.625$, $n = 500$

	MRE	$\mathbb{E}[W]$ $= 13.352$	$\mathbb{P}\{W \geq 50\}$ ≈ 0.0593	$\mathbb{P}\{W \geq 70\}$ ≈ 0.0283	$\mathbb{P}\{W \geq 90\}$ ≈ 0.0149
Theoretical	Fitted Exp	0.572	1.37	1.74	1.88
	Fitted Par	1.88E+53	15.9	34.3	66.1
	Empirical	0.227	0.411	0.602	0.842
Quasi-Empirical	NQE_{exp}	0.173	0.323	0.466	0.644
	NQE_{Par}	0.219	0.365	0.487	0.597
	CQE_{exp}	0.219	0.411	0.592	0.799
	CQE_{Par}	0.134	0.335	0.527	0.691
Hypothesis Test	$HT_{0.1}^{emp}$	0.219	0.365	0.487	0.597
	$HT_{0.1}^{NQE}$	0.219	0.365	0.487	0.597
	$HT_{0.5}^{emp}$	0.219	0.365	0.487	0.597
	$HT_{0.5}^{NQE}$	0.219	0.365	0.487	0.597
	$HT_{0.9}^{emp}$	0.219	0.365	0.487	0.597
	$HT_{0.9}^{NQE}$	0.219	0.365	0.487	0.597
Likelihood Ratio	LR_{exp}	0.177	0.329	0.479	0.667
	LR_{Par}	0.212	0.368	0.500	0.619
	LR_{cons}	0.177	0.329	0.479	0.667
Heuristics	KS	0.219	0.368	0.494	0.606
	B_3^{emp}	0.219	0.365	0.487	0.597
	B_3^{NQE}	0.219	0.365	0.487	0.597
	B_4^{emp}	0.209	0.371	0.509	0.671
	B_4^{NQE}	0.170	0.291	0.376	0.485
	$HT_{0.1}^{emp} + B_3^{emp}$	0.219	0.365	0.487	0.597
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.219	0.365	0.487	0.597
	$HT_{0.5}^{emp} + B_3^{emp}$	0.219	0.365	0.487	0.597
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.219	0.365	0.487	0.597
	$HT_{0.9}^{emp} + B_3^{emp}$	0.219	0.365	0.487	0.597
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.219	0.365	0.487	0.597
	$LR_{exp} + B_3^{NQE}$	0.177	0.329	0.479	0.667
	$LR_{Par} + B_3^{NQE}$	0.212	0.368	0.500	0.619
	$LR_{cons} + B_3^{NQE}$	0.177	0.329	0.479	0.667
	$KS + B_3^{NQE}$	0.219	0.368	0.494	0.606

Table 39: Mean Relative Error: 0.75 Unif(0, 10) + 0.25 Par(10,4), $\rho = 0.375$, $n = 500$

	MRE	$\mathbb{E}[W]$ $= 3.3107$	$\mathbb{P}\{W \geq 20\}$ ≈ 0.0384	$\mathbb{P}\{W \geq 30\}$ ≈ 0.0173	$\mathbb{P}\{W \geq 40\}$ ≈ 0.0091
Theoretical	Fitted Exp	0.361	0.849	0.794	0.489
	Fitted Par	8.02E+56	25.1	57.0	109
	Empirical	0.222	0.322	0.499	0.728
Quasi-Empirical	NQE_{exp}	0.129	0.264	0.355	0.513
	NQE_{Par}	0.282	0.319	0.429	0.544
	CQE_{exp}	0.214	0.322	0.498	0.716
	CQE_{Par}	0.094	0.262	0.543	0.743
Hypothesis Test	$HT_{0.1}^{emp}$	0.282	0.319	0.429	0.544
	$HT_{0.1}^{NQE}$	0.282	0.319	0.429	0.544
	$HT_{0.5}^{emp}$	0.282	0.319	0.429	0.544
	$HT_{0.5}^{NQE}$	0.282	0.319	0.429	0.544
	$HT_{0.9}^{emp}$	0.282	0.319	0.429	0.544
	$HT_{0.9}^{NQE}$	0.282	0.319	0.429	0.544
Likelihood Ratio	LR_{exp}	0.129	0.264	0.355	0.513
	LR_{Par}	0.212	0.285	0.443	0.640
	LR_{cons}	0.129	0.264	0.355	0.513
Heuristics	KS	0.245	0.294	0.423	0.574
	B_3^{emp}	0.282	0.319	0.429	0.544
	B_3^{NQE}	0.282	0.319	0.429	0.544
	B_4^{emp}	0.229	0.310	0.430	0.588
	B_4^{NQE}	0.151	0.258	0.242	0.305
	$HT_{0.1}^{emp} + B_3^{emp}$	0.282	0.319	0.429	0.544
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.282	0.319	0.429	0.544
	$HT_{0.5}^{emp} + B_3^{emp}$	0.282	0.319	0.429	0.544
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.282	0.319	0.429	0.544
	$HT_{0.9}^{emp} + B_3^{emp}$	0.282	0.319	0.429	0.544
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.282	0.319	0.429	0.544
	$LR_{exp} + B_3^{NQE}$	0.129	0.264	0.355	0.513
	$LR_{Par} + B_3^{NQE}$	0.212	0.285	0.443	0.640
	$LR_{cons} + B_3^{NQE}$	0.129	0.264	0.355	0.513
	$KS + B_3^{NQE}$	0.245	0.294	0.423	0.574

A.3 Numerical Results: One Thousand Observed Service Times

In this section, we provide the detailed results that are not presented in Section 5.3.2 for $n = 1000$ in Tables 40 to 49.

Table 40: Mean Relative Error: Exp(1), $\rho = 0.75$, $n = 1000$

	MRE	$\mathbb{E}[W]$ $= 3$	$\mathbb{P}\{W \geq 10\}$ ≈ 0.062	$\mathbb{P}\{W \geq 15\}$ ≈ 0.018	$\mathbb{P}\{W \geq 20\}$ ≈ 0.005
Theoretical	Fitted Exp	0.097	0.211	0.304	0.401
	Fitted Par	1.05E+78	15.23	55.58	196
	Empirical	0.098	0.216	0.314	0.416
Quasi-Empirical	NQE_{exp}	0.098	0.213	0.307	0.407
	NQE_{Par}	0.430	0.536	1.33	3.21
	CQE_{exp}	0.098	0.216	0.314	0.419
	CQE_{Par}	0.131	0.297	0.389	0.445
Hypothesis Test	$HT_{0.1}^{emp}$	0.430	0.536	1.33	3.21
	$HT_{0.1}^{NQE}$	0.430	0.536	1.33	3.21
	$HT_{0.5}^{emp}$	0.430	0.536	1.33	3.21
	$HT_{0.5}^{NQE}$	0.430	0.536	1.33	3.21
	$HT_{0.9}^{emp}$	0.430	0.536	1.33	3.21
	$HT_{0.9}^{NQE}$	0.430	0.536	1.33	3.21
Likelihood Ratio	LR_{exp}	0.098	0.213	0.307	0.407
	LR_{Par}	0.098	0.212	0.305	0.407
	LR_{cons}	0.098	0.213	0.307	0.407
Heuristics	KS	0.098	0.213	0.307	0.407
	B_3^{emp}	0.247	0.445	1.04	2.33
	B_3^{NQE}	0.247	0.445	1.04	2.33
	B_4^{emp}	0.108	0.242	0.428	0.725
	B_4^{NQE}	0.108	0.238	0.421	0.714
	$HT_{0.1}^{emp} + B_3^{emp}$	0.247	0.445	1.04	2.33
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.247	0.445	1.04	2.33
	$HT_{0.5}^{emp} + B_3^{emp}$	0.247	0.445	1.04	2.33
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.247	0.445	1.04	2.33
	$HT_{0.9}^{emp} + B_3^{emp}$	0.247	0.445	1.04	2.33
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.247	0.445	1.04	2.33
	$LR_{exp} + B_3^{NQE}$	0.098	0.213	0.307	0.407
	$LR_{Par} + B_3^{NQE}$	0.098	0.212	0.305	0.407
	$LR_{cons} + B_3^{NQE}$	0.098	0.213	0.307	0.407
	$KS + B_3^{NQE}$	0.098	0.213	0.307	0.407

Table 41: Mean Relative Error: Par(10,3.5), $\rho = 0.75$, $n = 1000$

	Mean Relative Error	$\mathbb{E}[W]$ $= 45$	$\mathbb{P}\{W \geq 150\}$ ≈ 0.056	$\mathbb{P}\{W \geq 200\}$ ≈ 0.032	$\mathbb{P}\{W \geq 250\}$ ≈ 0.020
Theoretical	Fitted Exp	0.182	0.505	0.434	0.453
	Fitted Par	0.146	0.244	0.288	0.317
	Empirical	0.906	0.584	0.887	1.23
Quasi-Empirical	NQE_{exp}	0.788	0.613	0.867	1.12
	NQE_{Par}	0.175	0.283	0.343	0.388
	CQE_{exp}	0.595	0.582	0.854	1.14
	CQE_{Par}	0.396	0.766	0.884	0.945
Hypothesis Test	$HT_{0.1}^{emp}$	0.213	0.326	0.415	0.492
	$HT_{0.1}^{NQE}$	0.186	0.310	0.374	0.420
	$HT_{0.5}^{emp}$	0.236	0.379	0.509	0.631
	$HT_{0.5}^{NQE}$	0.206	0.361	0.455	0.527
	$HT_{0.9}^{emp}$	0.267	0.423	0.589	0.755
	$HT_{0.9}^{NQE}$	0.222	0.402	0.515	0.606
Likelihood Ratio	LR_{exp}	0.774	0.551	0.792	1.04
	LR_{Par}	0.707	0.385	0.528	0.684
	LR_{cons}	0.774	0.551	0.792	1.04
Heuristics	KS	0.175	0.283	0.343	0.388
	B_3^{emp}	0.175	0.283	0.343	0.388
	B_3^{NQE}	0.175	0.283	0.343	0.388
	B_4^{emp}	0.906	0.584	0.887	1.23
	B_4^{NQE}	0.788	0.613	0.867	1.12
	$HT_{0.1}^{emp} + B_3^{emp}$	0.213	0.326	0.415	0.492
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.186	0.310	0.374	0.420
	$HT_{0.5}^{emp} + B_3^{emp}$	0.236	0.379	0.509	0.631
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.206	0.361	0.455	0.527
	$HT_{0.9}^{emp} + B_3^{emp}$	0.267	0.423	0.589	0.755
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.222	0.402	0.515	0.606
	$LR_{exp} + B_3^{NQE}$	0.774	0.551	0.792	1.04
	$LR_{Par} + B_3^{NQE}$	0.707	0.385	0.528	0.684
	$LR_{cons} + B_3^{NQE}$	0.774	0.551	0.792	1.04
	$KS + B_3^{NQE}$	0.175	0.283	0.343	0.388

Table 42: Mean Relative Error: Par(10,4), $\rho = 0.75$, $n = 1000$

	MRE	$\mathbb{E}[W]$ $= 30$	$\mathbb{P}\{W \geq 100\}$ ≈ 0.060	$\mathbb{P}\{W \geq 120\}$ ≈ 0.039	$\mathbb{P}\{W \geq 140\}$ ≈ 0.025
Theoretical	Fitted Exp	0.504	1.35	1.61	1.821
	Fitted Par	0.078	0.166	0.195	0.223
	Empirical	0.133	0.297	0.389	0.492
Quasi-Empirical	NQE_{exp}	0.131	0.329	0.412	0.495
	NQE_{Par}	0.086	0.175	0.208	0.240
	CQE_{exp}	0.123	0.293	0.376	0.465
	CQE_{Par}	0.217	0.507	0.600	0.681
Hypothesis Test	$HT_{0.1}^{emp}$	0.085	0.175	0.209	0.242
	$HT_{0.1}^{NQE}$	0.085	0.177	0.212	0.246
	$HT_{0.5}^{emp}$	0.109	0.223	0.281	0.343
	$HT_{0.5}^{NQE}$	0.100	0.223	0.269	0.314
	$HT_{0.9}^{emp}$	0.114	0.239	0.307	0.381
	$HT_{0.9}^{NQE}$	0.106	0.242	0.298	0.354
Likelihood Ratio	LR_{exp}	0.109	0.261	0.326	0.394
	LR_{Par}	0.088	0.181	0.214	0.248
	LR_{cons}	0.110	0.264	0.330	0.398
Heuristics	KS	0.086	0.175	0.208	0.240
	B_3^{emp}	0.086	0.175	0.208	0.240
	B_3^{NQE}	0.086	0.175	0.208	0.240
	B_4^{emp}	0.106	0.229	0.289	0.355
	B_4^{NQE}	0.091	0.216	0.265	0.316
	$HT_{0.1}^{emp} + B_3^{emp}$	0.085	0.175	0.209	0.242
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.085	0.177	0.212	0.246
	$HT_{0.5}^{emp} + B_3^{emp}$	0.109	0.223	0.281	0.343
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.100	0.223	0.269	0.314
	$HT_{0.9}^{emp} + B_3^{emp}$	0.114	0.239	0.307	0.381
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.106	0.242	0.298	0.354
	$LR_{exp} + B_3^{NQE}$	0.109	0.261	0.326	0.394
	$LR_{Par} + B_3^{NQE}$	0.088	0.181	0.214	0.248
	$LR_{cons} + B_3^{NQE}$	0.110	0.264	0.330	0.398
	$KS + B_3^{NQE}$	0.086	0.175	0.208	0.240

Table 43: Mean Relative Error: Par(10,5), $\rho \simeq 0.67$, $n = 1000$

	Mean Relative Error	$\mathbb{E}[W]$ $= 15$	$\mathbb{P}\{W \geq 50\}$ ≈ 0.0656	$\mathbb{P}\{W \geq 70\}$ ≈ 0.0262	$\mathbb{P}\{W \geq 100\}$ ≈ 0.0070
Theoretical	Fitted Exp	0.787	1.93	3.46	6.948
	Fitted Par	0.038	0.084	0.120	0.176
	Empirical	0.058	0.130	0.219	0.427
Quasi-Empirical	NQE_{exp}	0.099	0.246	0.410	0.748
	NQE_{Par}	0.040	0.089	0.129	0.194
	CQE_{exp}	0.056	0.129	0.215	0.392
	CQE_{Par}	0.086	0.211	0.314	0.468
Hypothesis Test	$HT_{0.1}^{emp}$	0.040	0.090	0.131	0.196
	$HT_{0.1}^{NQE}$	0.048	0.109	0.164	0.263
	$HT_{0.5}^{emp}$	0.045	0.099	0.153	0.263
	$HT_{0.5}^{NQE}$	0.062	0.147	0.235	0.407
	$HT_{0.9}^{emp}$	0.048	0.107	0.171	0.306
	$HT_{0.9}^{NQE}$	0.064	0.151	0.242	0.424
Likelihood Ratio	LR_{exp}	0.052	0.126	0.194	0.314
	LR_{Par}	0.042	0.094	0.138	0.209
	LR_{cons}	0.052	0.126	0.194	0.314
Heuristics	KS	0.040	0.090	0.130	0.195
	B_3^{emp}	0.040	0.089	0.129	0.194
	B_3^{NQE}	0.040	0.089	0.129	0.194
	B_4^{emp}	0.040	0.089	0.129	0.194
	B_4^{NQE}	0.040	0.089	0.129	0.194
	$HT_{0.1}^{emp} + B_3^{emp}$	0.040	0.090	0.131	0.196
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.048	0.109	0.164	0.263
	$HT_{0.5}^{emp} + B_3^{emp}$	0.045	0.099	0.153	0.263
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.062	0.147	0.235	0.407
	$HT_{0.9}^{emp} + B_3^{emp}$	0.048	0.107	0.171	0.306
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.064	0.151	0.242	0.424
	$LR_{exp} + B_3^{NQE}$	0.052	0.126	0.194	0.314
	$LR_{Par} + B_3^{NQE}$	0.042	0.094	0.138	0.209
	$LR_{cons} + B_3^{NQE}$	0.052	0.126	0.194	0.314
	$KS + B_3^{NQE}$	0.040	0.090	0.130	0.195

Table 44: Mean Relative Error: Par(10,10), $\rho = 0.5625$, $n = 1000$

	MRE	$\mathbb{E}[W]$ $= 7.35$	$\mathbb{P}\{W \geq 10\}$ ≈ 0.2787	$\mathbb{P}\{W \geq 30\}$ ≈ 0.0444	$\mathbb{P}\{W \geq 50\}$ ≈ 0.0071
Theoretical	Fitted Exp	0.971	0.369	2.95	10.4
	Fitted Par	0.010	0.010	0.026	0.044
	Empirical	0.011	0.010	0.029	0.048
Quasi-Empirical	NQE_{exp}	0.108	0.046	0.324	1.10
	NQE_{Par}	0.010	0.010	0.026	0.042
	CQE_{exp}	0.011	0.010	0.029	0.048
	CQE_{Par}	0.016	0.014	0.043	0.071
Hypothesis Test	$HT_{0.1}^{emp}$	0.010	0.010	0.026	0.043
	$HT_{0.1}^{NQE}$	0.019	0.013	0.053	0.139
	$HT_{0.5}^{emp}$	0.010	0.010	0.026	0.043
	$HT_{0.5}^{NQE}$	0.028	0.016	0.081	0.237
	$HT_{0.9}^{emp}$	0.011	0.010	0.027	0.045
	$HT_{0.9}^{NQE}$	0.037	0.020	0.108	0.333
Likelihood Ratio	LR_{exp}	0.081	0.036	0.242	0.808
	LR_{Par}	0.055	0.026	0.161	0.523
	LR_{cons}	0.081	0.036	0.242	0.808
Heuristics	KS	0.010	0.010	0.026	0.043
	B_3^{emp}	0.010	0.010	0.026	0.042
	B_3^{NQE}	0.010	0.010	0.026	0.042
	B_4^{emp}	0.010	0.010	0.026	0.042
	B_4^{NQE}	0.010	0.010	0.026	0.042
	$HT_{0.1}^{emp} + B_3^{emp}$	0.010	0.010	0.026	0.043
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.019	0.013	0.053	0.139
	$HT_{0.5}^{emp} + B_3^{emp}$	0.010	0.010	0.026	0.043
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.028	0.016	0.081	0.237
	$HT_{0.9}^{emp} + B_3^{emp}$	0.011	0.010	0.027	0.045
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.037	0.020	0.108	0.333
	$LR_{exp} + B_3^{NQE}$	0.081	0.036	0.242	0.808
	$LR_{Par} + B_3^{NQE}$	0.055	0.026	0.161	0.523
	$LR_{cons} + B_3^{NQE}$	0.081	0.036	0.242	0.808
	$KS + B_3^{NQE}$	0.010	0.010	0.026	0.043

Table 45: Mean Relative Error: Unif(0,1), $\rho = 0.25$, $n = 1000$

	MRE	$\mathbb{E}[W]$ $= 0.11$	$\mathbb{P}\{W \geq 0.05\}$ ≈ 0.2135	$\mathbb{P}\{W \geq 0.1\}$ ≈ 0.0919	$\mathbb{P}\{W \geq 0.5\}$ ≈ 0.0165
Theoretical	Fitted Exp	0.495	0.018	0.281	2.37
	Fitted Par	3.03E+77	3.68	9.89	59.7
	Empirical	0.026	0.017	0.031	0.053
Quasi-Empirical	NQE_{exp}	0.059	0.017	0.037	0.348
	NQE_{Par}	0.027	0.017	0.031	0.057
	CQE_{exp}	0.026	0.017	0.031	0.053
	CQE_{Par}	0.026	0.017	0.031	0.054
Hypothesis Test	$HT_{0.1}^{emp}$	0.027	0.017	0.031	0.057
	$HT_{0.1}^{NQE}$	0.033	0.017	0.033	0.116
	$HT_{0.5}^{emp}$	0.026	0.017	0.031	0.053
	$HT_{0.5}^{NQE}$	0.039	0.017	0.032	0.169
	$HT_{0.9}^{emp}$	0.026	0.017	0.031	0.053
	$HT_{0.9}^{NQE}$	0.057	0.017	0.037	0.325
Likelihood Ratio	LR_{exp}	0.059	0.017	0.037	0.348
	LR_{Par}	0.059	0.017	0.037	0.348
	LR_{cons}	0.059	0.017	0.037	0.348
Heuristics	KS	0.030	0.017	0.030	0.105
	B_3^{emp}	0.027	0.017	0.031	0.057
	B_3^{NQE}	0.027	0.017	0.031	0.057
	B_4^{emp}	0.027	0.017	0.031	0.057
	B_4^{NQE}	0.027	0.017	0.031	0.057
	$HT_{0.1}^{emp} + B_3^{emp}$	0.027	0.017	0.031	0.057
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.033	0.017	0.033	0.116
	$HT_{0.5}^{emp} + B_3^{emp}$	0.026	0.017	0.031	0.053
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.039	0.017	0.032	0.169
	$HT_{0.9}^{emp} + B_3^{emp}$	0.026	0.017	0.031	0.053
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.057	0.017	0.037	0.325
	$LR_{exp} + B_3^{NQE}$	0.059	0.017	0.037	0.348
	$LR_{Par} + B_3^{NQE}$	0.059	0.017	0.037	0.348
	$LR_{cons} + B_3^{NQE}$	0.059	0.017	0.037	0.348
	$KS + B_3^{NQE}$	0.030	0.017	0.030	0.105

Table 46: Mean Relative Error: 0.25 Unif(0,10) + 0.75Exp(1), $\rho = 0.475$, $n = 1000$

	MRE	$\mathbb{E}[W]$ $= 4.0235$	$\mathbb{P}\{W \geq 20\}$ ≈ 0.0377	$\mathbb{P}\{W \geq 30\}$ ≈ 0.0095	$\mathbb{P}\{W \geq 40\}$ ≈ 0.0024
Theoretical	Fitted Exp	1.14	3.17	8.54	20.9
	Fitted Par	6.26E+64	25.5	104	419
	Empirical	0.181	0.357	0.538	0.743
Quasi-Empirical	NQE_{exp}	0.182	0.358	0.539	0.746
	NQE_{Par}	0.184	0.364	0.551	0.765
	CQE_{exp}	0.181	0.357	0.538	0.744
	CQE_{Par}	0.172	0.331	0.494	0.677
Hypothesis Test	$HT_{0.1}^{emp}$	0.184	0.364	0.551	0.765
	$HT_{0.1}^{NQE}$	0.184	0.364	0.551	0.765
	$HT_{0.5}^{emp}$	0.184	0.364	0.551	0.765
	$HT_{0.5}^{NQE}$	0.184	0.364	0.551	0.765
	$HT_{0.9}^{emp}$	0.184	0.364	0.551	0.765
	$HT_{0.9}^{NQE}$	0.184	0.364	0.551	0.765
Likelihood Ratio	LR_{exp}	0.182	0.358	0.539	0.746
	LR_{Par}	0.182	0.358	0.539	0.746
	LR_{cons}	0.182	0.358	0.539	0.746
Heuristics	KS	0.182	0.359	0.542	0.750
	B_3^{emp}	0.184	0.364	0.551	0.765
	B_3^{NQE}	0.184	0.364	0.551	0.765
	B_4^{emp}	0.184	0.364	0.551	0.765
	B_4^{NQE}	0.184	0.364	0.551	0.765
	$HT_{0.1}^{emp} + B_3^{emp}$	0.184	0.364	0.551	0.765
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.184	0.364	0.551	0.765
	$HT_{0.5}^{emp} + B_3^{emp}$	0.184	0.364	0.551	0.765
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.184	0.364	0.551	0.765
	$HT_{0.9}^{emp} + B_3^{emp}$	0.184	0.364	0.551	0.765
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.184	0.364	0.551	0.765
	$LR_{exp} + B_3^{NQE}$	0.182	0.358	0.539	0.746
	$LR_{Par} + B_3^{NQE}$	0.182	0.358	0.539	0.746
	$LR_{cons} + B_3^{NQE}$	0.182	0.358	0.539	0.746
	$KS + B_3^{NQE}$	0.182	0.359	0.542	0.750

Table 47: Mean Relative Error: 0.75 Unif(0,10) + 0.25Exp(1), $\rho = 0.325$, $n = 1000$

	MRE	$\mathbb{E}[W]$ $= 1.5431$	$\mathbb{P}\{W \geq 5\}$ ≈ 0.1259	$\mathbb{P}\{W \geq 10\}$ ≈ 0.0412	$\mathbb{P}\{W \geq 20\}$ ≈ 0.0035
Theoretical	Fitted Exp	1.03	0.534	1.79	10.5
	Fitted Par	2.37E+69	6.94	23.3	282
	Empirical	0.330	0.351	0.446	1.02
Quasi-Empirical	NQE_{exp}	0.330	0.351	0.446	1.02
	NQE_{Par}	0.332	0.352	0.449	1.04
	CQE_{exp}	0.330	0.351	0.446	1.02
	CQE_{Par}	0.318	0.343	0.417	0.929
Hypothesis Test	$HT_{0.1}^{emp}$	0.332	0.352	0.449	1.04
	$HT_{0.1}^{NQE}$	0.332	0.352	0.449	1.04
	$HT_{0.5}^{emp}$	0.332	0.352	0.449	1.04
	$HT_{0.5}^{NQE}$	0.332	0.352	0.449	1.04
	$HT_{0.9}^{emp}$	0.332	0.352	0.449	1.04
	$HT_{0.9}^{NQE}$	0.332	0.352	0.449	1.04
Likelihood Ratio	LR_{exp}	0.330	0.351	0.446	1.02
	LR_{Par}	0.330	0.351	0.446	1.02
	LR_{cons}	0.330	0.351	0.446	1.02
Heuristics	KS	0.331	0.351	0.447	1.03
	B_3^{emp}	0.332	0.352	0.449	1.04
	B_3^{NQE}	0.332	0.352	0.449	1.04
	B_4^{emp}	0.332	0.352	0.449	1.04
	B_4^{NQE}	0.332	0.352	0.449	1.04
	$HT_{0.1}^{emp} + B_3^{emp}$	0.332	0.352	0.449	1.04
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.332	0.352	0.449	1.04
	$HT_{0.5}^{emp} + B_3^{emp}$	0.332	0.352	0.449	1.04
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.332	0.352	0.449	1.04
	$HT_{0.9}^{emp} + B_3^{emp}$	0.332	0.352	0.449	1.04
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.332	0.352	0.449	1.04
	$LR_{exp} + B_3^{NQE}$	0.330	0.351	0.446	1.02
	$LR_{Par} + B_3^{NQE}$	0.330	0.351	0.446	1.02
	$LR_{cons} + B_3^{NQE}$	0.330	0.351	0.446	1.02
	$KS + B_3^{NQE}$	0.331	0.351	0.447	1.03

Table 48: Mean Relative Error: 0.25 Unif(0,10) + 0.75 Par(10,4), $\rho = 0.625$, $n = 1000$

	MRE	$\mathbb{E}[W]$ $= 13.352$	$\mathbb{P}\{W \geq 50\}$ ≈ 0.0593	$\mathbb{P}\{W \geq 70\}$ ≈ 0.0283	$\mathbb{P}\{W \geq 90\}$ ≈ 0.0149
Theoretical	Fitted Exp	0.576	1.38	1.75	1.88
	Fitted Par	1.65E+86	15.9	34.3	66.1
	Empirical	0.178	0.336	0.451	0.602
Quasi-Empirical	NQE_{exp}	0.090	0.200	0.274	0.390
	NQE_{Par}	0.191	0.315	0.394	0.459
	CQE_{exp}	0.177	0.336	0.451	0.593
	CQE_{Par}	0.074	0.194	0.386	0.576
Hypothesis Test	$HT_{0.1}^{emp}$	0.191	0.315	0.394	0.459
	$HT_{0.1}^{NQE}$	0.191	0.315	0.394	0.459
	$HT_{0.5}^{emp}$	0.191	0.315	0.394	0.459
	$HT_{0.5}^{NQE}$	0.191	0.315	0.394	0.459
	$HT_{0.9}^{emp}$	0.191	0.315	0.394	0.459
	$HT_{0.9}^{NQE}$	0.191	0.315	0.394	0.459
Likelihood Ratio	LR_{exp}	0.095	0.208	0.283	0.395
	LR_{Par}	0.189	0.313	0.391	0.458
	LR_{cons}	0.095	0.208	0.283	0.395
Heuristics	KS	0.191	0.315	0.394	0.459
	B_3^{emp}	0.191	0.315	0.394	0.459
	B_3^{NQE}	0.191	0.315	0.394	0.459
	B_4^{emp}	0.177	0.320	0.403	0.502
	B_4^{NQE}	0.104	0.197	0.221	0.273
	$HT_{0.1}^{emp} + B_3^{emp}$	0.191	0.315	0.394	0.459
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.191	0.315	0.394	0.459
	$HT_{0.5}^{emp} + B_3^{emp}$	0.191	0.315	0.394	0.459
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.191	0.315	0.394	0.459
	$HT_{0.9}^{emp} + B_3^{emp}$	0.191	0.315	0.394	0.459
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.191	0.315	0.394	0.459
	$LR_{exp} + B_3^{NQE}$	0.095	0.208	0.283	0.395
	$LR_{Par} + B_3^{NQE}$	0.189	0.313	0.391	0.458
	$LR_{cons} + B_3^{NQE}$	0.095	0.208	0.283	0.395
	$KS + B_3^{NQE}$	0.191	0.315	0.394	0.459

Table 49: Mean Relative Error: 0.75 Unif(0,10) + 0.25Par(10,4), $\rho = 0.375$, $n = 1000$

	MRE	$\mathbb{E}[W]$ $= 3.3107$	$\mathbb{P}\{W \geq 20\}$ ≈ 0.0384	$\mathbb{P}\{W \geq 30\}$ ≈ 0.0173	$\mathbb{P}\{W \geq 40\}$ ≈ 0.0091
Theoretical	Fitted Exp	0.357	0.839	0.779	0.464
	Fitted Par	6.08E+70	25.1	57.0	109
	Empirical	0.225	0.237	0.306	0.460
Quasi-Empirical	NQE_{exp}	0.093	0.171	0.242	0.471
	NQE_{Par}	0.234	0.253	0.303	0.361
	CQE_{exp}	0.198	0.237	0.306	0.457
	CQE_{Par}	0.051	0.151	0.433	0.682
Hypothesis Test	$HT_{0.1}^{emp}$	0.234	0.253	0.303	0.361
	$HT_{0.1}^{NQE}$	0.234	0.253	0.303	0.361
	$HT_{0.5}^{emp}$	0.234	0.253	0.303	0.361
	$HT_{0.5}^{NQE}$	0.234	0.253	0.303	0.361
	$HT_{0.9}^{emp}$	0.234	0.253	0.303	0.361
	$HT_{0.9}^{NQE}$	0.234	0.253	0.303	0.361
Likelihood Ratio	LR_{exp}	0.093	0.171	0.242	0.471
	LR_{Par}	0.186	0.226	0.311	0.430
	LR_{cons}	0.093	0.171	0.242	0.471
Heuristics	KS	0.231	0.252	0.306	0.370
	B_3^{emp}	0.234	0.253	0.303	0.361
	B_3^{NQE}	0.234	0.253	0.303	0.361
	B_4^{emp}	0.233	0.235	0.275	0.367
	B_4^{NQE}	0.131	0.189	0.155	0.263
	$HT_{0.1}^{emp} + B_3^{emp}$	0.234	0.253	0.303	0.361
	$HT_{0.1}^{NQE} + B_3^{NQE}$	0.234	0.253	0.303	0.361
	$HT_{0.5}^{emp} + B_3^{emp}$	0.234	0.253	0.303	0.361
	$HT_{0.5}^{NQE} + B_3^{NQE}$	0.234	0.253	0.303	0.361
	$HT_{0.9}^{emp} + B_3^{emp}$	0.234	0.253	0.303	0.361
	$HT_{0.9}^{NQE} + B_3^{NQE}$	0.234	0.253	0.303	0.361
	$LR_{exp} + B_3^{NQE}$	0.093	0.171	0.242	0.471
	$LR_{Par} + B_3^{NQE}$	0.186	0.226	0.311	0.430
	$LR_{cons} + B_3^{NQE}$	0.093	0.171	0.242	0.471
	$KS + B_3^{NQE}$	0.231	0.252	0.306	0.370

REFERENCES

[1] Abate, J., Choudhury, G. L., and Whitt, W., "Waiting-time tail probabilities in queues with long-tail service-time distributions," *Queueing Systems*, vol. 16, no. 3-4, pp. 311–338, 1994.

[2] Alves, M. F., "A location invariant hill-type estimator," *Extremes*, vol. 4, no. 3, pp. 199–217, 2001.

[3] Andradóttir, S. and Kim, S.-H., "Fully sequential procedures for comparing constrained systems via simulation," *Naval Research Logistics*, vol. 57, no. 5, pp. 403–421, 2010.

[4] Asmussen, S. and Glynn, P., *Stochastic Simulation: Algorithms and Analysis*. Springer Verlag, NY, 2007.

[5] Asmussen, S. and Kroese, D. P., "Improved algorithms for rare event simulation with heavy tails," *Advances in Applied Probability*, vol. 38, no. 2, pp. 545–558, 2006.

[6] Baesler, F. F. and Sepúlveda, J. A., "Simulation optimization: Multi-response simulation optimization using stochastic genetic search within a goal programming framework," in *Proceedings of the 2000 Winter Simulation Conference*, pp. 788–794, IEEE Computer Society, 2000.

[7] Baesler, F. F. and Sepúlveda, J. A., "Multi-objective simulation optimization for a cancer treatment center," in *Proceedings of the 2001 Winter Simulation Conference*, pp. 1405–1411, IEEE Computer Society, 2001.

[8] Banks, J., Carson, J., Nelson, B. L., and Nicol, D. M., *Discrete-Event System Simulation*. Prentice Hall, fifth ed., 2010.

[9] Barabasi, A.-L., "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, no. 7039, pp. 207–211, 2005.

[10] Batur, D. and Choobineh, F. F., "Mean-variance based ranking and selection," in *Proceedings of the 2010 Winter Simulation Conference*, pp. 1160–1166, IEEE Computer Society, 2010.

[11] Bayes, A., "Statistical techniques for simulation models," *Australian Computer Journal*, vol. 2, no. 4, pp. 180–184, 1970.

[12] Bechhofer, R., "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *The Annals of Mathematical Statistics*, vol. 25, no. 1, pp. 16–39, 1954.

[13] Bechhofer, R., Elmaghraby, S., and Morse, N., "A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability," *The Annals of Mathematical Statistics*, pp. 102–119, 1959.

[14] Boxma, O. J. and Cohen, J., "The M/G/1 queue with heavy-tailed service time distribution," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 5, pp. 749–763, 1998.

[15] Bratley, P., Fox, B. L., and Schrage, L. E., *A Guide to Simulation*. Springer-Verlag, second ed., 1983.

[16] Butler, J., Morrice, D., and Mullarkey, P., "A multiple attribute utility theory approach to ranking and selection," *Management Science*, vol. 47, no. 6, pp. 800–816, 2001.

[17] Cérou, F. and Guyader, A., "Adaptive multilevel splitting for rare event analysis," *Stochastic Analysis and Applications*, vol. 25, no. 2, pp. 417–443, 2007.

[18] Cérou, F., LeGland, F., Moral, P., and Lezaud, P., "Limit theorems for the multilevel splitting algorithm in the simulation of rare events," in *Proceedings of the 37th Winter Simulation Conference*, pp. 682–691, IEEE Computer Society, 2005.

[19] Chen, C., "An effective approach to smartly allocate computing budget for discrete event simulation," in *Proceedings of the 34th IEEE Conference on Decision and Control*, vol. 3, pp. 2598–2603, IEEE, 1995.

[20] Chen, C., "A lower bound for the correct subset-selection probability and its application to discrete-event system simulations," *IEEE Transactions on Automatic Control*, vol. 41, no. 8, pp. 1227–1231, 1996.

[21] Chen, C., He, D., Fu, M., and Lee, L., "Efficient simulation budget allocation for selecting an optimal subset," *INFORMS Journal on Computing*, vol. 20, no. 4, pp. 579–595, 2008.

[22] Chen, C., Lin, J., Yücesan, E., and Chick, S., "Simulation budget allocation for further enhancing the efficiency of ordinal optimization," *Discrete Event Dynamic Systems*, vol. 10, no. 3, pp. 251–270, 2000.

[23] Chen, D., Turner, S., Cai, W., Gan, B., and Low, M., "Algorithms for HLA-based distributed simulation cloning," *ACM Transactions on Modeling and Computer Simulation*, vol. 15, no. 4, pp. 316–345, 2005.

[24] Chen, D., Turner, S., Cai, W., and Xiong, M., "A decoupled federate architecture for high level architecture-based distributed simulation," *Journal of Parallel and Distributed Computing*, vol. 68, no. 11, pp. 1487–1503, 2008.

[25] Chen, E. and Lee, L., "A multi-objective selection procedure of determining a Pareto set," *Computers & Operations Research*, vol. 36, no. 6, pp. 1872–1879, 2009.

[26] Chen, H., Dai, L., Chen, C., and Yücesan, E., "New development of optimal computing budget allocation for discrete event simulation," in *Proceedings of the 1997 Winter Simulation Conference*, pp. 334–341, IEEE Computer Society, 1997.

[27] Chick, S., "New two-stage and sequential procedures for selecting the best simulated system," *Operations Research*, vol. 49, no. 5, pp. 732–743, 2001.

[28] Chick, S. and Inoue, K., "New procedures to select the best simulated system using common random numbers," *Management Science*, vol. 47, no. 8, pp. 1133–1149, 2001.

[29] Clauset, A., Shalizi, C. R., and Newman, M. E., "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661–703, 2009.

[30] De Haan, L. and Peng, L., "Comparison of tail index estimators," *Statistica Neerlandica*, vol. 52, no. 1, pp. 60–70, 1998.

[31] Del Moral, P., *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer Verlag, 2004.

[32] Del Moral, P. and Miclo, L., "Genealogies and increasing propagation of chaos for Feynman-Kac and genetic models," *Annals of Applied Probability*, vol. 11, no. 4, pp. 1166–1198, 2001.

[33] Dudewicz, E. and Dalal, S., "Allocation of observations in ranking and selection with unequal variances," *Sankhyā: The Indian Journal of Statistics, Series B*, pp. 28–78, 1975.

[34] Embrechts, P., Klüppelberg, C., and Mikosch, T., *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.

[35] Feldman, R. and Taqqu, M., *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Springer, 1998.

[36] Feldmann, A. and Whitt, W., "Fitting mixtures of exponentials to long-tail distributions to analyze network performance models," in *IEEE INFOCOM'97*, vol. 3, pp. 1096–1104, IEEE, 1997.

[37] Fishman, G. S., *Discrete-event Simulation: Modeling, Programming, and Analysis*. Springer, first ed., 2001.

[38] Frazier, P. I., Powell, W. B., and Dayanik, S., "A knowledge-gradient policy for sequential information collection," *SIAM Journal on Control and Optimization*, vol. 47, no. 5, pp. 2410–2439, 2008.

[39] Garvels, M., *The Splitting Method in Rare Event Simulation.* PhD thesis, Universiteit Twente, Enschede, October 2000.

[40] Garvels, M. and Kroese, D., "A comparison of RESTART implementations," in *Proceedings of the 30th Winter Simulation Conference*, pp. 601–608, IEEE Computer Society, 1998.

[41] Garvels, M., Van Ommeren, J., and Kroese, D., "On the importance function in splitting simulation," *European Transactions on Telecommunications*, vol. 13, no. 4, pp. 363–371, 2002.

[42] Gillespie, C. S., *Fitting Heavy Tailed Distributions: The poweRlaw Package*, 2013. R package version 0.20.2, available at `https://github.com/csgillespie/poweRlaw`, date accessed: Nov 2013.

[43] Glasserman, P., Heidelberger, P., Shahabuddin, P., and Zajic, T., "Multilevel splitting for estimating rare event probabilities," *Operations Research*, vol. 47, no. 4, pp. 585–600, 1999.

[44] Glasserman, P., Heidelberger, P., Shahabuddin, P., and Zajic, T., "A large deviations perspective on the efficiency of multilevel splitting," *IEEE Transactions on Automatic Control*, vol. 43, no. 12, pp. 1666–1679, 2002.

[45] Glasserman, P. and Staum, J., "Resource allocation among simulation time steps," *Operations Research*, vol. 51, no. 6, pp. 908–921, 2003.

[46] Glynn, P. and Juneja, S., "A large deviations perspective on ordinal optimization," in *Proceedings of the 2004 Winter Simulation Conference*, vol. 1, pp. 577–585, IEEE Computer Society, 2004.

[47] Glynn, P. and Whitt, W., "The asymptotic efficiency of simulation estimators," *Operations Research*, vol. 40, no. 3, pp. 505–520, 1992.

[48] Gomes, M. I., De Haan, L., and Rodrigues, L. H., "Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 31–52, 2008.

[49] Görg, C. and Schreiber, F., "The RESTART/LRE method for rare event simulation," in *Proceedings of the 28th Winter Simulation Conference*, pp. 390–397, IEEE Computer Society, 1996.

[50] Hartmann, M., "An improvement on Paulson's procedure for selecting the poprlation with the largest mean from k normal populations with a common unknown variance," *Sequential Analysis*, vol. 10, no. 1-2, pp. 1–16, 1991.

[51] Healey, C., Andradóttir, S., and Kim, S.-H., "Efficient comparison of constrained systems using dormancy," *European Journal of Operational Research*, vol. 224, no. 2, pp. 340–352, 2013.

[52] Healey, C., Andradóttir, S., and Kim, S.-H., "Selection procedures for simulations with multiple constraints under independent and correlated sampling," *ACM Transactions on Modeling and Computer Simulation*, to appear.

[53] Henderson, S., Chiera, B., and Cooke, R., "Generating dependent quasi-random numbers," in *Proceedings of the 32nd Winter Simulation Conference*, vol. 1, pp. 527–536, IEEE Computer Society, 2000.

[54] Heyde, C. C. and Kou, S. G., "On the controversy over tailweight of distributions," *Operations Research Letters*, vol. 32, no. 5, pp. 399–408, 2004.

[55] Hill, B. M., "A simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, vol. 3, no. 5, pp. 1163–1174, 1975.

[56] Hopmans, A. and Kleijnen, J., "Importance sampling in systems simulation: A practical failure?," *Mathematics and Computers in Simulation*, vol. 21, no. 2, pp. 209–220, 1979.

[57] Hunter, S. R. and Pasupathy, R., "Optimal sampling laws for stochastically constrained simulation optimization on finite sets," *INFORMS Journal on Computing*, vol. 25, no. 3, pp. 527–542, 2013.

[58] Hunter, S. R., Pujowidianto, N. A., Pasupathy, R., Lee, L. H., and Chen, C.-H., "Constrained and correlated simulation optimization on finite sets," *Under Review*, 2012.

[59] Hybinette, M., "Just-in-time cloning," in *Proceedings of the 18th Workshop on Parallel and Distributed Simulation*, pp. 45–51, ACM, 2004.

[60] Hybinette, M. and Fujimoto, R., "Cloning parallel simulations," *ACM Transactions on Modeling and Computer Simulation*, vol. 11, no. 4, pp. 378–407, 2001.

[61] Hybinette, M. and Fujimoto, R., "Scalability of parallel simulation cloning," in *Proceedings of the 35th Annual Simulation Symposium*, pp. 275–282, IEEE, 2002.

[62] Kahn, H. and Harris, T., "Estimation of particle transmission by random sampling," *National Bureau of Standards Applied Mathematics Series*, vol. 12, pp. 27–30, 1951.

[63] Kim, S.-H. and Nelson, B., "A fully sequential procedure for indifference-zone selection in simulation," *ACM Transactions on Modeling and Computer Simulation*, vol. 11, no. 3, pp. 251–273, 2001.

[64] Lagnoux, A., "Rare event simulation," *Probability in the Engineering and Informational Sciences*, vol. 20, no. 01, pp. 45–66, 2006.

[65] Lagnoux-Renaudie, A., "Effective branching splitting method under cost constraint," *Stochastic Processes and their Applications*, vol. 118, no. 10, pp. 1820–1851, 2008.

[66] Lagnoux-Renaudie, A., "A two-step branching splitting model under cost constraint for rare event analysis," *Journal of Applied Probability*, vol. 46, no. 2, pp. 429–452, 2009.

[67] Law, A. M. and Kelton, W. D., *Simulation Modeling and Analysis*. McGraw-Hill, third ed., 1997.

[68] L'Ecuyer, P., Demers, V., and Tuffin, B., "Splitting for rare-event simulation," in *Proceedings of the 38th Winter Simulation Conference*, pp. 137–148, IEEE Computer Society, 2006.

[69] L'Ecuyer, P., Demers, V., and Tuffin, B., "Rare events, splitting, and quasi-Monte Carlo," *ACM Transactions on Modeling and Computer Simulation*, vol. 17, no. 2, p. Article 9, 2007.

[70] Lee, L., Pujowidianto, N., Li, L., Chen, C., and Yap, C., "Approximate simulation budget allocation for selecting the best design in the presence of stochastic constraints," *IEEE Transactions on Automatic Control*, vol. 57, pp. 2940 – 2945, 2012.

[71] Lee, L., Chew, E., and Teng, S., "Integration of statistical selection with search mechanism for solving multi-objective simulation-optimization problems," in *Proceedings of the 2006 Winter Simulation Conference*, pp. 294–303, IEEE Computer Society, 2006.

[72] Lee, L., Chew, E., Teng, S., and Goldsman, D., "Finding the non-dominated Pareto set for multi-objective simulation models," *IIE Transactions*, vol. 42, no. 9, pp. 656–674, 2010.

[73] Lindley, D. V., "The theory of queues with a single server," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 48, pp. 277–289, 1952.

[74] Mandelbrot, B., "The variation of some other speculative prices," *Journal of Business*, vol. 40, no. 4, pp. 393–413, 1967.

[75] Miller, J., Nelson, B., and Reilly, C., "Getting more from the data in a multinomial selection problem," in *Proceedings of the 1996 Winter Simulation Conference*, pp. 287–294, IEEE Computer Society, 1996.

[76] Müller, U. A., Dacorogna, M. M., and Pictet, O. V., "Heavy tails in high-frequency financial data," in *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (Feldman, R. and Taqqu, M., eds.), Birkhaüser, 1998.

[77] Nelson, B., Swann, J., Goldsman, D., and Song, W., "Simple procedures for selecting the best simulated system when the number of alternatives is large," *Operations Research*, vol. 49, no. 6, pp. 950–963, 2001.

[78] Newman, M. E., "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, 2005.

[79] Nolan, J., *Stable Distributions: Models for Heavy-tailed Data*. Birkhauser, 2003.

[80] Paulson, E., "A sequential procedure for selecting the population with the largest mean from k normal populations," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 174–180, 1964.

[81] Resnick, S. and Stărică, C., "Smoothing the hill estimator," *Advances in Applied Probability*, vol. 29, no. 1, pp. 271–293, 1997.

[82] Resnick, S. I., "Heavy tail modeling and teletraffic data: Special invited paper," *The Annals of Statistics*, vol. 25, no. 5, pp. 1805–1869, 1997.

[83] Resnick, S. I., *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. Springer, 2007.

[84] Rinott, Y., "On two-stage selection procedures and related probability-inequalities," *Communications in Statistics–Theory and Methods*, vol. 7, no. 8, pp. 799–811, 1978.

[85] Rubinstein, R. and Kroese, D., *Simulation and the Monte Carlo Method*. Wiley, Hoboken, NJ, 2007.

[86] Santner, T. and Tamhane, A., "Designing experiments for selecting a normal population with a large mean and a small variance," tech. rep., Cornell University, Operations Research and Industrial Engineering, 1983.

[87] Sigman, K., "Appendix: A primer on heavy-tailed distributions," *Queueing Systems*, vol. 33, no. 1, pp. 261–275, 1999.

[88] Teng, S., Lee, L., and Chew, E., "Integration of indifference-zone with multi-objective computing budget allocation," *European Journal of Operational Research*, vol. 203, no. 2, pp. 419–429, 2010.

[89] Villén-Altamirano, M. and Villén-Altamirano, J., "Accelerated simulation of rare events using RESTART method with hysteresis," in *Proceedings of the International Teletraffic Conference Specialists' Seminar on Telecommunication Services for Developing Economies. University of Mining and Metallurgy, Krakov, Poland*, pp. 675–686, elsevier, 1991.

[90] Villén-Altamirano, M. and Villén-Altamirano, J., "RESTART: a straightforward method for fast simulation of rare events," in *Proceedings of the 26th Winter Simulation Conference*, pp. 282–289, IEEE Computer Society, 1994.

[91] Villén-Altamirano, M. and Villen-Altamirano, J., "Analysis of RESTART simulation: Theoretical basis and sensitivity study," *European Transactions on Telecommunications*, vol. 13, no. 4, pp. 373–385, 2002.

[92] Villén-Altamirano, M. and Villén-Altamirano, J., "On the efficiency of RESTART for multidimensional state systems," *ACM Transactions on Modeling and Computer Simulation*, vol. 16, no. 3, pp. 251–279, 2006.

[93] Villén-Altamirano, M. and Villén-Altamirano, J., "The Rare Event Simulation Method RESTART: Efficiency Analysis and Guidelines for Its Application," *Network Performance Engineering*, pp. 509–547, 2011.

[94] Vuong, Q. H., "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica*, vol. 57, no. 2, pp. 307–333, 1989.

[95] Weissman, I., "Estimation of parameters and large quantiles based on the k largest observations," *Journal of the American Statistical Association*, vol. 73, no. 364, pp. 812–815, 1978.

[96] Willinger, W., Taqqu, M. S., Leland, W. E., and Wilson, D. V., "Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements," *Statistical Science*, vol. 10, no. 1, pp. 67–85, 1995.