DECISION MAKING IN THE PRESENCE OF SUBJECTIVE STOCHASTIC CONSTRAINTS

A Thesis
Presented to
The Academic Faculty

by

Yuwei Zhou

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

Georgia Institute of Technology
December 2021

DECISION MAKING IN THE PRESENCE OF SUBJECTIVE STOCHASTIC CONSTRAINTS

Approved by:

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

Professor Seong-Hee Kim, Advisor
School of Industrial and System Engineering
Georgia Institute of Technology

Professor Yajun Mei
School of Industrial and System Engineering
Georgia Institute of Technology

Professor Enlu Zhou
School of Industrial and System Engineering
Georgia Institute of Technology

Professor Chuljin Park
Department of Industrial Engineering
Hanyang University

Date Approved: 1 October 2021

There is only one true heroism in the world: to see the world as it is, and to love it.

— Roman Rolland

To my family and friends.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my research advisors, Dr. Sigrún Andradóttir and Dr. Seong-Hee Kim for their caring, support, and guidance during my Ph.D. study. Their encouragement and patience were always key for me to conquer all the obstacles. I am also extremely grateful for their thoughtful suggestions for my research projects and meaningful advice for my future career. Their passion and persistence for research would be a forever inspiration for me, both for life and career. I am fortunate to be advised by them.

I also hope to extend my thanks to Dr. Enlu Zhou, Dr. Yajun Mei, and Dr. Chuljin Park for serving on my thesis committee and providing valuable advice during my proposal and defense. Their detailed and highly constructive comments help this thesis become more complete. I want to especially thank Dr. Chuljin Park for being a co-author. His effort and feedback were critical as the research project was born and matured.

Besides my thesis committee, I am indebted to all the faculty members for their help and support. I am fortunate to take many excellent courses which were helpful for my research. Special thanks go to Dr. Brani Vidakovic and Dr. Roshan Joseph for their constant help and guidance when I served as the head teaching assistant for their course for two and a half years. The experience was invaluable for my future career.

Life during my Ph.D. study would be much more difficult without my wonderful friends. I am grateful for their caring, encouragement, and support. I want to especially thank my fellow students Yuang Chen for his generous support and valuable advice, Kaizhao Sun for those great times of stress relief activities, Pornpawee Bumpensanti for cheering me up when I was stressed, Guanlin Chen for frequently sharing life stories together, and Xinyu Liu for warm encouragement. I also appreciate the constant support from my friends Liangci Xue, Dongyu Lin, and Min Gong for being on my side and helping me whenever I needed it. I am fortunate to have those friends in my life.

Last but not the least, my sincere thanks go to all my family members. I want to thank my paternal grandparents, Peitian Wo and Yikang Zhou, and maternal grandparents, Lanying Liu and Mingli Luo. I truly appreciate the love and support they gave me ever since my childhood and feel sorry for not being able to accompany them frequently. Most important, I would like to give special thanks to my parents Zhiyun Luo and Hao Zhou for their unconditional love and for always being supportive of my decisions. Whenever I faced difficulty, felt depressed and homesick, they were always by my side. This thesis could not be done without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Constrained Ranking and Selection considers optimizing a primary performance measure over a finite set of alternatives subject to constraints on secondary performance measures. When the constraints are stochastic, the corresponding performance measures should be estimated by simulation. When the constraints are subjective, the decision maker is willing to consider multiple constraint threshold values. In this thesis, we consider three problem formulations when subjective stochastic constraints are present.

In Chapter 2, we consider the problem of finding a set of feasible or near-feasible systems among a finite number of simulated systems in the presence of subjective stochastic constraints. A decision maker may want to test multiple constraint threshold values for the feasibility check, or she may want to determine how a set of feasible systems changes as constraints become more strict with the objective of pruning systems or finding the system with the best performance. We present indifference-zone procedures that recycle observations for the feasibility check and provide an overall probability of correct decision for all threshold values. Our numerical experiments show that the proposed procedures perform well in reducing the required number of observations relative to four alternative procedures (that either restart feasibility check from scratch with respect to each set of thresholds or with the Bonferroni inequality applied in a conservative way) while providing a statistical guarantee on the probability of correct decision.

Chapter 3 considers the problem of finding a system with the best primary performance measure among a finite number of simulated systems in the presence of subjective stochastic constraints on secondary performance measures. When no feasible system exists, the decision maker may be willing to relax some constraint thresholds. We take multiple threshold values for each constraint as a user's input and propose indifference-zone procedures that perform the phases of feasibility check and selection-of-the-best sequentially or simultaneously. We prove that the proposed procedures yield the best system in the most desirable

feasible region possible with at least a pre-specified probability. Our experimental results show that our procedures perform well with respect to the number of observations required to make a decision, as compared with straightforward procedures that repeatedly solve the problem for each set of constraint thresholds.

In Chapter 4, we consider the problem of finding a portfolio of systems with the best primary performance measure among finitely many simulated systems as stochastic constraints on secondary performance measures are relaxed. By finding a portfolio of the best systems under a variety of constraint thresholds, the decision maker can identify a robust solution with respect to the constraints or consider the trade-off between the primary performance measure and the level of feasibility of the secondary performance measures. We propose indifference-zone procedures that perform the phases of feasibility check and selection-of-the-best sequentially and simultaneously, and prove that the proposed procedures identify the portolio of the best systems with at least a pre-specified probability. Our proposed procedures show a significant reduction in the required number of observations compared with straightforward procedures that repeatedly identify the best system with respect to each set of constraint thresholds.

CHAPTER I

INTRODUCTION

Analyzing the performance of stochastic systems, such as identifying a set of feasible systems or a system with the best performance, can be challenging due to the randomness in output data. Ranking and Selection (R&S) is one of the classical and actively studied problems in the simulation community. When a finite number of systems are considered and stochastic simulation is used to estimate the performance of the systems, R&S procedures are statistical approaches for feasibility determination and selection-of-the-best. Two major issues for R&S procedures are efficiency and validity, with the goal of reducing the number of necessary observations to reach a decision while guaranteeing a nominal probability of correct selection (PCS). [9] and [7] provide a good literature review of R&S.

When the problem requires not only selecting the best system with respect to a primary performance measure but also determining the feasibility with respect to stochastic constraints on secondary performance measures, it becomes constrained R&S. There are three major approaches in solving constrained R&S, namely the optimal computing budget allocation (OCBA) approach, the Bayesian approach, and the indifference-zone (IZ) approach. [13], [8], and [15] propose sampling frameworks that can approximate the OCBA considering stochastic constraints. [4] provide a procedure to solve a constrained R&S problem from the OCBA perspective. [18] discuss a fully sequential policy for allocating simulation effort to determine a set of several simulated systems that have mean performance exceeding a threshold of known value via a Bayesian formulation. Among the procedures that use the indifference-zone (IZ) approach, [1] propose constrained R&S procedures that find the system with the best primary performance measure among a finite number of simulated systems in the presence of a stochastic constraint on a single real-valued secondary performance measure. [2] propose fully sequential feasibility check procedures that are designed to find a feasible set in the presence of multiple constraints. [6] use dormancy to perform constrained R&S more efficiently and [5] extend the work of [1] to the problem of finding the best feasible system in the presence of multiple constraints. [14] propose an Adaptive Feasibility Checking Procedure (AFCP) that can self-adjust the tolerance level of all constraints to avoid the need for the decision maker to pre-determine IZ parameters.

Classical constrained R&S procedures consider stochastic constraints with fixed threshold values. If the decision maker wants to identify feasible systems with respect to a threshold, the procedures return a set of feasible systems or an empty set if there do not exist feasible systems with respect to the threshold. On the other hand, if she wants to select the best system in terms of the primary performance that also satisfies the constraints on the secondary performance measures, the procedures identify the best feasible system or declare there exists no feasible system with respect to the stochastic constraints. However, it is natural for a decision maker to consider varying threshold values instead of fixed values for the following reasons:

- When multiple systems are declared feasible with respect to the fixed threshold values, the decision maker may want to consider "tighter" threshold values in order to prune systems with worse performance on the stochastic constraints.

- When there do not exist feasible systems with respect to fixed threshold values and the decision maker is flexible with some constraint thresholds, she may want to consider "looser" threshold values such that some feasible systems may be identified.

- When the objective is to find the system with the best primary performance measure but no system satisfies the stochastic constraints on the secondary performance measures with respect to the fixed threshold values, the decision maker may want to relax the constraints so that she can identify the best system that is feasible to "looser" threshold values.

- When the decision maker is willing to consider multiple threshold values, she may want to identify a portfolio of the best systems with respect to each set of threshold values as this information is helpful in identifying a robust solution with respect to the constraints or considering the trade-off between the primary performance measure and the level of feasibility on the stochastic constraints.

For the problem of finding feasible systems, if each constraint threshold value is fixed to one constant, procedures such as [2] can be used. When the objective is to find the system with the best primary performance measure with fixed constraint thresholds, procedures due to [1] and [5] can be applied to systems with single or multiple constraints, respectively. When multiple threshold values are considered, one may restart a constrained R&S procedure "from scratch" for each threshold of each constraint and each system or over each possible combination of threshold values. However, this approach is inefficient in terms of computational costs because restarting from scratch wastes information contained in the observations collected from previously completed executions of a constrained R&S procedure with respect to different threshold values.

Given that there is no change in the simulation model of the systems under consideration, it is natural to re-use observations collected from a feasibility check with different sets of threshold values and obtain additional observations only when the changed set of threshold values requires more observations to reach a decision. However, the guarantee of PCS with respect to all the threshold values becomes questionable due to repeated tests on the same data sets. Although the Bonferroni inequality can be used to avoid dependence among tests, it makes the procedure overly conservative.

In this thesis, we introduce subjective constraints and consider the concept of recycling observations for three problem formulations when the threshold values change. We aim to develop procedures that avoid the overly conservative use of the Bonferroni inequality by exploiting the fact that the underlying distribution of observations does not change. To the best of our knowledge, [3] propose the idea of recycling simulation observations in computer experiments by using pre-existing simulation outputs to improve the efficiency of the current experiment when experiments are repeated periodically (e.g., a periodic credit risk evaluation problem). They consider importance sampling and multiple importance sampling as two particular output recycling implementations and show that multiple importance sampling is an effective and robust way to recycle and reuse pre-existing outputs. But their focus is on estimation rather than feasibility check.

The rest of the thesis is organized as follows: Chapter 2 proposes a procedure that performs feasibility checks using recycled observations when multiple thresholds are considered. Chapter 3 proposes procedures that select the best system in the most desirable feasible region possible by recycling simulation observations. In Chapter 4, we propose procedures that identify a portfolio of best systems when the stochastic constraints are relaxed. Directions for further research are provided in Chapter 5.

We also note that the notation used in each chapter is self-contained. As each chapter considers a different problem formulation, we redefine some notation within each chapter to avoid an overly complicated system of notation.

CHAPTER II

FINDING FEASIBLE SYSTEMS FOR SUBJECTIVE CONSTRAINTS USING RECYCLED OBSERVATIONS

2.1 Introduction

We consider the problem of finding a set of feasible systems among a finite number of simulated systems when the values of constraint thresholds vary. For example, a manager may want to control an inventory level with respect to two performance measures using an (s, S) inventory policy (namely ordering products to increase the inventory level to S when the inventory level at a review period is below s, with no order placed when the inventory level is greater or equal to s). Specifically, she wants to identify the feasible combinations of the values of s and S among finitely many choices such that the probability that a shortage occurs between two successive review periods is less than or equal to q_1, while the expected cost per review period is less than or equal to q_2 dollars, where q_1 and q_2 are some constant threshold values. The manager may want to try different values for q_1 (such as 0.01, 0.05, and 0.1) and q_2 (such as 115, 120, 125, 130, and 135) rather than two fixed values, and observe the effects on the feasible set. Changing threshold values can also be used for finding the best system. For example, if one wants to find a combination of the values of s and S that results in the smallest expected cost per review period, one can start feasibility check with a large threshold value such as 135 and then repeat the feasibility check by adjusting the threshold value until there is only one feasible solution left. The decision maker can also perform the same technique with the other performance measures, so that she can find her preferred solution with respect to all the performance measures.

In this chapter, we consider the concept of recycling observations in the context of finding a set of feasible or near-feasible systems when the threshold values change. Our aim is to develop a procedure that avoids the overly conservative use of the Bonferroni inequality by exploiting the fact that underlying distributions of observations do not change. To the best of our knowledge, [3] propose the idea of recycling simulation observations in computer experiments by using pre-existing simulation outputs to improve the efficiency of the current experiment when experiments are repeated periodically (e.g., a periodic credit risk evaluation problem). They consider importance sampling and multiple importance sampling as two particular output recycling implementations, and show that multiple importance sampling is an effective and robust way to recycle and reuse pre-existing outputs. But their focus is on estimation rather than comparison or feasibility check.

We propose a procedure that recycles all observations collected from the previous feasibility determinations (if needed to make a decision), prove that our procedure provides the desired overall probability of correct decision, and show that it saves a significant number of observations as compared with four alternative statistically valid procedures that perform feasibility checks with multiple thresholds: three with restart from scratch under different circumstances and one with recycled observations but with the Bonferroni inequality applied in a conservative way. In our procedure, we use the green simulation concept of recycling observations in consecutively applied feasibility checks. The specific contributions of the paper are three-fold, namely, we (1) suggest a summary statistic which can be used to implement feasibility check over all threshold values in consideration or sequentially added thresholds, (2) propose an efficient procedure based on the summary statistic when threshold values change, and (3) mathematically and empirically document the statistical guarantee and efficiency of the proposed procedure.

The rest of this paper is organized as follows: In Section 2.2, we define our problem and notation. Section 2.3 introduces our procedure. Section 2.4 then provides the statistical guarantee of our procedure. Experimental results for a single system or multiple systems with either one constraint or multiple correlated constraints and multiple threshold values are shown in Section 2.5, together with an illustration of how the procedure could be applied to solve the inventory control example mentioned at the beginning of this section. Concluding remarks are provided in Section 2.6. Finally, the Appendix contains some of our proofs and derivations, a discussion of a special case when applying our proposed procedure, and the descriptions and proofs of statistical validity of four alternative procedures.

2.2 Problem and Notation

In this section, we describe our problem and notation.

We consider k systems whose s performance measures can be estimated through stochastic simulation. Let Θ denote the index set of all possible systems (i.e., $\Theta = \{1, \ldots, k\}$). Let $Y_{i\ell n}$, where $i = 1, \ldots, k, \ell = 1, \ldots, s$, and $n = 1, 2, \ldots$, be the nth observation of the ith system for the ℓth performance measure. The expected value for system i regarding performance measure ℓ is denoted as $y_{i\ell} = \mathrm{E}[Y_{i\ell n}]$ and the variance for system i regarding performance measure ℓ is denoted as $\sigma^2_{i\ell} = \mathrm{Var}(Y_{i\ell n})$. Observations are assumed to satisfy the following normality assumption:

Assumption 1. *For each $i = 1, 2, \ldots, k$,*

$$\begin{bmatrix} Y_{i1n} \\ \vdots \\ Y_{isn} \end{bmatrix} \overset{iid}{\sim} N_s \left(\begin{bmatrix} y_{i1} \\ \vdots \\ y_{is} \end{bmatrix}, \Sigma_i \right), \qquad n = 1, 2, \ldots$$

where $\overset{iid}{\sim}$ denotes independent and identically distributed, N_s denotes s-dimensional multivariate normal, and Σ_i is the $s \times s$ covariance matrix of the vector $(Y_{i1n}, \ldots, Y_{isn})$.

Normally distributed observations are a common assumption used in many R&S procedures because Assumption 1 can be justified by the Central Limit Theorem when observations are either within-replication averages or batch means [12]. The observations of different performance measures from a system are usually correlated, such as throughput and expected waiting time for a queueing system, or expected cost and shortage probability in an inventory example.

For a given threshold vector $\boldsymbol{q} = (q_1, \ldots, q_s)$, [2] introduce procedure $\mathcal{F}_B$ to determine a set of systems with $y_{i\ell} \leq q_\ell$ for all $\ell = 1, 2, \ldots, s$. In this paper, instead of a fixed threshold vector $\boldsymbol{q}$, we consider a case where the threshold values change. We let d_ℓ denote the number of threshold values for performance measure ℓ that the decision maker is interested in, and let q_ℓ^m denote the threshold value for performance measure ℓ with index m, where $m = 1, \ldots, d_\ell$.

In order to check the feasibility of each system with respect to one constraint and a fixed threshold, namely q_ℓ^m, where $m = 1, \ldots, d_\ell$, Andradóttir and Kim (2010) introduce a tolerance level, denoted by ϵ_ℓ for constraint ℓ, that is a positive real number specified by the decision maker. Any system with $y_{i\ell} \leq q_\ell^m - \epsilon_\ell$ is considered desirable and feasible with respect to constraint ℓ and threshold q_ℓ^m. The set of all desirable systems with respect to constraint ℓ and threshold q_ℓ^m is denoted as $D_\ell(q_\ell^m)$. Systems with $y_{i\ell} \geq q_\ell^m + \epsilon_\ell$ are unacceptable and infeasible with respect to constraint ℓ and threshold q_ℓ^m, placing them in the set $U_\ell(q_\ell^m)$. Systems that fall within the tolerance level of q_ℓ^m, so that $q_\ell^m - \epsilon_\ell < y_{i\ell} < q_\ell^m + \epsilon_\ell$ are acceptable, and are placed in the set $A_\ell(q_\ell^m)$:

$$\begin{aligned} D_\ell(q_\ell^m) &= \{i \in \Theta \mid y_{i\ell} \leq q_\ell^m - \epsilon_\ell\}; \\ U_\ell(q_\ell^m) &= \{i \in \Theta \mid y_{i\ell} \geq q_\ell^m + \epsilon_\ell\}; \text{ and} \\ A_\ell(q_\ell^m) &= \{i \in \Theta \mid q_\ell^m - \epsilon_\ell < y_{i\ell} < q_\ell^m + \epsilon_\ell\}. \end{aligned}$$

When performing feasibility check, we use $\mathrm{CD}_{i\ell}(q_\ell^m)$ to denote a correct decision event of system i with respect to constraint ℓ for threshold q_ℓ^m, which is an event such that system i is declared to be feasible with respect to constraint ℓ if $i \in D_\ell(q_\ell^m)$ and infeasible if $i \in U_\ell(q_\ell^m)$. For $i \in A_\ell(q_\ell^m)$, any decision is considered as a correct decision.

We define $\mathrm{CD}_{i\ell}$, the correct decision event for system i with respect to constraint ℓ, as correctly determining feasibility for all possible thresholds q_ℓ^m where $m = 1, \ldots, d_\ell$, i.e. $\mathrm{CD}_{i\ell} = \cap_{m=1}^{d_\ell} \mathrm{CD}_{i\ell}(q_\ell^m)$. Then, a statistically-valid procedure that determines the feasibility for all combinations of the threshold values with respect to all performance measures should satisfy the following statement:

$$\mathrm{PCD} = \Pr\left(\cap_{i=1}^k \cap_{\ell=1}^s \mathrm{CD}_{i\ell}\right) \geq 1 - \alpha,$$

where α is the nominal confidence level for the feasibility check.

Throughout the paper, we need additional notation defined below:

$n_0 \equiv$ the initial sample size for each system ($n_0 \geq 2$);

$r_i \equiv$ the number of observations obtained so far for system i ($r_i \geq n_0$);

$S_{i\ell}^2(n_0) \equiv$ the sample variance of $Y_{i\ell 1}, \ldots, Y_{i\ell n_0}$ for system $i = 1, 2, \ldots, k$ and constraint

$$\ell = 1, 2, \ldots, s;$$

$$R(r_i; v, w, z) \equiv \max\left\{0, \frac{(n_0-1)wz}{v} - \frac{v}{2c} r_i\right\} \text{ for } v, w, z \in \mathbb{R}^+ \text{ and } c \in \{1, 2, \cdots, \infty\};$$

$$|\cdot| \equiv \text{the cardinality of a set.}$$

The feasibility check procedures we present use the non-negative function $R(r_i; \cdot)$ to specify an interval $\Big(-R(r_i; \cdot), R(r_i; \cdot)\Big)$ called the continuation region after r_i observations have been collected. To determine the continuation region, we need to choose the value of c (which will impact the value of w as will become clear in Section 2.3.1). The shape of the continuation region defined by $\Big(-R(r_i; \cdot), R(r_i; \cdot)\Big)$ becomes a longer and narrower triangle as c increases (Kim and Nelson, 2001) and eventually becomes two parallel lines $(-R_{i\ell}, R_{i\ell})$ for $c = \infty$, where $R_{i\ell} = \frac{(n_0-1)wz}{v}$ (the values of v, z depend on the system i and constraint ℓ as $v = \epsilon_\ell$ and $z = S^2_{i\ell}(n_0)$).

For feasibility check, we consider the monitoring statistic for system i and threshold q^m_ℓ on constraint ℓ as the cumulative sum of the difference between $Y_{i\ell n}$ and the threshold q^m_ℓ, i.e., $\sum_{n=1}^{r_i}(Y_{i\ell n} - q^m_\ell)$. When the monitoring statistic exits through the upper boundary of the continuation region, we declare system i is infeasible for constraint ℓ and threshold value q^m_ℓ. On the other hand, if the monitoring statistic exits through the lower boundary of the continuation region, system i is declared feasible for constraint ℓ and threshold value q^m_ℓ.

2.3 Feasibility Check Procedures with Recycled Observations

In this section, we discuss the generic procedure with recycled observations in Section 2.3.1, the implementation parameters for our proposed procedure in Section 2.3.2, and the addition of threshold values in Section 2.3.3.

2.3.1 Generic Procedure

In this section, we first discuss our approach for determining the feasibility of systems when the threshold vector changes by recycling observations obtained so far. We then present the details of our proposed procedure for multiple systems and multiple constraints.

For a given system i, constraint ℓ, and thresholds $\{q_\ell^1, q_\ell^2, \ldots, q_\ell^{d_\ell}\}$, the parameters $r_i, \epsilon_\ell, \eta_\ell$, and $S_{i\ell}^2(n_0)$ of $R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))$ do not depend on the threshold values. Moreover, $r_i, \epsilon_\ell, \eta_\ell$ are constants and $S_{i\ell}^2(n_0)$ does not change when we recycle observations for different thresholds. Thus $R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))$ remains the same.

For a given threshold q_ℓ^m, we declare system i as

$$\begin{cases} \text{feasible} & \text{if } \sum_{n=1}^{r_i}(Y_{i\ell n} - q_\ell^m) \leq -R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0)), \\ \text{infeasible} & \text{if } \sum_{n=1}^{r_i}(Y_{i\ell n} - q_\ell^m) \geq R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0)), \end{cases} \tag{1}$$

which is equivalent to declaring system i as

$$\begin{cases} \text{feasible} & \text{if } \bar{Y}_{i\ell}(r_i) + \dfrac{R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))}{r_i} \leq q_\ell^m, \\ \text{infeasible} & \text{if } \bar{Y}_{i\ell}(r_i) - \dfrac{R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))}{r_i} \geq q_\ell^m, \end{cases} \tag{2}$$

where we use $\bar{Y}_{i\ell}(r_i)$ to denote the average value of r_i observations taken from system i with respect to constraint ℓ, i.e., $\bar{Y}_{i\ell}(r_i) = \sum_{n=1}^{r_i} Y_{i\ell n}/r_i$. Even though (1) and (2) are equivalent, we choose (2) to determine the feasibility of each system with respect to each constraint for different thresholds as (1) requires shifting the sample path $\sum_{n=1}^{r_i}(Y_{i\ell n} - q_\ell^m)$ when the threshold value changes while (2) does not. Thus for a given set of thresholds $\{q_\ell^1, q_\ell^2, \ldots, q_\ell^{d_\ell}\}$ for constraint $\ell = 1, 2, \ldots, s$, we update $\bar{Y}_{i\ell}(r_i) \pm R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/r_i$ for system i as we get more observations and declare the system is feasible or infeasible with respect to threshold value q_ℓ^m when one of the inequalities in (2) is satisfied. In other words, we form an interval $\left(\bar{Y}_{i\ell}(r_i) - R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/r_i, \bar{Y}_{i\ell}(r_i) + R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/r_i\right)$ for each system with respect to each constraint and make feasibility check with respect to this constraint if the threshold falls outside of the interval. Otherwise, we take more observations and make a feasibility decision when (2) is satisfied.

Figure 1 shows the behavior of the interval $\left(\bar{Y}_{i\ell}(r_i) - R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/r_i, \bar{Y}_{i\ell}(r_i) + R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/r_i\right)$ as a function of the number of observations r_i, as well as feasibility decisions for two threshold values q_ℓ^1 and q_ℓ^2 with $q_\ell^1 < q_\ell^2$. We can easily see that the system i is declared infeasible with respect to q_ℓ^1 at r_i^1 observations as q_ℓ^1 falls below $\bar{Y}_{i\ell}(r_i^1) - R(r_i^1; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/r_i^1$, while the feasibility with respect to q_ℓ^2 is undeclared at time r_i^1 because q_ℓ^2 is still within the interval $\Big(\bar{Y}_{i\ell}(r_i^1) - R(r_i^1;\epsilon_\ell,\eta_\ell,S_{i\ell}^2(n_0))/r_i^1, \bar{Y}_{i\ell}(r_i^1) + R(r_i^1;\epsilon_\ell,\eta_\ell,S_{i\ell}^2(n_0))/r_i^1\Big)$. The feasibility decision with respect to q_ℓ^2 is made at r_i^2 observations as q_ℓ^2 is above $\bar{Y}_{i\ell}(r_i^2) + R(r_i^2;\epsilon_\ell,\eta_\ell,S_{i\ell}^2(n_0))/r_i^2$.

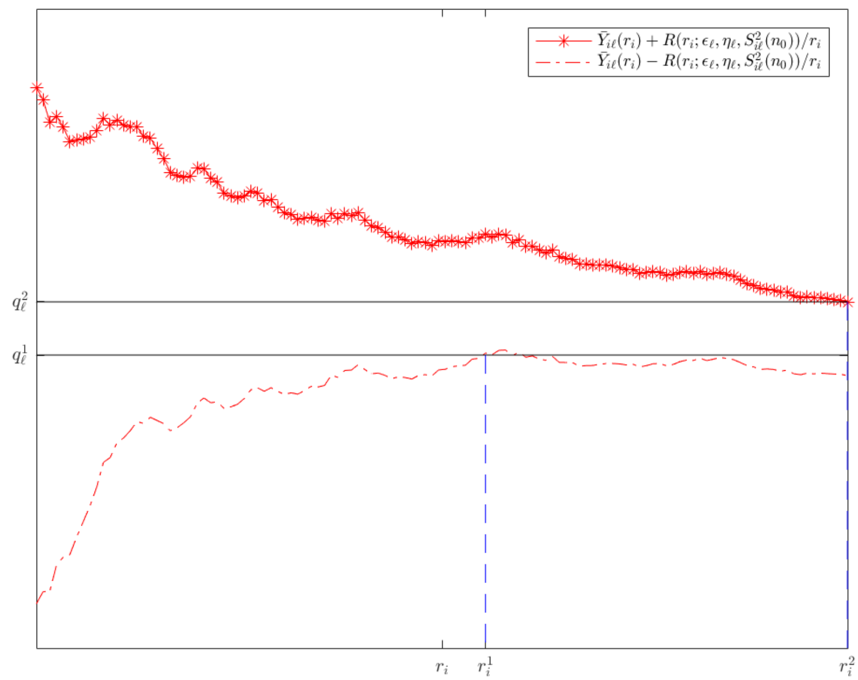


Figure 1: Changes in the interval $\Big(\bar{Y}_{i\ell}(r_i) - R(r_i;\epsilon_\ell,\eta_\ell,S_{i\ell}^2(n_0))/r_i, \bar{Y}_{i\ell} + R(r_i;\epsilon_\ell,\eta_\ell,S_{i\ell}^2(n_0))/r_i\Big)$

In Section 2.3.3, we discuss the addition of new threshold values after the execution of Algorithm 1. This suggests that we might want to keep all the values of $\bar{Y}_{i\ell}(r_i)$ until the feasibility determination is completed. However, this is not desirable due to a data storage problem. Instead, we keep the following two quantities while system i is simulated:

$$v_{i\ell}^{\mathrm{UB}} \equiv \min\left\{ \bar{Y}_{i\ell}(r') + \frac{R(r';\epsilon_\ell,\eta_\ell,S_{i\ell}^2(n_0))}{r'} \,\middle|\, n_0 \le r' \le r_i \right\};$$
$$v_{i\ell}^{\mathrm{LB}} \equiv \max\left\{ \bar{Y}_{i\ell}(r') - \frac{R(r';\epsilon_\ell,\eta_\ell,S_{i\ell}^2(n_0))}{r'} \,\middle|\, n_0 \le r' \le r_i \right\}.$$

Before we give the full description of our procedure, we need two additional definitions:

$$g(\eta) \equiv \begin{cases} \sum_{j=1}^{c}(-1)^{j+1}\left(1-\frac{1}{2}\mathcal{I}(j=c)\right)\times\left(1+\frac{2\eta(2c-j)j}{c}\right)^{-(n_0-1)/2}, & c\in\mathbb{N}^+, \\ \int_0^\infty \frac{1}{1+\exp(2\eta x)}\times\frac{1}{2^{(n_0-1)/2}\Gamma((n_0-1)/2)}x^{(n_0-1)/2-1}e^{-x/2}dx, & c=\infty, \end{cases} \quad (3)$$

$$\beta \equiv \begin{cases} 1-(1-\alpha)^{1/k}, & \text{when systems are independent,} \\ \alpha/k, & \text{when systems are dependent,} \end{cases} \quad (4)$$

where $\Gamma(\cdot)$ is the gamma function and $\mathcal{I}(\cdot)$ is the indicator function.

Algorithm 1 Procedure $\mathcal{RF}$

[**Setup**:] Choose confidence level $1-\alpha$, tolerance level ϵ_ℓ, and thresholds $\left\{q_\ell^1, q_\ell^2, \ldots, q_\ell^{d_\ell}\right\}$ for constraint $\ell = 1, 2, \ldots, s$. Also, choose the value of c and set $\Theta = \{1, 2, \ldots, k\}$. For $\ell = 1, \ldots, s$, set η_ℓ such that $g(\eta_\ell) = \beta_\ell$, where β satisfies (4), and either

(i) $\beta_\ell = (\beta/s)\cdot\mathcal{I}(d_\ell = 1) + [\beta/(2s)]\cdot\mathcal{I}(d_\ell > 1)$ for $\ell = 1, 2, \ldots, s$, or

(ii) $\beta_\ell = \beta/D$; $D = \sum_{\ell=1}^{s}\min\{d_\ell, 2\}$ for $\ell = 1, \ldots, s$.

for each system $i \in \Theta$ **do**

[**Initialization**:]

- Obtain n_0 observations $Y_{i\ell 1}, Y_{i\ell 2}, \ldots, Y_{i\ell n_0}$ for $\ell = 1, 2, \ldots, s$.
- Compute $\bar{Y}_{i\ell}(n_0)$ and $S_{i\ell}^2(n_0)$.
- Set $r_i = n_0$, $\text{ON} = \{1, 2, \ldots, s\}$, and $\text{ON}_\ell = \{1, 2, \ldots, d_\ell\}$ for $\ell = 1, 2, \ldots, s$.
- Set $v_{i\ell}^{\text{UB}} = \infty$ and $v_{i\ell}^{\text{LB}} = -\infty$ for $\ell = 1, 2, \ldots, s$.

[**Feasibility Check**:]

for $\ell \in \text{ON}$ **do**

$v_{i\ell}^{\text{UB}} = \min(v_{i\ell}^{\text{UB}}, \bar{Y}_{i\ell}(r_i) + R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/r_i)$. If $v_{i\ell}^{\text{UB}}$ is updated, set $\text{LAST}_{i\ell} = \text{UB}$.

$v_{i\ell}^{\text{LB}} = \max(v_{i\ell}^{\text{LB}}, \bar{Y}_{i\ell}(r_i) - R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/r_i)$. If $v_{i\ell}^{\text{LB}}$ is updated, set $\text{LAST}_{i\ell} = \text{LB}$.

for $m \in \text{ON}_\ell$ **do**,

If $v_{i\ell}^{\text{UB}} \le q_\ell^m$, set $Z_{i\ell}^m = 1$ and $\text{ON}_\ell = \text{ON}_\ell \setminus \{m\}$.
If $v_{i\ell}^{\text{LB}} \ge q_\ell^m$, set $Z_{i\ell}^m = 0$ and $\text{ON}_\ell = \text{ON}_\ell \setminus \{m\}$.

end for

If $\text{ON}_\ell = \emptyset$, set $\text{ON} = \text{ON} \setminus \{\ell\}$.

end for

[**Stopping Condition**:]

- If $\text{ON} = \emptyset$, return $Z_{i\ell}^m$ for $\ell = 1, 2, \ldots, s$ and $m = 1, 2, \ldots, d_\ell$.
- Otherwise, set $r_i = r_i + 1$, take one additional observation $Y_{i\ell r_i}$ and update $\bar{Y}_{i\ell}(r_i)$ for $\ell \in \text{ON}$, then go to [**Feasibility Check**].

end for

The full description of our procedure $\mathcal{RF}$ is provided in Algorithm 1 for k systems, s

constraints, and thresholds $\{q_\ell^1, q_\ell^2, \ldots, q_\ell^{d_\ell}\}$, where $\ell = 1, \ldots, s$. Note that if each constraint has one threshold value ($d_\ell = 1$ for all ℓ), then Procedure $\mathcal{RF}$ becomes the same as a feasibility check procedure presented in [5]. Also, in the full description of our procedure, the variable $Z_{i\ell}^m$ indicates the feasibility of system i with respect to threshold q_ℓ^m on constraint ℓ. More specifically, $Z_{i\ell}^m = 1$ means system i is feasible with respect to threshold q_ℓ^m on constraint ℓ, and $Z_{i\ell}^m = 0$ otherwise. To declare the feasibility of the system with respect to all the constraints for a given threshold vector $\boldsymbol{q} = (q_1^{m_1}, \ldots, q_s^{m_s})$, we check the product of $Z_{i\ell}^{m_\ell}$ for all $\ell = 1, \ldots, s$, where m_ℓ is the index of the threshold value with respect to constraint ℓ. That is, we declare system i is feasible with respect to $\boldsymbol{q} = (q_1^{m_1}, \ldots, q_s^{m_s})$ if $\prod_{\ell=1}^s Z_{i\ell}^{m_\ell} = 1$, and declare system i is infeasible if $\prod_{\ell=1}^s Z_{i\ell}^{m_\ell} = 0$. The variables $\text{LAST}_{i\ell}$, where $i = 1, \ldots, k$ and $\ell = 1, \ldots, s$, are needed if we would like to be able to add threshold values efficiently after the execution of Algorithm 1 (see Section 2.3.3).

2.3.2 Implementation Parameters

In this section, we discuss the two choices for β_ℓ in Algorithm 1. Comparison between the two choices is provided in Section 2.5.2.

First, note that β can be interpreted as the nominal probability of error for each system. This error β needs to be split among different constraints and different threshold values for each system. Theorem 1 in Section 2.4.1 shows that the *effective* number of threshold values for each constraint with multiple threshold values is two. The first choice (i) in Algorithm 1 is to split the error β among constraints equally for each system and then split it further among threshold values. The second choice (ii) is to count the total effective number of threshold values over all constraints and split β equally among them. Thus the first choice (i) gives β/s to each constraint and then it is further split among the effective number of threshold values (thus β/s to a constraint with only one threshold value and $\beta/(2s)$ to a constraint with two or more threshold values). On the other hand, the total number of effective threshold values is D (i.e., $D = \sum_{\ell=1}^s \min\{d_\ell, 2\}$) and the second choice (ii) is to split β equally among them.

For example, suppose we consider a feasibility check problem among k systems and two performance measures. The first constraint has one threshold q_1^1 but the second constraint has two thresholds q_2^1 and q_2^2. We set β as in (4). Then the first choice (i) sets $\beta_1 = \beta/2$ and $\beta_2 = \beta/4$, while the second choice (ii) sets $\beta_1 = \beta_2 = \beta/3$. Notice that choices (i) and (ii) become identical if all the constraints have either one or multiple thresholds to be tested (i.e., $d_\ell = 1$ for $\ell = 1, \ldots, s$ or $d_\ell \geq 2$ for all $\ell = 1, \ldots, s$).

2.3.3 Adding More Thresholds

A decision maker may want to add a new threshold value $q_\ell^{d_\ell+1}$ once feasibility with respect to thresholds $\{q_\ell^1, q_\ell^2, \ldots, q_\ell^{d_\ell}\}$ is determined. For example, suppose that a manager wants to identify an (s, S) inventory policy whose expected cost per review period is less than 115 or 120 dollars. If there is no feasible policy with these two thresholds, the manager may want to test a new threshold value, say, 125 dollars. We call the feasibility check for the initial set of threshold values $\{q_\ell^1, \ldots, q_\ell^{d_\ell}\}$, where $\ell = 1, \ldots, s$, the first pass, and the feasibility check for the newly-added threshold $q_\ell^{d_\ell+1}$ the second pass.

Suppose first that $d_\ell \geq 2$. Then there is no change in the values of $\beta_1, \ldots, \beta_s$ in $\mathcal{RF}$ when more thresholds are added. When a new threshold value $q_\ell^{d_\ell+1}$ is added after the first pass, one can compare the values of $v_{i\ell}^{\mathrm{UB}}$ and $v_{i\ell}^{\mathrm{LB}}$ with $q_\ell^{d_\ell+1}$. More specifically, if $v_{i\ell}^{\mathrm{UB}} \leq q_\ell^{d_\ell+1}$ and $v_{i\ell}^{\mathrm{LB}} < q_\ell^{d_\ell+1}$ ($v_{i\ell}^{\mathrm{LB}} \geq q_\ell^{d_\ell+1}$ and $v_{i\ell}^{\mathrm{UB}} > q_\ell^{d_\ell+1}$), we immediately declare system i is feasible (infeasible) with respect to $q_\ell^{d_\ell+1}$ and there is no need for additional observations. If $v_{i\ell}^{\mathrm{UB}} > q_\ell^{d_\ell+1}$ and $v_{i\ell}^{\mathrm{LB}} < q_\ell^{d_\ell+1}$, we take additional observations until (2) is satisfied.

When the first pass of feasibility decisions is over, since we calculate the values of $v_{i\ell}^{\mathrm{LB}}$ and $v_{i\ell}^{\mathrm{UB}}$ only at integer times (for $r_i = n_0, n_0 + 1, \ldots$) rather than continuously in time, it is possible that $v_{i\ell}^{\mathrm{UB}} \leq v_{i\ell}^{\mathrm{LB}}$ happens at the *last* stage (i.e., the integer time when the first pass of feasibility decisions are concluded for constraint ℓ on system i). If the new threshold value $q_\ell^{d_\ell+1}$ satisfies $v_{i\ell}^{\mathrm{UB}} \leq q_\ell^{d_\ell+1} \leq v_{i\ell}^{\mathrm{LB}}$, we need to know which value was updated last in the first pass. Consider the example in Figure 2. At the *last* stage of the first pass (shown as the feasibility check for $q_\ell^{d_\ell}$), we have $v_{i\ell}^{\mathrm{UB}} \leq v_{i\ell}^{\mathrm{LB}}$. When $q_\ell^{d_\ell+1}$ is added after the first pass, we have $v_{i\ell}^{\mathrm{UB}} \leq q_\ell^{d_\ell+1} \leq v_{i\ell}^{\mathrm{LB}}$ and $v_{i\ell}^{\mathrm{UB}}$ is the last updated value in the first pass. As

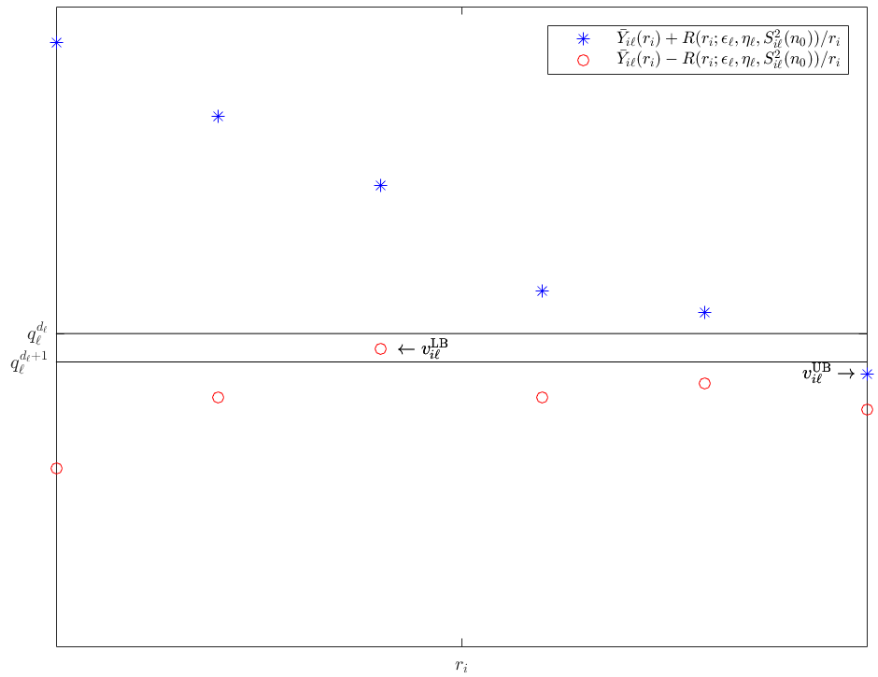


Figure 2: Crossing of $v_{i\ell}^{\text{UB}}$ and $v_{i\ell}^{\text{LB}}$

shown in Figure 2, the upper bound of the interval $\left(\bar{Y}_{i\ell}(r_i) - R(r_i; \epsilon_\ell, \eta_\ell, S^2_{i\ell}(n_0))/r_i, \bar{Y}_{i\ell} + R(r_i; \epsilon_\ell, \eta_\ell, S^2_{i\ell}(n_0))/r_i\right)$ is greater than $v_{i\ell}^{\text{LB}}$ before the last stage of the first pass and $q_\ell^{d_\ell+1}$ would have satisfied $v_{i\ell}^{\text{LB}} \geq q_\ell^{d_\ell+1}$ before it satisfied $v_{i\ell}^{\text{UB}} \leq q_\ell^{d_\ell+1}$ if it had been included in the initial set of threshold values. Thus we should declare system i infeasible with respect to $q_\ell^{d_\ell+1}$. In general, when $v_{i\ell}^{\text{UB}} \leq q_\ell^{d_\ell+1} \leq v_{i\ell}^{\text{LB}}$, if the last updated value in the first pass is $v_{i\ell}^{\text{UB}}$, we declare the system infeasible with respect to $q_\ell^{d_\ell+1}$ and we declare the system feasible with respect to $q_\ell^{d_\ell+1}$ if the last updated value in the first pass is $v_{i\ell}^{\text{LB}}$.

The description of the procedure when $d_\ell \geq 2$ and a new threshold $q_\ell^{d_\ell+1}$ is added for constraint ℓ is shown in Algorithm 2. We find the value of $Z_{i\ell}^{d_\ell+1}$ and declare the feasibility with respect to $\boldsymbol{q} = (q_1^{m_1}, \ldots, q_\ell^{d_\ell+1}, \ldots, q_s^{m_s})$ by checking the value of $Z_{i\ell}^{d_\ell+1} \cdot \prod_{\ell' \neq \ell, \ell'=1,2,\ldots,s} Z_{i\ell'}^{m_{\ell'}}$. When $\boldsymbol{q}$ contains multiple additional thresholds, let L denote the set of constraints that have new thresholds. The feasibility with respect to $\boldsymbol{q}$ is declared based on the value of $\prod_{\ell \in L} Z_{i\ell}^{d_\ell+1} \cdot \prod_{\ell \notin L} Z_{i\ell}^{m_\ell}$.

However, if ℓ is such that $d_\ell = 1$, then adding new threshold values to constraint ℓ

Algorithm 2 When $d_\ell \geq 2$ and $q_\ell^{d_\ell+1}$ is added

for each system $i \in \Theta$ **do**
  Set ON $= \{\ell\}$ and $\text{ON}_\ell = \{d_\ell + 1\}$.
  If $v_{i\ell}^{\text{UB}} \leq q_\ell^{d_\ell+1}$ and $v_{i\ell}^{\text{LB}} < q_\ell^{d_\ell+1}$, set $Z_{i\ell}^{d_\ell+1} = 1$ and $\text{ON}_\ell = \text{ON}_\ell \backslash \{d_\ell + 1\}$;
  Else if $v_{i\ell}^{\text{LB}} \geq q_\ell^{d_\ell+1}$ and $v_{i\ell}^{\text{UB}} > q_\ell^{d_\ell+1}$, set $Z_{i\ell}^{d_\ell+1} = 0$ and $\text{ON}_\ell = \text{ON}_\ell \backslash \{d_\ell + 1\}$;
  Else if $v_{i\ell}^{\text{UB}} \leq q_\ell^{d_\ell+1} \leq v_{i\ell}^{\text{LB}}$,
   set $Z_{i\ell}^{d_\ell+1} = 0$ if $\text{LAST}_{i\ell} = \text{UB}$, and set $\text{ON}_\ell = \text{ON}_\ell \backslash \{d_\ell + 1\}$;
   set $Z_{i\ell}^{d_\ell+1} = 1$ if $\text{LAST}_{i\ell} = \text{LB}$, and set $\text{ON}_\ell = \text{ON}_\ell \backslash \{d_\ell + 1\}$.
  If $\text{ON}_\ell = \emptyset$, set ON $=$ ON$\backslash\{\ell\}$.
  If ON $= \emptyset$, return $Z_{i\ell}^{d_\ell+1}$; otherwise, set $r_i = r_i + 1$, take one additional observation $Y_{i\ell r_i}$ and update $\bar{Y}_{i\ell}(r_i)$ for $\ell \in$ ON, then go to [**Feasibility Check**] of Procedure $\mathcal{RF}$.
end for

requires updating $\beta_1, \ldots, \beta_s$ if they are set as in (ii) in Algorithm 1 or just β_ℓ if $\beta_1, \ldots, \beta_s$ are set as in (i) in Algorithm 1. A detailed discussion for this case is provided in Appendix A.1.

2.4 Statistical Validity

We prove the statistical validity of $\mathcal{RF}$ in this section. We first address statistical validity for a single system, one or more constraints, and multiple constraint thresholds in Section 2.4.1, and then discuss the overall probability of correct decision in Section 2.4.2 to illustrate that our procedure for handling multiple constraint thresholds can be generalized to multiple systems and multiple constraints.

We first introduce the following lemma which is critical in proving the statistical validity of Algorithm 1. Recall that $\mathcal{RF}$ ensures that η_ℓ satisfies $g(\eta_\ell) = \beta_\ell$ where $g(\cdot)$ is defined in (3). Discussion about the existence of η_ℓ for $c \in \{1, \infty\}$ is provided in Appendix A.2.

Lemma 1. *For system i and constraint ℓ with specific threshold value q_ℓ^m, $\mathcal{RF}$ guarantees $\Pr(\text{CD}_{i\ell}(q_\ell^m)) \geq 1 - \beta_\ell$.*

Proof. For $c \in \mathbb{N}^+$, if system i and constraint ℓ with threshold value q_ℓ^m are considered separately, then it is easy to see that Procedure $\mathcal{RF}$ is essentially the same as a statistically-valid feasibility check procedure in [1] for a single system and a single constraint with one threshold value and confidence level $1-\beta_\ell$. Thus the lemma holds. For $c = \infty$, see Appendix

A.3. □

2.4.1 Statistical Validity for a Single System

In this section, we first prove the statistical validity of $\mathcal{RF}$ for a single system with a single constraint and then for a single system with multiple constraints. For simplicity, we use $\{q^1, q^2, \ldots, q^d\}$ to denote the changing threshold values with respect to the single constraint. Thus, we drop the subscripts in some notation as follows:

$$\begin{aligned} y &= \text{Performance measure;} \\ \epsilon &= \text{Tolerance level for the single constraint;} \\ \text{CD}(q) &= \text{Correct decision event with constraint threshold } q. \end{aligned}$$

In the following theorem, we use $\Pr(\text{CD}(y-\epsilon), \text{CD}(y+\epsilon))$ to denote the joint probability of the events $\text{CD}(y-\epsilon)$ and $\text{CD}(y+\epsilon)$.

Theorem 1. *Given a single system and a single constraint with threshold constants $\{q^1, q^2, \ldots, q^d\}$, the joint probability of correct decision with respect to thresholds $y-\epsilon$ and $y+\epsilon$ is a lower bound on the joint probability of correct decision with respect to all thresholds, i.e.,*

$$\Pr(\cap_{m=1}^{d} \text{CD}(q^m)) \geq \Pr(\text{CD}(y-\epsilon), \text{CD}(y+\epsilon)).$$

Proof. Without loss of generality, we can assume $q^1 < \cdots < q^d$. We consider the following three cases.

When $q^1 < \cdots < q^d \leq y - \epsilon$, the correct decision is to declare the system infeasible for all threshold constants $q^1, \ldots, q^d$. By (2), we have $\text{CD}(y-\epsilon) \subseteq \text{CD}(q^d) \subseteq \cdots \subseteq \text{CD}(q^1)$. Thus, we have

$$\Pr(\cap_{m=1}^{d} \text{CD}(q^m)) \geq \Pr(\text{CD}(y-\epsilon)) \geq \Pr(\text{CD}(y-\epsilon), \text{CD}(y+\epsilon)).$$

When $y + \epsilon \leq q^1 < \cdots < q^d$, by similar arguments and (2), we have $\text{CD}(y+\epsilon) \subseteq \text{CD}(q^1) \subseteq \cdots \subseteq \text{CD}(q^d)$ and

$$\Pr(\cap_{m=1}^{d} \text{CD}(q^m)) \geq \Pr(\text{CD}(y+\epsilon)) \geq \Pr(\text{CD}(y-\epsilon), \text{CD}(y+\epsilon)).$$

In general, there exist $\underline{m}$ and $\overline{m}$ where $0 \le \underline{m} < \overline{m} \le d+1$ such that $q^1 < \cdots < q^{\underline{m}} \le y - \epsilon < q^{\underline{m}+1} < \cdots < q^{\overline{m}-1} < y + \epsilon \le q^{\overline{m}} < \cdots < q^d$. Then the correct decision is to declare the system infeasible for threshold constants $q^1, \ldots, q^{\underline{m}}$ and feasible for threshold constants $q^{\overline{m}}, \ldots, q^d$. By (2), it is clear that $\mathrm{CD}(y-\epsilon) \subseteq \mathrm{CD}(q^{\underline{m}}) \subseteq \cdots \subseteq \mathrm{CD}(q^1)$ and $\mathrm{CD}(y+\epsilon) \subseteq \mathrm{CD}(q^{\overline{m}}) \subseteq \cdots \subseteq \mathrm{CD}(q^d)$. It is also clear that the procedure always makes correct decisions for acceptable systems, i.e., $\mathrm{CD}(q) = 1$ when $q \in \{q^{\underline{m}+1}, \ldots, q^{\overline{m}-1}\}$, and thus the overall probability of correct decision does not decrease due to acceptable systems. Finally, we have

$$\Pr(\cap_{m=1}^{d} \mathrm{CD}(q^m)) \ge \Pr(\mathrm{CD}(y-\epsilon), \mathrm{CD}(y+\epsilon))$$

as claimed. □

Based on the above theorem, we know that the PCD over $\{q^1, q^2, \ldots, q^d\}$ has as lower bound the PCD with two threshold values, $y - \epsilon$ and $y + \epsilon$. Thus, when there are multiple threshold values, the CD event is achieved as long as we make a correct decision for these two threshold values, and thus the effective number of threshold values for any constraint with multiple threshold values is just two.

Based on Theorem 1, we introduce the following theorem that gives the lower bound on the PCD with respect to all thresholds for a single system and s constraints.

Theorem 2. *For system i with s constraints and threshold constants $\{q_\ell^1, q_\ell^2, \ldots, q_\ell^{d_\ell}\}$ for $\ell = 1, 2, \ldots, s$, the $\mathcal{RF}$ procedure guarantees $\Pr(\cap_{\ell=1}^{s} \mathrm{CD}_{i\ell}) \ge 1 - \beta$.*

Proof. We first discuss the case when β_ℓ in Algorithm 1 is determined based on (i), so that

$$\beta_\ell = \begin{cases} \beta/s, & \text{if } d_\ell = 1; \\ \beta/(2s), & \text{if } d_\ell > 1. \end{cases}$$

For ℓ such that $d_\ell = 1$,

$$\Pr(\mathrm{CD}_{i\ell}) = \Pr(\mathrm{CD}_{i\ell}(q_\ell^1)) \ge 1 - \beta_\ell = 1 - (\beta/s),$$

where the inequality holds due to Lemma 1. For ℓ such that $d_\ell > 1$,

$$\begin{aligned}\Pr(\mathrm{CD}_{i\ell}) &= \Pr(\cap_{m=1}^{d_\ell}\mathrm{CD}_{i\ell}(q_\ell^m)) \\ &\geq \Pr(\mathrm{CD}_{i\ell}(y_\ell - \epsilon_\ell), \mathrm{CD}_{i\ell}(y_\ell + \epsilon_\ell)) \\ &\geq \Pr(\mathrm{CD}_{i\ell}(y_\ell - \epsilon_\ell)) + \Pr(\mathrm{CD}_{i\ell}(y_\ell + \epsilon_\ell)) - 1 \\ &\geq 1 - \beta_\ell + 1 - \beta_\ell - 1 = 1 - 2\beta_\ell = 1 - 2(\beta/(2s)) = 1 - (\beta/s),\end{aligned}$$

where the first inequality holds due to Theorem 1 and the third inequality holds due to Lemma 1. Thus, the Bonferroni inequality yields

$$\Pr(\cap_{\ell=1}^{s}\mathrm{CD}_{i\ell}) \geq 1 - \sum_{\ell=1}^{s}\Pr(\mathrm{ICD}_{i\ell}) \geq 1 - s(\beta/s) = 1 - \beta.$$

We then discuss the case when β_ℓ in Algorithm 1 is determined based on (ii). We let s_1 denote the number of constraints that have one threshold (i.e., the constraints with $d_\ell = 1$), and let s_2 denote the number of constraints that have two or more thresholds (i.e., the constraints with $d_\ell > 1$). Then $\beta_\ell = \beta/D$, where $D = \sum_{\ell=1}^{s}\min\{d_\ell, 2\} = s_1 + 2s_2$. Then, based on a similar argument as for case (i), we have that for ℓ such that $d_\ell = 1$,

$$\Pr(\mathrm{CD}_{i\ell}) \geq 1 - \beta_\ell = 1 - \frac{\beta}{s_1 + 2s_2}.$$

For ℓ such that $d_\ell > 1$,

$$\Pr(\mathrm{CD}_{i\ell}) \geq 1 - 2\beta_\ell = 1 - \frac{2\beta}{s_1 + 2s_2}.$$

Thus, the Bonferroni inequality yields

$$\Pr(\cap_{\ell=1}^{s}\mathrm{CD}_{i\ell}) \geq 1 - \sum_{\ell=1}^{s}\Pr(\mathrm{ICD}_{i\ell}) \geq 1 - \left(s_1\frac{\beta}{s_1 + 2s_2} + s_2\frac{2\beta}{s_1 + 2s_2}\right) = 1 - \beta. \qquad \square$$

2.4.2 Statistical Validity for Multiple Systems and Multiple Constraints

In this section, we extend Theorem 2 (which covers a single system with multiple constraints) to the general case with multiple systems and multiple constraints in the following theorem.

Theorem 3. *Procedure $\mathcal{RF}$ guarantees* $\mathrm{PCD} \geq 1 - \alpha$.

Proof. When systems are simulated with common random numbers (CRN), we have

$$\begin{aligned}
\text{PCD} &= \Pr\left(\cap_{i=1}^{k} \cap_{\ell=1}^{s} \text{CD}_{i\ell} \right) \\
&\geq 1 - \sum_{i=1}^{k} \Pr\left(\cap_{\ell=1}^{s} \text{CD}_{i\ell} \right)^{c} \\
&\geq 1 - k\beta \\
&= 1 - k\frac{\alpha}{k} = 1 - \alpha,
\end{aligned}$$

where the first inequality follows from the Bonferroni inequality, the second inequality holds due to Theorem 2, and we have used equation (4).

Similarly, when systems are simulated independently, we have

$$\begin{aligned}
\text{PCD} &= \Pr\left(\cap_{i=1}^{k} \cap_{\ell=1}^{s} \text{CD}_{i\ell} \right) \\
&= \prod_{i=1}^{k} \Pr\left(\cap_{\ell=1}^{s} \text{CD}_{i\ell} \right) \\
&\geq [1-\beta]^{k} = \left[1 - (1 - (1-\alpha)^{1/k})\right]^{k} = 1 - \alpha,
\end{aligned}$$

where the inequality follows from Theorem 2 and we have used equation (4). □

2.5 Experiments

In this section, we provide the results of our numerical experiments to demonstrate the performance of procedure $\mathcal{RF}$ compared with the performance of four alternative statistically valid procedures: (a) recycling observations but with the Bonferroni inequality for all systems, all constraints, and all thresholds (Recycle$^{\mathcal{B}}$); (b) performing feasibility check for each combination of thresholds of all constraints (Restart$^{\text{prod}}$) and restarting procedure $\mathcal{F}_B$ "from scratch" for each combination of thresholds; (c) performing feasibility check for all systems, all constraints, and all thresholds based on a pre-defined order of constraints and thresholds and restarting procedure $\mathcal{F}_B$ every time "from scratch" (Restart$^{\text{sum}}$); and (d) combining the thresholds into $\max_{\ell=1,\ldots,s} d_\ell$ threshold vectors (with at most one threshold for each constraint) and performing feasibility check for each threshold vector (Restart$^{\max}$). Specifically, for the Restart$^{\text{sum}}$ procedure, we assume that a decision maker runs the feasibility check procedure $\mathcal{F}_B$ for constraints $\ell = 1, 2, \ldots, s$ and thresholds $m = 1, 2, \ldots, d_\ell$ in the stated order, restarting each time a constraint or a threshold is changed. We provide the detailed descriptions of the Recycle$^{\mathcal{B}}$, Restart$^{\text{prod}}$, Restart$^{\text{sum}}$, and Restart$^{\text{max}}$ procedures and proofs for their statistical validity in Appendices A.4, A.5, A.6, and A.7, respectively.

Notice that $\mathcal{RF}$ and Recycle$^{\mathcal{B}}$ become identical (except Recycle$^{\mathcal{B}}$ is not designed for adding constraint thresholds) when all constraints have either one or two threshold values and thus the two procedures will perform the same in those cases. When there are constraints with more than two threshold values, $\mathcal{RF}$ is expected to perform better than Recycle$^{\mathcal{B}}$. Similarly, $\mathcal{RF}$ is expected to perform better than Restart$^{\text{prod}}$, Restart$^{\text{sum}}$, and Restart$^{\text{max}}$ if there exist any constraints that have multiple thresholds.

We first discuss feasibility check for a single system with a single constraint and multiple threshold values in Section 2.5.1. We then address the choice of the implementation parameter β_ℓ in Section 2.5.2 and compare Restart$^{\text{prod}}$ and Restart$^{\text{sum}}$ with the other procedures in Section 2.5.3. Section 2.5.4 provides results for a single system with multiple correlated constraints and multiple threshold values. Section 2.5.5 shows results for multiple systems with multiple constraints, and Section 2.5.6 illustrates the performance of the different procedures for an inventory example (see Section 2.1). All the experiments are based on 100,000 replications with $\alpha = 0.05$ and $n_0 = 20$, and the experimental results are shown for a triangular-shaped ($c = 1$) or a straight-line ($c = \infty$) continuation region or both. We report estimated PCD and average total number of observations (OBS) for each experiment, where we consider $(Y_{i1n}, \ldots, Y_{isn})$ (and any subset thereof) as one observation for system i. With 100,000 replications, the estimated PCD shown in our tables are meaningful up to the 0.001th digit, while the first four digits of OBS are meaningful.

2.5.1 A Single System with a Single Constraint

We consider a single system and a single constraint with feasibility checks for two, four, and one hundred threshold values. The observations from the single system are from a standard normal distribution and we set $\epsilon = 0.1$ in this subsection. As there are multiple thresholds in all cases, the choices (i) and (ii) for setting $\beta_1, \ldots, \beta_s$ in $\mathcal{RF}$, Recycle$^{\mathcal{B}}$, and Restart$^{\text{sum}}$ are identical. Since the number of combinations of thresholds is just the number of thresholds on the single constraint, $\text{Restart}^{\text{prod}}$, $\text{Restart}^{\text{sum}}$ and $\text{Restart}^{\max}$ are identical. We only report the results of $\text{Restart}^{\max}$ for all three procedures.

We first consider two threshold values $q^1 = -\epsilon$ and $q^2 = \epsilon$. The system is infeasible with respect to q^1 and feasible with respect to q^2. This case is considered as the "most difficult" case for a single system with a single constraint as the mean performance ($y = 0$) is at the boundary of the unacceptable or desirable regions. For four threshold values, we consider $q^1 = -1.1\epsilon, q^2 = -\epsilon, q^3 = \epsilon$, and $q^4 = 1.1\epsilon$. The estimated PCD and OBS required to complete the feasibility checks are shown in Table 1 for both triangular and straight-line continuation regions.

Table 1: Average number of observations and estimated PCD for $k = 1$ system and $s = 1$ constraint with two and four threshold values

	Two threshold values				Four threshold values			
	Triangular		Straight-line		Triangular		Straight-line	
	PCD	OBS	PCD	OBS	PCD	OBS	PCD	OBS
$\mathcal{RF}$	0.9539	305.09	0.9552	313.86	0.9542	307.26	0.9555	316.90
$\text{Recycle}^{\mathcal{B}}$	0.9539	305.09	0.9552	313.86	0.9769	387.18	0.9773	390.27
$\text{Restart}^{\max}$	0.9526	465.32	0.9555	436.37	0.9608	1164.85	0.9622	1055.30

It is clear that $\mathcal{RF}, \text{Recycle}^{\mathcal{B}}$, and $\text{Restart}^{\max}$ guarantee PCD greater than $1 - \alpha$. The estimated PCD of $\mathcal{RF}$ is around 0.95 for both two and four thresholds. For $\text{Recycle}^{\mathcal{B}}$ and $\text{Restart}^{\max}$, the PCD increases to around 0.97 and 0.96 for the four-value case from about 0.95 for the two-value case, respectively.

In terms of the number of observations, $\mathcal{RF}$ performs identically in the two-value case compared with $\text{Recycle}^{\mathcal{B}}$ as expected, and spends 28.07%–34.43% fewer observations than $\text{Restart}^{\max}$. In the four-value case, $\mathcal{RF}$ saves 18.80%–20.64% and 69.97%–73.62% on the number of observations compared with $\text{Recycle}^{\mathcal{B}}$ and $\text{Restart}^{\max}$, respectively.

We then consider one hundred threshold values q^m, where $m = 1, \ldots, 100$, as follows:

$$q^m = \begin{cases} -6\epsilon + 0.1m \times \epsilon, & \text{if } m \leq 50, \\ \epsilon + 0.1(m - 51) \times \epsilon, & \text{if } m \geq 51. \end{cases}$$

The experimental results are shown in Table 2.

Table 2: Average number of observations and estimated PCD for $k = 1$ system and $s = 1$ constraint with one hundred threshold values

	Triangular		Straight-line	
	PCD	OBS	PCD	OBS
$\mathcal{RF}$	0.9534	310.54	0.9557	323.03
Recycle$^{\mathcal{B}}$	0.9992	765.71	0.9990	813.80
Restart$^{\max}$	0.9982	30360.27	0.9977	21433.57

When the number of thresholds increases dramatically, while $\mathcal{RF}$ still performs similar to the case when the number of thresholds is two, Recycle$^{\mathcal{B}}$ and Restart$^{\max}$ require a lot more OBS and are also more conservative in terms of PCD. It is expected that $\mathcal{RF}$ needs much fewer OBS than Recycle$^{\mathcal{B}}$ and Restart$^{\max}$ if the number of thresholds is significantly large for both triangular and straight-line continuation regions. We see that the straight-line continuation region requires more observations compared with the triangular continuation region in $\mathcal{RF}$ and Recycle$^{\mathcal{B}}$ while it requires fewer observations in Restart$^{\max}$. The experiments in the remaining sections are based on triangular continuation regions (except for one experiment in Section 2.5.5.1).

2.5.2 Implementation Parameters for Systems with Multiple Constraints

In this section, we provide a numerical example and a discussion about the performance of $\mathcal{RF}$, Recycle$^{\mathcal{B}}$, and Restart$^{\text{sum}}$ as a function of the two different ways of setting the implementation parameters β_ℓ (see Section 2.3.2). As Restart$^{\text{prod}}$ and Restart$^{\max}$ only have one way of setting the implementation parameter, we omit them in this section. We use $\mathcal{RF}_1$, Recycle$_1^{\mathcal{B}}$, and Restart$_1^{\text{sum}}$ to denote the versions of the procedures that set the parameters $\beta_1, \ldots, \beta_s$ based on choice (i), and use $\mathcal{RF}_2$, Recycle$_2^{\mathcal{B}}$, and Restart$_2^{\text{sum}}$ to denote the corresponding procedures with choice (ii) in each algorithm.

We first test the performance of the four procedures applied to four configurations. We consider a single system with two constraints, where the first constraint has one fixed threshold and the second constraint has two thresholds. In all the configurations shown below, we choose $\boldsymbol{y} = (0, 0)$ and $\epsilon = 0.1$. The observations of the two constraints are

Table 3: Average number of observations and observed PCD (reported in parentheses) for implementation parameters (i) and (ii)

	$\mathcal{RF}_1(\text{Recycle}_1^{\mathcal{B}})$	$\mathcal{RF}_2(\text{Recycle}_2^{\mathcal{B}})$	$\text{Restart}_1^{\text{sum}}$	$\text{Restart}_2^{\text{sum}}$
Configuration 1	408.45 (0.954)	396.11 (0.954)	833.34 (0.953)	815.37 (0.953)
Configuration 2	285.62 (0.977)	303.28 (0.983)	596.86 (0.975)	602.17 (0.983)
Configuration 3	242.40 (0.976)	276.55 (0.984)	435.10 (0.977)	456.13 (0.984)
Configuration 4	385.50 (0.977)	352.42 (0.968)	679.86 (0.977)	635.75 (0.968)

independent standard normal random variables.

Configuration 1: Set $q_1^1 = -\epsilon, q_2^1 = -\epsilon$, and $q_2^2 = \epsilon$.

Configuration 2: Set $q_1^1 = -\epsilon, q_2^1 = -2\epsilon$, and $q_2^2 = 2\epsilon$.

Configuration 3: Set $q_1^1 = -\epsilon, q_2^1 = -4\epsilon$, and $q_2^2 = 4\epsilon$.

Configuration 4: Set $q_1^1 = -4\epsilon, q_2^1 = -\epsilon$, and $q_2^2 = \epsilon$.

One may notice that the mean performance of the first constraint is at the boundary of the unacceptable region in the first three configurations, which is a "most difficult" case for one constraint with a single threshold. Configuration 1 sets the mean performance of the second constraint at either the boundary of the unacceptable region (q_2^1) or the boundary of the desirable region (q_2^2), while Configurations 2 and 3 set the mean performance further from the boundaries of the unacceptable (desirable) region. Configuration 4 has the same (difficult) thresholds on the second constraint as in Configuration 1 but sets the mean performance of the first constraint far from the boundary of the unacceptable region. Notice that the performance of $\mathcal{RF}$ and $\text{Recycle}^{\mathcal{B}}$ are expected to be identical as $d_\ell \leq 2$, where $\ell = 1, 2$. Table 3 shows the estimated PCD and OBS of the $\mathcal{RF}, \text{Recycle}^{\mathcal{B}}$, and $\text{Restart}^{\text{sum}}$ procedures under triangular-shaped continuation regions for all configurations.

We see that under both choices (i) and (ii), all the procedures guarantee statistical validity of all configurations. Choice (i) is dominated by choice (ii) for under Configuration

1 and 4 for all three procedures, while choice (i) performs better than choice (ii) under Configuration 2 and 3.

In Configuration 1, since all the thresholds for both constraints are considered as "most difficult", assigning error evenly to each feasibility check, which follows choice (ii), is plausible. However, as Configuration 2 has a "difficult" threshold on the first constraint but "easy" thresholds on the second constraint, allocating more error to the threshold for the first constraint and less error to the threshold for the second constraint, which follows choice (i), is beneficial. Configuration 3 has even "easier" thresholds on the second constraint, suggesting that choice (i) is more beneficial for all three procedures. In these cases, choosing between (i) and (ii) depends on the difficulty of the feasibility checks. Although Configuration 4 has the same number of thresholds as the other three configurations, it has an "easy" threshold on the first constraint but two "difficult" thresholds on the second constraint. Choice (ii) allows more error allocation to the second constraint and less error allocation to the first constraint, which performs better than choice (i).

It is clear that the total number of required observations depends on the difficulty of the feasibility checks and the number of thresholds on each constraint. Of course, the decision maker may not have the information about the mean configuration before she performs feasibility check. Thus it is difficult to predict in advance whether (i) or (ii) will result in better performance. Choice (i) also has value if it may be of interest to add thresholds later (see Appendix A.1).

2.5.3 Comparison between Restart$^{\text{prod}}$, Restart$^{\text{sum}}$, and the other Procedures

In this section, we compare the performance of Restart$^{\text{prod}}$ and Restart$^{\text{sum}}$ with the other procedures. Based on the decriptions shown in Appendix A.5 for Restart$^{\text{prod}}$, A.6 for Restart$^{\text{sum}}$ and A.7 for Restart$^{\text{max}}$, the number of "restarts" depends highly on the number of thresholds on each constraint. As Restart$^{\text{prod}}$ performs feasibility checks for each combination of thresholds of all constraints, it requires $\prod_{\ell=1}^{s} d_\ell$ "restarts" to determine feasibility for one system, whereas Restart$^{\text{sum}}$ performs feasibility checks independently for each system, each constraint, and each threshold, and hence requires $\sum_{\ell=1}^{s} d_\ell$ "restarts" for one system. On the other hand, $\text{Restart}^{\max}$ performs feasibility check by restarting independently for threshold vectors that contain thresholds from all constraints with thresholds that have not yet received feasibility decisions until the feasibility of each threshold on each constraint is determined. This requires $\max_{\ell=1,\dots,s} d_\ell$ "restarts" for one system.

We consider three configurations of a single system with two constraints and independent standard normal observations. The first configuration has one threshold on the first constraint and two thresholds on the second constraints. The second configuration has two thresholds on both constraints, and the third configuration has two thresholds on the first constraint and four thresholds on the second constraint. In all configurations, we set $\boldsymbol{y} = (0, 0)$ and $\epsilon = 0.1$. More specifically, we have

Configuration 1 (C1): $\boldsymbol{q}^1 = (-\epsilon), \boldsymbol{q}^2 = (-\epsilon, \epsilon)$.

Configuration 2 (C2): $\boldsymbol{q}^1 = (-\epsilon, \epsilon), \boldsymbol{q}^2 = (-\epsilon, \epsilon)$.

Configuration 3 (C2): $\boldsymbol{q}^1 = (-\epsilon, \epsilon), \boldsymbol{q}^2 = (-1.25\epsilon, -\epsilon, \epsilon, 1.25\epsilon)$.

The experimental results for the average number of observations and observed PCD for $\mathcal{RF}$, $\text{Recycle}^{\mathcal{B}}$,

$\text{Restart}^{\text{sum}}$, $\text{Restart}^{\max}$ and $\text{Restart}^{\text{prod}}$ under triangular-shaped continuation regions are shown in Table 4.

Table 4: Average number of observations and observed PCD (reported in parentheses) for $\text{Restart}^{\text{prod}}$, $\text{Restart}^{\text{sum}}$ and the other procedures

	$\mathcal{RF}_1$	$\mathcal{RF}_2$	$\text{Recycle}_1^{\mathcal{B}}$	$\text{Recycle}_2^{\mathcal{B}}$	$\text{Restart}^{\max}$	$\text{Restart}_1^{\text{sum}}$	$\text{Restart}_2^{\text{sum}}$	$\text{Restart}^{\text{prod}}$
C1	408.45 (0.954)	396.11 (0.954)	408.45 (0.954)	396.11 (0.954)	622.71 (0.953)	833.34 (0.953)	815.37 (0.953)	765.38 (0.953)
C2	466.57 (0.954)	466.57 (0.954)	466.57 (0.954)	466.57 (0.954)	766.26 (0.954)	1201.05 (0.954)	1201.05 (0.954)	1871.00 (0.954)
C3	467.17 (0.953)	467.17 (0.953)	523.40 (0.965)	524.67 (0.969)	1476.79 (0.964)	1981.91 (0.960)	1953.23 (0.964)	4315.16 (0.962)

We can easily see that Configuration 1 has two combinations of thresholds, while Configurations 2 and 3 have four and eight combinations, respectively. This means that $\text{Restart}^{\text{prod}}$ needs to perform two, four, and eight restarts to conclude feasibility checks for Configuration 1, 2, and 3, respectively. However, $\text{Restart}^{\text{sum}}$ performs three, four, and six restarts as it performs feasibility check independently for each system, constraint, and threshold. One the other hand, Restart$^{\max}$ performs two restarts for Configurations 1 and 2, and performs four restarts for Configuration 3. One may notice that when $\sum_{\ell=1}^{s} d_\ell$ is smaller than $\prod_{\ell=1}^{s} d_\ell$, Restart$^{\text{sum}}$ is likely to perform better than Restart$^{\text{prod}}$. As $\max_{\ell=1,\ldots,s} d_\ell$ is always smaller than $\sum_{\ell=1}^{s} d_\ell$ and $\prod_{\ell=1}^{s} d_\ell$, Restart$^{\max}$ is superior compared with Restart$^{\text{prod}}$ and Restart$^{\text{sum}}$.

In the remaining experimental results, we omit the Restart$^{\text{prod}}$ and Restart$^{\text{sum}}$ procedures and only demonstrate the perfomance of the $\mathcal{RF}$, Recycle$^{\mathcal{B}}$, and Restart$^{\max}$ procedures.

2.5.4 A Single System and Multiple Constraints under Correlations

We consider a single system with multiple constraints when the constraints are correlated. The observations satisfy Assumption 1 and the covariance matrix is chosen as a square matrix with diagonal elements σ_ℓ^2 for $\ell = 1, 2, \ldots, s$ and non-diagonal elements $\rho\sigma_\ell\sigma_\nu$ for $\ell \neq \nu$ where $\ell, \nu = 1, 2, \ldots, s$. Thus, we consider equal correlation, ρ, between each pair of constraints, where the value of ρ is chosen over $\{-0.25, -0.15, 0, 0.3, 0.7\}$ which ensures Σ_i to be positive definite. The marginal variances σ_ℓ^2 are in one of three configurations: constant variances (CV), increasing variances (IV), and decreasing variances (DV). The variances σ_ℓ^2 in the CV configuration are all set to 1, while the variances in IV and DV are set to $1 + (\ell - 1)\epsilon_\ell$ and $1 + (s - \ell)\epsilon_\ell$, where $\ell = 1, 2, \ldots, s$, respectively. We provide experimental results for a single system and multiple constraints, with either an equal or a varying number of threshold values for each constraint, in Section 2.5.4.1 and Section 2.5.4.2, respectively.

2.5.4.1 Multiple correlated constraints with the same number of thresholds for each constraint

We consider the case of a single system and five constraints where each constraint has four threshold values. As each constraint has the same number of thresholds, choices (i) and (ii) for β_ℓ are identical for $\mathcal{RF}$ and Recycle$^{\mathcal{B}}$. We choose the mean vector $\boldsymbol{y} = (0, 0, 0, 0, 0)$, $\boldsymbol{q}^1 = (-1.25\epsilon, -1.25\epsilon, -1.25\epsilon, -1.25\epsilon, -1.25\epsilon)$, $\boldsymbol{q}^2 = (-\epsilon, -\epsilon, -\epsilon, -\epsilon, -\epsilon)$, $\boldsymbol{q}^3 = (\epsilon, \epsilon, \epsilon, \epsilon, \epsilon)$, and

$\boldsymbol{q}^4 = (1.25\epsilon, 1.25\epsilon, 1.25\epsilon, 1.25\epsilon, 1.25\epsilon)$, where $\epsilon = 1/\sqrt{n_0}$. Table 5 shows that all the procedures are statistically valid and that the savings of $\mathcal{RF}$ in the average number of observations is 14.21% – 14.71% and 72.15% – 74.16% compared with Recycle$^{\mathcal{B}}$ and Restart$^{\max}$, respectively.

Table 5: Average number of observations and observed PCD (reported in parentheses) for correlated constraint observations and four threshold values for each constraint

	CV			IV			DV		
ρ	$\mathcal{RF}$	Recycle$^{\mathcal{B}}$	Restart$^{\max}$	$\mathcal{RF}$	Recycle$^{\mathcal{B}}$	Restart$^{\max}$	$\mathcal{RF}$	Recycle$^{\mathcal{B}}$	Restart$^{\max}$
-0.25	144.14	168.06	557.80	224.77	262.33	855.46	224.66	262.91	855.82
	(0.958)	(0.979)	(0.974)	(0.957)	(0.978)	(0.973)	(0.958)	(0.978)	(0.972)
-0.15	145.45	169.55	555.37	226.05	264.48	852.84	226.13	264.35	853.37
	(0.957)	(0.979)	(0.974)	(0.957)	(0.977)	(0.972)	(0.957)	(0.978)	(0.973)
0	146.17	170.50	549.59	226.91	265.58	846.63	227.03	265.59	845.74
	(0.957)	(0.978)	(0.974)	(0.956)	(0.978)	(0.972)	(0.957)	(0.978)	(0.972)
0.3	143.69	167.74	529.31	224.32	262.31	819.99	224.37	262.25	820.27
	(0.958)	(0.979)	(0.974)	(0.957)	(0.978)	(0.973)	(0.957)	(0.978)	(0.974)
0.7	132.38	154.81	477.05	210.93	247.13	757.28	210.84	247.20	757.13
	(0.967)	(0.982)	(0.977)	(0.967)	(0.982)	(0.978)	(0.965)	(0.982)	(0.977)

We also consider the case where each constraint has ten threshold values. That is, for each constraint ℓ, where $\ell = 1, \ldots, 5$, we choose $y_\ell = 0$ and $q_\ell^1 = -1.4\epsilon_\ell, q_\ell^2 = -1.3\epsilon, q_\ell^3 = -1.2\epsilon, q_\ell^4 = -1.1\epsilon, q_\ell^5 = -\epsilon, q_\ell^6 = \epsilon, q_\ell^7 = 1.1\epsilon, q_\ell^8 = 1.2\epsilon, q_\ell^9 = 1.3\epsilon$, and $q_\ell^{10} = 1.4\epsilon$, where $\epsilon = 1/\sqrt{n_0}$. Table 6 shows the savings from $\mathcal{RF}$ increases significantly when the number of thresholds is increased up to ten. While all three procedures guarantee statistical validity, $\mathcal{RF}$ saves 28.68%–29.39% observations compared with Recycle$^{\mathcal{B}}$, and 90.49%–91.03% observations compared with Restart$^{\max}$. The relative performance of $\mathcal{RF}$ compared with Recycle$^{\mathcal{B}}$ and Restart$^{\max}$ is about the same among all variance configurations with all three methods performing better under CV than under IV and DV (which have larger variances than CV). The effect of ρ in both cases is only significant for $\rho = 0.7$.

2.5.4.2 Multiple correlated constraints with a different number of thresholds for each constraint

We consider the setting when $s = 5$ and each constraint has the threshold values shown in Table 7. In this setting, the first two constraints have three thresholds, the third constraint has four thresholds, and the fourth and fifth constraints have two thresholds. We choose $\boldsymbol{y} = (0, 0, 0, 0, 0)$ and set $\epsilon = 1/\sqrt{n_0}$.

Table 6: Average number of observations and observed PCD (reported in parentheses) for correlated constraint observations and ten threshold values for each constraint

	CV			IV			DV		
ρ	$\mathcal{RF}$	Recycle$^{\mathcal{B}}$	Restart$^{\max}$	$\mathcal{RF}$	Recycle$^{\mathcal{B}}$	Restart$^{\max}$	$\mathcal{RF}$	Recycle$^{\mathcal{B}}$	Restart$^{\max}$
−0.25	144.15	202.12	1606.55	224.86	316.52	2477.48	224.66	316.85	2477.89
	(0.958)	(0.992)	(0.983)	(0.957)	(0.991)	(0.982)	(0.957)	(0.991)	(0.983)
−0.15	145.46	204.06	1603.13	226.14	318.57	2472.91	226.13	318.95	2472.10
	(0.957)	(0.992)	(0.983)	(0.957)	(0.991)	(0.983)	(0.957)	(0.992)	(0.982)
0	146.18	204.95	1590.65	226.97	319.55	2457.41	227.10	319.42	2456.58
	(0.957)	(0.991)	(0.982)	(0.956)	(0.991)	(0.983)	(0.957)	(0.991)	(0.983)
0.3	143.69	201.68	1537.17	224.31	316.01	2390.19	224.37	316.32	2392.14
	(0.958)	(0.992)	(0.983)	(0.957)	(0.991)	(0.983)	(0.957)	(0.991)	(0.983)
0.7	132.39	186.27	1391.54	210.91	298.69	2216.82	210.87	298.38	2217.65
	(0.967)	(0.992)	(0.985)	(0.966)	(0.993)	(0.984)	(0.965)	(0.992)	(0.984)

Table 7: Threshold configurations for a single system

Constraint	Threshold values of constraint ℓ
$\ell = 1$	$-2\epsilon, -1.25\epsilon, -\epsilon$
$\ell = 2$	$\epsilon, 1.25\epsilon, 2\epsilon$
$\ell = 3$	$-1.25\epsilon, -\epsilon, \epsilon, 1.25\epsilon$
$\ell = 4$	$-2\epsilon, 2\epsilon$
$\ell = 5$	$-\epsilon, \epsilon$

We show the estimated PCD and OBS under CV, IV, and DV with correlation $\rho = -0.25, 0, 0.25$ for the two choices of the parameter β_ℓ in Table 8. As all five constraints have at least two thresholds, the two choices of β_ℓ for $\mathcal{RF}$ are identical in this setting.

All the procedures are statistically valid and $\mathcal{RF}$ still shows savings up to 10.89% and 69.61% compared with Recycle$^{\mathcal{B}}$ and Restart$^{\max}$, respectively. The relative performance of $\mathcal{RF}$, Recycle$^{\mathcal{B}}$, and Restart$^{\max}$ is similar in all cases.

We see that $\mathcal{RF}$, Recycle$^{\mathcal{B}}$, and Restart$^{\max}$ require fewer observations compared with the case when all constraints have four thresholds (shown in Table 5) or ten thresholds (shown in Table 6). Similarly, the PCD in Table 8 is higher than in Table 5. This is consistent with the fact that the configuration of the thresholds in this case is easier.

For Recycle$^{\mathcal{B}}$, choice (ii) dominates choice (i) under CV and DV while it is dominated by choice (i) under IV. Constraints 1 and 2 have one "most difficult" threshold while constraints 3 and 5 have two "most difficult" thresholds. IV makes the later constraints more difficult, while DV makes the later constraints less difficult. Choice (i) sets larger β_ℓ for constraints

Table 8: Average number of observations and observed PCD (reported in parentheses) for correlated constraint observations and two, three, or four threshold values per constraint

		$\mathcal{RF}$	$\text{Recycle}_1^{\mathcal{B}}$	$\text{Recycle}_2^{\mathcal{B}}$	$\text{Restart}^{\max}$
CV	$\rho = -0.25$	130.75	143.58	141.44	415.66
		(0.975)	(0.981)	(0.981)	(0.980)
	$\rho = 0$	132.42	145.08	143.18	420.75
		(0.975)	(0.981)	(0.982)	(0.980)
	$\rho = 0.25$	131.37	144.14	142.25	421.08
		(0.975)	(0.981)	(0.982)	(0.980)
IV	$\rho = -0.25$	207.89	219.21	224.82	616.50
		(0.973)	(0.980)	(0.981)	(0.979)
	$\rho = 0$	209.24	220.77	226.37	621.88
		(0.974)	(0.980)	(0.981)	(0.979)
	$\rho = 0.25$	208.24	219.89	224.72	621.64
		(0.974)	(0.981)	(0.981)	(0.979)
DV	$\rho = -0.25$	197.29	221.40	213.86	636.15
		(0.974)	(0.982)	(0.981)	(0.978)
	$\rho = 0$	199.85	223.73	216.56	643.54
		(0.973)	(0.981)	(0.981)	(0.979)
	$\rho = 0.25$	199.49	223.30	216.61	644.06
		(0.974)	(0.982)	(0.981)	(0.979)

4 and 5 compared with choice (ii), which is beneficial under the IV configuration. As in Section 2.5.4.1, the methods perform better under CV than under IV and DV. The effect of ρ in all cases is not significant.

2.5.5 Multiple Systems and Multiple Constraints under Correlation

In this section, we consider multiple independent systems with multiple correlated constraints, and provide experimental results for the cases of equal and different number of thresholds per constraint in Sections 2.5.5.1 and Section 2.5.5.2, respectively.

2.5.5.1 Multiple systems and multiple correlated constraints with the same number of thresholds for each constraint

In this section, we consider $k = 12$ systems with $s = 5$ constraints. For each system, we choose the same mean and threshold configurations as in the four threshold case in Section 2.5.4.1. Three of the twelve systems are from the CV, IV, and DV variance configurations, respectively. We test correlations $\rho = -0.25, 0, 0.25$, and 0.7 under both triangular-shaped and straight line continuation regions. The results are provided in Table 9.

Table 9 shows that all the procedures have estimated PCD greater than the nominal value and $\mathcal{RF}$ achieve savings up to 11.43% and 73.48% compared with Recycle$^{\mathcal{B}}$ and Restart$^{\max}$, respectively. The relative savings compared with Recycle$^{\mathcal{B}}$ is sightly less than for a single system, five constraints, and four thresholds (shown in Table 5), while it is similar compared with Restart$^{\max}$. As in Tables 5 and 6, the effect of ρ is only significant for $\rho = 0.7$.

Table 9: Average number of observations and observed PCD (reported in parentheses) of $k = 12$ systems and $s = 5$ constraints with four thresholds per constraint

	Triangular			Straight-line		
	$\mathcal{RF}$	Recycle$^{\mathcal{B}}$	Restart$^{\max}$	$\mathcal{RF}$	Recycle$^{\mathcal{B}}$	Restart$^{\max}$
$\rho = -0.25$	3915.09	4420.49	14760.83	3826.57	4253.85	13292.81
	(0.954)	(0.977)	(0.974)	(0.962)	(0.980)	(0.976)
$\rho = 0$	3956.83	4463.58	14685.11	3876.77	4306.32	13029.37
	(0.956)	(0.977)	(0.975)	(0.961)	(0.980)	(0.976)
$\rho = 0.25$	3921.11	4426.11	14395.44	3835.64	4262.10	12613.39
	(0.955)	(0.977)	(0.974)	(0.960)	(0.980)	(0.975)
$\rho = 0.7$	3676.17	4147.50	13260.89	3550.95	3955.76	11340.57
	(0.961)	(0.980)	(0.976)	(0.964)	(0.982)	(0.978)

2.5.5.2 *Multiple systems and multiple correlated constraints with a different number of thresholds for each constraint*

In this section, we consider $k = 12$ systems with $s = 5$ constraints, and the constraints for each system have the thresholds shown in Table 7. Four of the 12 systems have each of the variance configurations CV, IV, and DV, respectively. As each constraint has multiple thresholds, the two ways of choosing β_ℓ for $\mathcal{RF}$ are identical. However, as the numbers of thresholds on each constraint are different, the two ways of setting β_ℓ are different in Recycle$^{\mathcal{B}}$. We test the performance of $\mathcal{RF}$ and Restart$^{\max}$, and the performance for the two different choices of β_ℓ for Recycle$^{\mathcal{B}}$. Table 10 shows the estimated PCD and OBS to perform the feasibility check for all twelve systems under correlation $\rho \in \{-0.25, 0, 0.25\}$.

All three procedures guarantee statistical validity and the savings from $\mathcal{RF}$ are 5.86%–6.26% compared with Recycle$^{\mathcal{B}}$ and 68.36%–68.76% compared with Restart$^{\max}$. The savings are similar to the single system case (Table 8). Setting β_ℓ based on choice (ii) dominates (i) in this case for Recycle$^{\mathcal{B}}$. In the results from the case of single system with different number of thresholds per constraint (Table 8), we see that (ii) dominates (i) under both CV and DV for Recycle$^{\mathcal{B}}$ and (i) dominates (ii) only under IV. Therefore, it is expected that choice (ii) dominates (i) here as more systems come from CV and DV configurations. The correlation ρ does not have a significant impact in this case.

Table 10: Average number of observations and observed PCD (reported in parentheses) of $k = 12$ systems and $s = 5$ constraints with two, three, or four thresholds per constraint

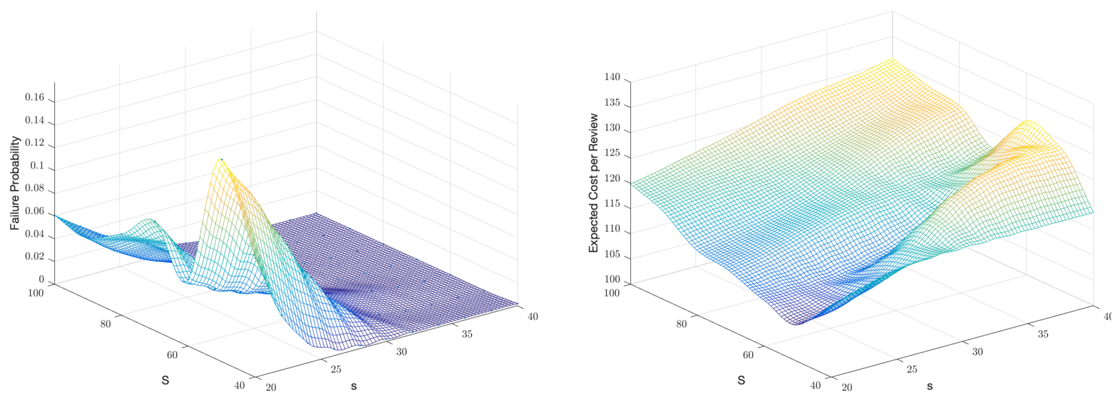
	$\mathcal{RF}$	Recycle$_1^{\mathcal{B}}$	Recycle$_2^{\mathcal{B}}$	Restart$^{\max}$
$\rho = -0.25$	3579.42	3816.45	3804.25	11313.39
	(0.974)	(0.980)	(0.980)	(0.981)
$\rho = 0$	3612.68	3853.91	3839.10	11416.95
	(0.972)	(0.980)	(0.980)	(0.981)
$\rho = 0.25$	3595.26	3832.25	3819.15	11403.96
	(0.973)	(0.980)	(0.981)	(0.981)

2.5.6 Inventory Policy Example

In this section, we test $\mathcal{RF}$, Recycle$^{\mathcal{B}}$, and Restart$^{\max}$ on the (s, S) inventory policy problem discussed in Section 2.1. The problem is from [11]. Two performance measures are considered: the failure probability ($\ell = 1$), which is the probability that a shortage occurs between two successive review periods, and the expected cost per review period ($\ell = 2$), which denotes the average total cost for each review period, including ordering cost, holding cost, and penalty cost when demand is more than the inventory level. The ordering cost is 3 per item as well as a fixed ordering cost of 32 per order. The holding cost is 1 per item per review period, and the penalty cost is 5 per item of unsatisfied demand. As in Section 2.1, consider three threshold values for the first constraint ($q_1 \in \{0.01, 0.05, 0.1\}$) and five threshold values for the second constraint ($q_2 \in \{115, 120, 125, 130, 135\}$).

Demand during each review period is assumed to follow a Poisson distribution with mean 25 and is independent for different review periods. We consider $\Theta = \{(s, S)|s = 20 + 2m, S = 40 + 10n$ where $m = 0, 1, 2, \ldots, 10$ and $n = 0, 1, 2, \ldots, 6\}$, which consists of 77 systems. We compute the two performance measures based on 30 review periods. The analytical results for both performance measures can be obtained by using a Markov chain model. The analytical result for the failure probability is presented in Figure 3a and the expected cost per review is provided in Figure 3b.

Observations corresponding to the two performance measures from each replication are not necessarily normally distributed and they are correlated. The values of the correlation between these two performance measures among all 77 systems are estimated based on simulation with 1,000,000 replications. The estimated correlations range from -0.23 to 0.52, and the results are shown in Figure 4.



(a) The values of the failure probability (b) The values of the expected cost per review

Figure 3: The values of the failure probability and expected cost per review

We apply procedures $\mathcal{RF}$, Recycle$^{\mathcal{B}}$, and Restart$^{\max}$ and obtain the average values of PCD and OBS, as well as their standard errors, based on 100,000 replications. For Recycle$^{\mathcal{B}}$, we set β_ℓ based on the second choice (ii) for the purpose of demonstration (the two choices are identical for $\mathcal{RF}$). Table 11 shows the results with and without CRN applied across systems. By applying CRN across systems, the correlation between systems with respect to the observed shortage between review periods has the range from -0.29 to 1, and the correlation between systems with respect to the observed cost per review period has the range from 0.25 to 1. These two correlations are shown in Figures 5a and 5b, respectively, where we order the 77 systems first by their values of S and then by their values of s (i.e., we set the index of system (20,40) as 1, the index of system (20,50) as 2, the index of system (22,40) as 8 and so on).

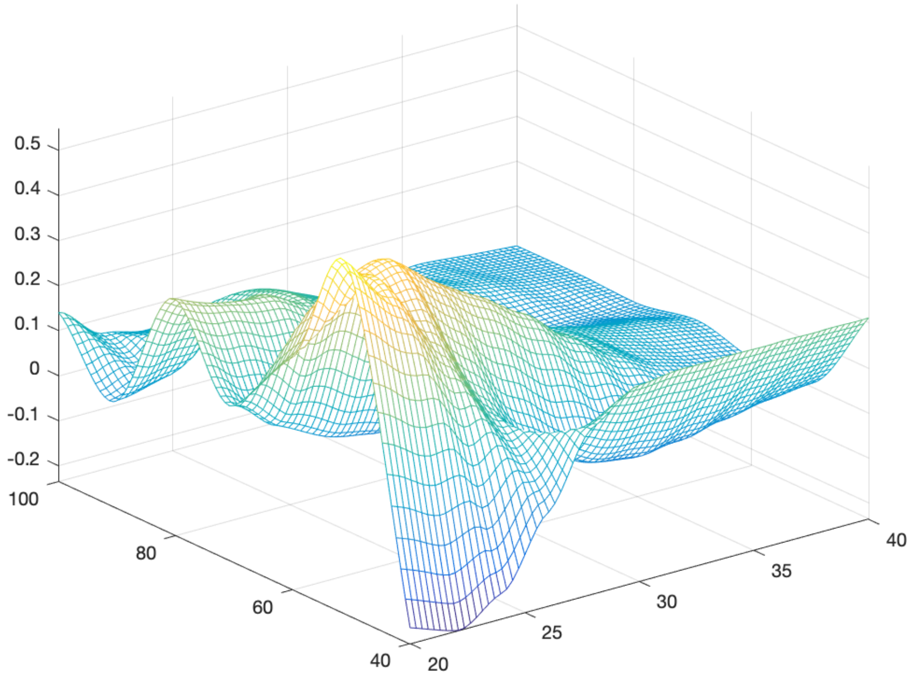
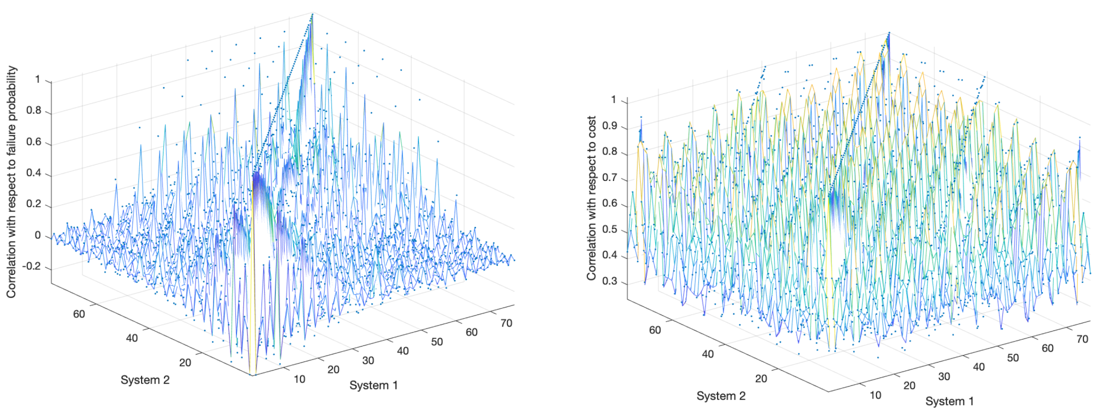


Figure 4: Estimated correlations between failure probability and cost per review



(a) The correlation between systems with respect to the
failure probability

(b) The correlation between systems with respect to the expected cost per review

Figure 5: The correlation between systems with respect to the failure probability and expected cost per review

From Table 11, we observe that the $\mathcal{RF}$, Recycle$^{\mathcal{B}}$, and Restart$^{\max}$ procedures are statistical valid and that $\mathcal{RF}$ achieves savings around 8% OBS compared with Recycle$^{\mathcal{B}}$, and

Table 11: Average number of observations and observed PCD (mean values and standard errors) with or without CRN applied across systems

		Without CRN		With CRN	
		Mean	Standard Error	Mean	Standard Error
$\mathcal{RF}$	PCD	0.998	0.00014	0.998	0.00015
	OBS	92475	24	92945	32
Recycle$^{\mathcal{B}}$	PCD	0.998	0.00013	0.998	0.00013
	OBS	100616	26	101119	35
Restart$^{\max}$	PCD	0.998	0.00012	0.998	0.00013
	OBS	148739	28	148883	48

around 38% compared with Restart$^{\max}$. The case using CRN requires slightly more OBS compared with the case that does not use CRN. This is due to the fact that the value of β_ℓ when CRN applied is set to incorporate the correlation between systems as compared with the case without CRN (see equations (4) and (13)), but CRN does not yield the same benefit for feasibility check as it does for comparing systems. The estimated PCD equals 0.998 in all cases.

Finally, we illustrate how to use the above results for finding the best system for a multi-objective problem. In our example, we want to minimize the failure probability and the expected cost per review period. From one single replication of the (s, S) policy example, one can enumerate all the systems that are feasible or infeasible with respect to each constraint and for each threshold using the $Z_{i\ell}^m$ variables (see Algorithm 1). Table 12 shows the number of feasible systems with respect to each combination of constraint thresholds for one particular run. As $q_1 = 0.01$ and $q_2 = 115$ corresponds to the tightest constraints, the systems that are feasible with respect to both thresholds should be considered as the best system. In our example, $(s, S) = (32, 60)$ is the best system.

If there are multiple feasible systems for the tightest threshold values, one may add additional thresholds that make the constraints tighter and run Algorithm 2. For example, there are 63 systems declared feasible with respect to $q_1 = 0.05$ and $q_2 = 130$. By considering a tighter threshold 0.01 for q_1, one will obtain 42 feasible systems, which is fewer than for $q_1 = 0.05$ and $q_2 = 130$. On the other hand, if there is no feasible system for the tightest constraints, then one may add easier thresholds and run Algorithm 2. For example, if a

Table 12: Number of feasible systems with respect to all combinations of constraint thresholds

q_1 \ q_2	115	120	125	130	135
0.01	1	17	27	42	44
0.05	8	33	48	63	65
0.1	14	41	56	71	73

decision maker uses tigher thresholds instead of $q_1 = 0.01$ and $q_2 = 115$, then it is likely that there do not exist any feasible systems. In this case, one may consider using "looser" thresholds on either constraint to see whether any systems are declared feasible.

2.6 Conclusions

In this chapter, we consider the problem of determining a feasible set of systems among finitely many systems when threshold constants in one or more constraints are subjective. We propose an indifference-zone procedure that recycles observations and performs feasibility check with respect to all thresholds for each constraint simultaneously. We prove the statistical validity of the proposed procedure and show by experiments that the procedure saves a significant number of observations compared with four alternative statistical valid procedures and also scales well with respect to the number of thresholds. We also explain that our procedure is useful in finding the best system for multi-objective optimization problems based on an application of determining an optimal inventory policy.

CHAPTER III

SELECTION OF THE BEST IN THE PRESENCE OF SUBJECTIVE STOCHASTIC CONSTRAINTS

3.1 Introduction

We consider the problem of selecting the best or near-best system with respect to a primary performance measure among a finite number of simulated systems while also satisfying stochastic constraints on one or more secondary performance measures. When no feasible system exists with respect to a given set of threshold values, the decision maker may be willing to relax the threshold values of some constraints so that a feasible system can be found. In that sense, constraints with varying thresholds can be considered as subjective constraints. We illustrate this problem with an example.

Suppose a decision maker wants to design an inventory policy such that the expected fill rate within each review period is maximized. She considers using an (s, S) inventory policy (namely ordering products to increase the inventory level up to S when the inventory level at a review period is below s and placing no order, otherwise). Two constraints exist, namely the probability that a shortage occurs between two successive review periods should be less than or equal to $q_1 = 1\%$ and the expected cost per review period should be less than or equal to q_2, where the value of q_2 is small. The decision maker thinks $q_2 = \$100{,}000$ is small but is willing to relax the threshold to \$105,000 or \$110,000 if no feasible system can be found with $q_2 = \$100{,}000$. If there is still no feasible systems with respect to $q_2 = \$110{,}000$, then the decision maker is willing to raise the threshold q_1 to 5%, still with three possible values for q_2.

In this chapter, we adopt the concept of recycling simulation observations in the context of constrained R&S when constraint thresholds vary. We provide fully sequential procedures that return the best feasible system with respect to the most preferred threshold values possible, where the preference order among thresholds is specified by the user. The threshold values for constraints are relaxed until there is at least one feasible solution. We prove that our procedures achieve a desired overall probability of correct selection and also perform well in reducing the required number of observations until a decision is made compared with straight-forward repeating procedures, namely applying procedures of [1] or [5] iteratively to each possible set of threshold values depending on whether the problem has a single constraint or multiple constraints.

The rest of this chapter is organized as follows: Section 3.2 provides the background for our problem. Sections 3.3 and 3.4 propose and analyze sequentially-running and simultaneously-running procedures, respectively, for the feasibility check and comparison phases. Section 3.5 discusses three major preference orders of the constraint thresholds and demonstrates how each preference order can be constructed automatically from users' inputed threshold values for each constraint. In Section 3.6, we present numerical results for the proposed procedures and compare their performances with the straight-forward procedures which apply existing constrained R&S procedures repeatedly to each set of thresholds. Concluding remarks are provided in Section 3.7. A discussion of two alternative procedures that we compare with our proposed procedures is included in Appendices B.1 and B.2.

3.2 Background

In this section, we formulate our problem in Section 3.2.1 and discuss how we define the correct selection event in Section 3.2.2. The necessary assumptions for the statistical validity of our proposed procedures are presented in Section 3.2.3.

3.2.1 Problem Formulation

We consider k systems whose primary performance measures, as well as s secondary performance measures, can be estimated through stochastic simulation. Let Γ denote the index set of all possible systems (i.e., $\Gamma = \{1, \ldots, k\}$). Let X_{in} be the observation associated with the primary performance measure of system i from replication n, and $Y_{i\ell n}$ be the observation associated with the ℓth stochastic constraint of system i from replication n, where $\ell = 1, \ldots, s$. We also define the expected values of the primary and secondary performance measures for each system $i \in \Gamma$ and constraint $\ell = 1, \ldots, s$ as $x_i = \mathrm{E}[X_{in}]$ and $y_{i\ell} = \mathrm{E}[Y_{i\ell n}]$, respectively. Constrained R&S is to select

$$\begin{array}{lll} \arg\max_{i \in \Gamma} & x_i & \\ \text{s.t.} & y_{i\ell} \leq q_\ell & \text{for all } \ell = 1, \ldots, s, \end{array}$$

where q_ℓ denotes the constraint threshold for constraint ℓ.

For a given threshold vector $\boldsymbol{q} = (q_1, \ldots, q_s)$, procedures due to [1] can be used to find the best system if there is only one constraint. If there are multiple constraints, procedures due to [5] are suitable. In this paper, we assume that the decision maker has a list of possible threshold values in consideration for each constraint and hopes to select the best system with respect to the most preferable thresholds possible. We let d_ℓ denote the number of distinct threshold values and q_ℓ^m denote the mth distinct threshold value on constraint ℓ, where $m = 1, \ldots, d_\ell$ and $\ell = 1, \ldots, s$. We assume $q_\ell^1 < \cdots < q_\ell^{d_\ell}$, where $\ell = 1, \ldots, s$.

The threshold values for individual constraints are combined into an ordered list of vectors of threshold values $\{\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(d)}\}$, where d denotes the total number of threshold vectors that the decision maker is interested to test. We assume that $\boldsymbol{q}^{(1)}$ is preferred to $\boldsymbol{q}^{(2)}$, $\boldsymbol{q}^{(2)}$ is preferred to $\boldsymbol{q}^{(3)}$, and so on. For the implementation of our procedures, a decision maker can input the ordered list of threshold vectors, or the decision maker can input an ordered list of threshold values for each constraint and a mechanism for constructing an ordered list of threshold vectors from the inputed threshold values (see Section 3.5). We let $q_\ell^{(\theta)}$ be the threshold value on constraint ℓ in $\boldsymbol{q}^{(\theta)}$, where $\theta = 1, \ldots, d$ and $\ell = 1, \ldots, s$. Then we introduce the threshold index vector $\boldsymbol{I}^{(\theta)}$ to include the indices of the threshold values that form $\boldsymbol{q}^{(\theta)}$. Similar to the definition of $q_\ell^{(\theta)}$, $I_\ell^{(\theta)}$ represents the threshold index on constraint ℓ.

Consider the example of selecting the best inventory control policy discussed in Section 2.1. Then $s = 2, d_1 = 2$ (i.e., two threshold values for the first constraint), $d_2 = 3$ (i.e., three threshold values for the second constraint), $q_1^1 = 1, q_1^2 = 5$, and $q_2^1 = 100000, q_2^2 = 105000$, and $q_2^3 = 110000$. Moreover, we consider the following $d = 6$ ordered threshold vectors

$$\boldsymbol{q}^{(1)} = \begin{bmatrix} 1 \\ 100000 \end{bmatrix}, \quad \boldsymbol{q}^{(2)} = \begin{bmatrix} 1 \\ 105000 \end{bmatrix}, \quad \boldsymbol{q}^{(3)} = \begin{bmatrix} 1 \\ 110000 \end{bmatrix},$$

$$\boldsymbol{q}^{(4)} = \begin{bmatrix} 5 \\ 100000 \end{bmatrix}, \quad \boldsymbol{q}^{(5)} = \begin{bmatrix} 5 \\ 105000 \end{bmatrix}, \text{ and } \boldsymbol{q}^{(6)} = \begin{bmatrix} 5 \\ 110000 \end{bmatrix}.$$

Note that $q_1^{(1)} = q_1^{(2)} = q_1^{(3)} = 1, q_1^{(4)} = q_1^{(5)} = q_1^{(6)} = 5$, while $q_2^{(1)} = q_2^{(4)} = 100000, q_2^{(2)} = q_2^{(5)} = 105000$, and $q_2^{(3)} = q_2^{(6)} = 110000$. The threshold index vectors are

$$\boldsymbol{I}^{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{I}^{(2)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{I}^{(3)} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad \boldsymbol{I}^{(4)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{I}^{(5)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \text{ and } \boldsymbol{I}^{(6)} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Hence $I_1^{(1)} = I_1^{(2)} = I_1^{(3)} = 1, I_1^{(4)} = I_1^{(5)} = I_1^{(6)} = 2$, while $I_2^{(1)} = I_2^{(4)} = 1, I_2^{(2)} = I_2^{(5)} = 2$, and $I_2^{(3)} = I_2^{(6)} = 3$.

For $\theta \leq d$, we use A_θ to denote the region that is feasible under threshold vector $\boldsymbol{q}^{(\theta)}$ but not under threshold vectors $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta-1)}$ (if $\theta > 1$), and use A_{d+1} to denote the region that is infeasible to all $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(d)}$. More specifically, we let

$$A_\theta = \begin{cases} \left\{(z_1, z_2, \ldots, z_s) : z_\ell \leq q_\ell^{(\theta)}, \ell = 1, 2, \ldots, s\right\}, & \text{if } \theta = 1; \\ \left\{(z_1, z_2, \ldots, z_s) : z_\ell \leq q_\ell^{(\theta)}, \ell = 1, 2, \ldots, s\right\} \setminus \cup_{\kappa=1}^{\theta-1} A_\kappa, & \text{if } \theta = 2, \ldots, d; \\ \mathbb{R}^s \setminus \cup_{\kappa=1}^{d} A_\kappa, & \text{if } \theta = d+1. \end{cases} \tag{5}$$

With this definition of A_θ, we can say that the decision maker wants to find the best among systems whose constraint mean configurations fall in A_1 but would consider systems in A_2 if no systems fall in A_1. She would further consider systems in A_3 if no systems fall in A_1 and A_2 and $d \geq 3$, etc.

We assume that the ordered list of threshold vectors is such that when there is no trade-off, the decision maker always prefers "tighter" combinations of threshold values. Consider a case where there are two (non-negative) constraints, the first constraint has three thresholds, and the second constraint has two thresholds. Then it is not possible for the decision maker to prefer (q_1^3, q_2^1) to (q_1^2, q_2^1) in the preference order. Figure 6 shows $A_1, \ldots, A_5$ for an example with $d = 4$ combinations of threshold vectors. We see that $\boldsymbol{q}^{(1)} = (q_1^2, q_2^1)$ does not correspond to the "tightest" combination of threshold values (i.e., (q_1^1, q_2^1)), and similarly $\boldsymbol{q}^{(d)} = (q_1^3, q_2^1)$ does not correspond to the "weakest" combination of threshold values (i.e., (q_1^3, q_2^2)).

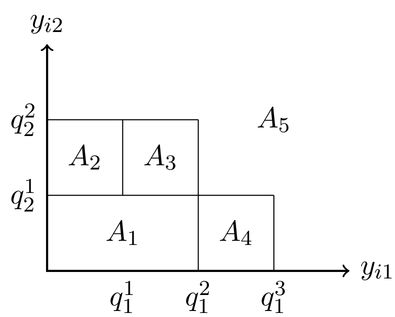


Figure 6: A preference order where the "tightest" ("weakest") combination of thresholds is not "most" ("least") preferred

The following definition will facilitate the efficient implementation of our approaches.

Definition 1. *Constraint ℓ has an increasing preference if $q_\ell^{(\theta)} \leq q_\ell^{(\theta')}$ for any $\theta, \theta' = 1, 2, \ldots, d$ with $\theta < \theta'$.*

We consider the following two examples to further explain Definition 1. Figure 7 shows two preference orders of threshold vectors for two (non-negative) constraints with $d_1 = d_2 = 3$. Based on our definition of threshold vectors, we have $d = 3, \boldsymbol{q}^{(1)} = (q_1^1, q_2^1), \boldsymbol{q}^{(2)} = (q_1^2, q_2^2)$, and $\boldsymbol{q}^{(3)} = (q_1^3, q_2^3)$ in Figure 7a, which satisfies Definition 1 for both constraints. On the other hand, Figure 7b formulates the threshold vectors as $\boldsymbol{q}^{(1)} = (q_1^1, q_2^1), \boldsymbol{q}^{(2)} = (q_1^1, q_2^2), \boldsymbol{q}^{(3)} = (q_1^1, q_2^3), \boldsymbol{q}^{(4)} = (q_1^2, q_2^1)$, etc. We see that constraint 1 has increasing preference whereas constraint 2 does not have increasing preference. Finally, in Figure 6, neither constraint has increasing preference.

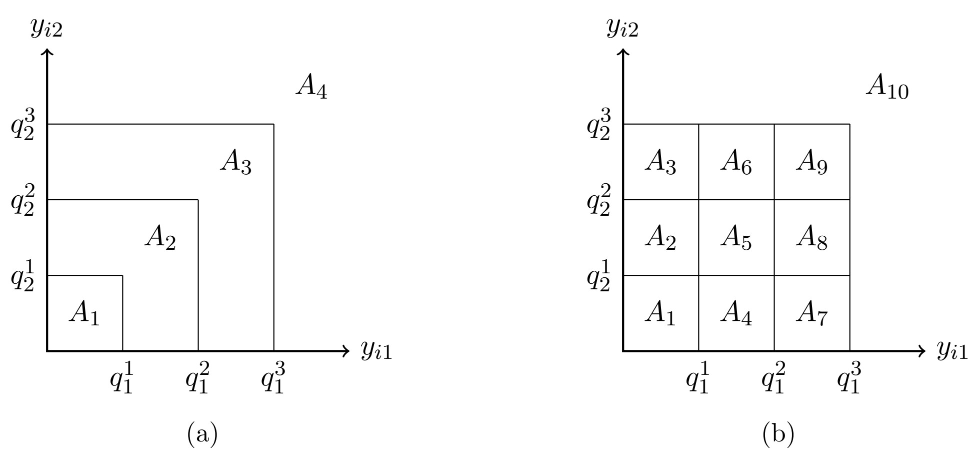


Figure 7: Two preference orders of threshold vectors

3.2.2 Correct Selection

To solve the constrained R&S problem with subjective constraints described in Section 3.2.1, we consider two phases, namely Phase I to identify feasible systems and Phase II to select a system with the largest x_i based on a comparison among feasible systems. These phases are designed to correctly select the best feasible system with respect to the most preferred threshold vector possible, as described in this section.

To check the feasibility of each system with respect to constraint ℓ, [1] introduce a tolerance level, namely $\epsilon_\ell > 0$, for constraint ℓ, which is a positive real value predefined by the decision maker. This is often interpreted as the amount the decision maker is willing to be off from a given threshold value. Consider a threshold value q_ℓ^m for $m = 1, 2, \ldots, d_\ell$. Any systems with $y_{i\ell} \leq q_\ell^m - \epsilon_\ell$ are considered as desirable systems with respect to constraint ℓ and threshold value q_ℓ^m. We let $D_\ell(q_\ell^m)$ denote the set of desirable systems for constraint ℓ and q_ℓ^m. Systems with $y_{i\ell} \geq q_\ell^m + \epsilon_\ell$ are considered as unacceptable systems for constraint ℓ and threshold q_ℓ^m, and are placed in set $U_\ell(q_\ell^m)$. Systems that fall within a tolerance level of q_ℓ^m, which means $q_\ell^m - \epsilon_\ell < y_{i\ell} < q_\ell^m + \epsilon_\ell$, are considered as acceptable systems, placing them in the set $A_\ell(q_\ell^m)$. More specifically,

$$
\begin{aligned}
&D_\ell(q_\ell^m) = \{i \in \Gamma | y_{i\ell} \leq q_\ell^m - \epsilon_\ell\}; \\
&U_\ell(q_\ell^m) = \{i \in \Gamma | y_{i\ell} \geq q_\ell^m + \epsilon_\ell\}; \text{ and} \\
&A_\ell(q_\ell^m) = \{i \in \Gamma | q_\ell^m - \epsilon_\ell < y_{i\ell} < q_\ell^m + \epsilon_\ell\}.
\end{aligned}
$$

When feasibility check is performed to completion (until a decision is made), we let $\text{CD}_{i\ell}(q_\ell^m)$ denote the correct decision event of system i with respect to constraint ℓ and threshold q_ℓ^m, which is defined as declaring system i as feasible if $i \in D_\ell(q_\ell^m)$ and as infeasible if $i \in U_\ell(q_\ell^m)$. Any feasibility decision is considered correct if $i \in A_\ell(q_\ell^m)$. For any threshold vector $\boldsymbol{q}^{(\theta)}$, we say that system i is desirable with respect to $\boldsymbol{q}^{(\theta)}$ when it is desirable with respect to all the constraints, i.e., $i \in D_\ell(q_\ell^{(\theta)})$ for all $\ell = 1, \ldots, s$. System i is unacceptable with respect $\boldsymbol{q}^{(\theta)}$ if it is unacceptable with respect to at least one constraint, i.e., there exists ℓ such that $i \in U_\ell(q_\ell^{(\theta)})$. When system i is acceptable to some (or all) the constraints and desirable with respect to the other constraints, system i is called acceptable with respect to $\boldsymbol{q}^{(\theta)}$.

To select the best system with respect to the primary performance measure in Phase II, the decision maker needs to choose an indifference-zone parameter δ, which is the smallest absolute difference that the decision maker considers significant. More specifically, any system whose primary performance measure is at least δ smaller (larger) than system i is considered as inferior (superior) to system i.

Let θ^* be the smallest θ such that $D_\ell\left(q_\ell^{(\theta)}\right) \neq \emptyset$ for all ℓ. If for each $\theta = 1, \ldots, d$, there exists at least one constraint ℓ_θ such that $D_{\ell_\theta}\left(q_{\ell_\theta}^{(\theta)}\right) = \emptyset$, i.e., θ^* does not exist, then we set $\theta^* = d+1$. If $\theta^* \leq d$, then $\boldsymbol{q}^{(\theta^*)}$ is the most preferable threshold vector possible where at least one desirable system exists. Further, let B denote the set of desirable systems with respect to $\boldsymbol{q}^{(\theta^*)}$ (i.e., $B = \cap_{\ell=1}^s D_\ell\left(q_\ell^{(\theta^*)}\right)$) and let $[b]$ be the index of the best system among the systems in B, so that $x_{[b]} \geq x_i$ for $i, [b] \in B$. Then if $\theta^* \leq d$, the correct selection event is to select a desirable or acceptable system with respect to $\boldsymbol{q}^{(\theta^*)}$ whose primary performance is not inferior to the best system, or an acceptable system with respect to a preferred threshold vector. More specifically,

$$\text{CS} = \Bigg\{ \text{select } i \text{ such that either } i \in \cap_{\ell=1}^s \left(D_\ell\left(q_\ell^{(\theta^*)}\right) \cup A_\ell\left(q_\ell^{(\theta^*)}\right)\right) \text{ and } x_i > x_{[b]} - \delta$$
$$\text{or } i \in \cup_{\theta<\theta^*} \cap_{\ell=1}^s \left(D_\ell\left(q_\ell^{(\theta)}\right) \cup A_\ell\left(q_\ell^{(\theta)}\right)\right) \Bigg\}.$$

If $\theta^* = d+1$, CS is to either declare that no feasible systems exist or identify any acceptable system with respect to any of the threshold vectors $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(d)}$.

3.2.3 Notation and Assumptions

Throughout the paper, we let $\mathbb{1}(\cdot)$ be the indicator function and use the additional notation defined below:

$n_0 \equiv$ initial sample size for each system ($n_0 \geq 2$);

$r_i \equiv$ number of observations so far for system i ($r_i \geq n_0$);

$\bar{X}_i(r_i) \equiv$ average value of $X_{i1}, \ldots, X_{ir_i}$ for system i;

$\bar{Y}_{i\ell}(r_i) \equiv$ average value of $Y_{i\ell 1}, \ldots, Y_{i\ell r_i}$ for system i and constraint ℓ;

$$S^2_{X_{ij}}(n_0) \equiv \text{sample variance of } X_{i1} - X_{j1}, \ldots, X_{in_0} - X_{jn_0} \text{ between system } i \text{ and } j;$$

$$S^2_{Y_{i\ell}}(n_0) \equiv \text{sample variance of } Y_{i\ell 1}, \ldots, Y_{i\ell n_0} \text{ for system } i \text{ and constraint } \ell;$$

$$R(r_i; v, w, z) \equiv \max\left\{0, \frac{(n_0-1)wz}{v} - \frac{v}{2c} r_i\right\} \text{ for } v, w, z \in \mathbb{R}^+ \text{ and } c \in \{1, 2, \ldots\};$$

$$g(\eta) \equiv \sum_{j=1}^{c} (-1)^{j+1} \left(1 - \frac{1}{2}\mathbb{1}(j = c)\right) \times \left(1 + \frac{2\eta(2c-j)j}{c}\right)^{-(n_0-1)/2};$$

$$\alpha \equiv \text{overall nominal error for a procedure under consideration.}$$

Note that an integer parameter c is required for both $R(r_i; v, w, z)$ and $g(\eta)$. This is a user-defined parameter that impacts the shape of the continuation region defined by $(-R(r_i; v, w, z), R(r_i; v, w, z))$ (it becomes a longer triangle as c increases). The choice $c = 1$ is recommended as it guarantees a unique and easy solution when computing the implementation parameter η from $g(\eta)$. [10] also suggest that $c = 1$ is a good choice when the decision maker does not have information about the systems' mean configuration. The experimental results in the paper are based on $c = 1$.

Our statistical analysis of our proposed procedures will rely on the following two assumptions.

Assumption 2. *For each system i, where $i = 1, \ldots, k$, we have*

$$\begin{bmatrix} X_{in} \\ Y_{i1n} \\ \vdots \\ Y_{isn} \end{bmatrix} \overset{iid}{\sim} N_{s+1}\left(\begin{bmatrix} x_i \\ y_{i1} \\ \vdots \\ y_{is} \end{bmatrix}, \Sigma_i\right), \qquad n = 1, 2, \ldots$$

where $\overset{iid}{\sim}$ denotes independent and identically distributed, N_{s+1} denotes $(s+1)$-dimensional multivariate normal, and Σ_i is the $(s+1)\times(s+1)$ covariance matrix of the vector $(X_{in}, Y_{i1n}, \ldots, Y_{isn})$.

Normally distributed data is a common assumption used in many R&S procedures due to the fact that it can be justified by the Central Limit Theorem when observations are either within-replication averages or batch means [12]. Moreover, primary and secondary performance measures are usually correlated. When common random numbers (CRN) are introduced in simulating observations from each system, observations between systems are correlated. Our formulation allows correlations between both performance measures and systems.

Assumption 3. *If* $\theta^* \leq d$*, then for any system* $i \in \cap_{\ell=1}^{s} \left(D_\ell(q_\ell^{(\theta^*)}) \cup A_\ell(q_\ell^{(\theta^*)}) \right)$*, where* $i \neq [b]$*, we assume* $x_i \leq x_{[b]} - \delta$*.*

Assumption 3 implies that there exists only one best system $[b]$ and any systems that are desirable or acceptable with respect to $q_\ell^{(\theta^*)}$ for all constraint $\ell = 1, \ldots, s$ are inferior to system $[b]$. This assumption is a standard assumption for proving the statistical validity of IZ approaches in the R&S literature.

3.3 Sequentially-Running Procedures

In this section, we present procedures that implement Phases I and II sequentially. The outline of this section is as follows. Section 3.3.1 describes a sequentially-running procedure and Section 3.3.2 proves its statistical validity. A variation of the sequentially-running procedure is discussed in Section 3.3.3.

3.3.1 Procedure $\mathcal{ZAK}^{\mathcal{R}}$

A proposed sequentially-running procedure named $\mathcal{ZAK}^{\mathcal{R}}$ ("restart") is described in Algorithm 3 (we use $|S|$ to denote the cardinality of a set S). $\mathcal{ZAK}^{\mathcal{R}}$ starts by executing Phase I for all systems to identify the most preferred threshold vector possible, $\boldsymbol{q}^{(\theta^*)}$, as well as the feasible systems with respect to $\boldsymbol{q}^{(\theta^*)}$. The parameter θ keeps track of our current estimate of θ^* (initially $\theta = d$), M is a set of systems that are in consideration (initially M contains all the systems, i.e., $M = \Gamma$), and F is a set of systems that are declared feasible with respect to threshold vector $\boldsymbol{q}^{(\theta)}$ (initially $F = \emptyset$). The procedure returns $Z_{i\ell}^m = 1 (Z_{i\ell}^m = 0)$ if system i is declared feasible (infeasible) with respect to constraint ℓ and threshold q_ℓ^m, and $Z_{i\ell}^m = 2$ if no decision is made about the feasibility of system i with respect to threshold q_ℓ^m on constraint ℓ. Notice that once a system is declared feasible with respect to threshold vector $\boldsymbol{q}^{(\theta)}$ where $1 \leq \theta \leq d-1$, we do not need to check feasibility for any systems with respect to the less preferred threshold vectors $\boldsymbol{q}^{(\theta+1)}, \ldots, \boldsymbol{q}^{(d)}$. Similarly, once a system is declared feasible with respect to $\boldsymbol{q}^{(\theta)}$ where $1 \leq \theta \leq d$, then we do not need to collect additional observations from any inferior systems whose feasibility with respect to $\boldsymbol{q}^{(\theta)}$ is known and that are infeasible with respect to all the preferred threshold vectors $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta-1)}$.

Our approach for handling multiple threshold values builds on the work of $\mathcal{RF}$ discussed in Chapter 2, an efficient fully-sequential procedure for checking the feasibility of all systems with respect to all constraints and all thresholds simultaneously. Theorem 1 in Chapter 2 shows that once a system i is declared feasible with respect to a threshold q_ℓ^m such that $q_\ell^m \geq y_{i\ell}+\epsilon_\ell$, this system will be declared feasible with respect to all thresholds $q_\ell^{m+1}, \ldots, q_\ell^{d_\ell}$ on constraint ℓ. Similarly, if a system i is declared infeasible with respect to a threshold q_ℓ^m such that $q_\ell^m \leq y_{i\ell} - \epsilon_\ell$, then this system will be declared infeasible with respect to all the thresholds $q_\ell^1, \ldots, q_\ell^{m-1}$. This fact is essential in our proposed procedures.

The sequentially-running procedure $\mathcal{ZAK}^{\mathcal{R}}$ performs Phase II on the surviving systems from the completion of Phase I. More specifically, it selects the best system with respect to the primary performance measure among the subset of systems that are declared feasible with respect to the most preferred threshold vector possible identified in Phase I. In order to prove the statistical validity of our proposed sequentially-running procedure and avoid storing simulation results, we avoid the correlation between the primary and secondary performance measures by not recycling any observations from Phase I and instead restarting "from scratch" when implementing comparisons in Phase II. Moreover, when CRN are used to compare systems in Phase II, we assume that the implementation of CRN is such that the simulation results for any survising system in Phase II do not depend on the set of surviving systems F.

Remark: In $\mathcal{ZAK}^{\mathcal{R}}$, it is possible to use r rather than r_i in Phase I. However, in Section 3.3.3 we propose a more efficient heuristic version of $\mathcal{ZAK}^{\mathcal{R}}$ (namely $\mathcal{ZAK}$) that recycles observations from Phase I in Phase II and that procedure requires the number of observations taken in Phase I for each system. Therefore, we state $\mathcal{ZAK}^{\mathcal{R}}$ with r_i rather than r in Phase I.

Algorithm 3 Procedure $\mathcal{ZAK}^{\mathcal{R}}$

[**Setup**:] Select the overall nominal confidence level $1-\alpha$ and choose $\alpha_f, \alpha_c > 0$ such that $(1-\alpha_f)(1-\alpha_c) = 1-\alpha$. Choose tolerance levels $\epsilon_1, \ldots, \epsilon_s$, indifference-zone parameter δ, threshold vectors $\{\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(d)}\}$, and associated index vectors $\{\boldsymbol{I}^{(1)}, \boldsymbol{I}^{(2)}, \ldots, \boldsymbol{I}^{(d)}\}$. Set $M = \Gamma$ and $Z_{i\ell}^m = 2$ for all $i \in M, \ell = 1, \ldots, s$, and $m = 1, \ldots, d_\ell$. Set $F = \emptyset$ and $\theta = d$. Set η_f such that $g(\eta_f) = \alpha_f'$, where α_f' is set as the solution to

$$\left(1 - \min\{s,d\}\alpha_f'\right)^{k-1} \times (1 - s\alpha_f') = 1 - \alpha_f, \text{ if systems are simulated independently;}$$

and set as

$$\alpha_f' = \alpha_f / \left[(k-1)\min\{s,d\} + s\right], \text{ if systems are simulated under CRN.}$$

Add any constraint ℓ, where $\ell = 1, \ldots, s$, with increasing preference to set IP.

[**Initialization for Phase I**:]

for each system $i \in M$ **do**

- Obtain n_0 observations $Y_{i\ell 1}, Y_{i\ell 2}, \ldots, Y_{i\ell n_0}$ for $\ell = 1, 2, \ldots, s$.
- Compute $\bar{Y}_{i\ell}(n_0)$ and $S^2_{Y_{i\ell}}(n_0)$.
- Set $r_i = n_0, \mathrm{ON}_i = \{1, 2, \ldots, s\}$, and $\mathrm{ON}_{i\ell} = \{1, \ldots, d_\ell\}$ for $\ell = 1, 2, \ldots, s$.

end for

[**Feasibility Check**:]

for each system $i \in M$ **do**
  for $\ell \in \mathrm{ON}_i$ **do**
    for $m \in \mathrm{ON}_{i\ell}$ **do**,
      If $\bar{Y}_{i\ell}(r_i) + R(r_i; \epsilon_\ell, \eta_f, S^2_{Y_{i\ell}}(n_0))/r_i \le q_\ell^m$, set $Z_{i\ell}^m = 1$ and $\mathrm{ON}_{i\ell} = \mathrm{ON}_{i\ell} \setminus \{m\}$.
      If $\bar{Y}_{i\ell}(r_i) - R(r_i; \epsilon_\ell, \eta_f, S^2_{Y_{i\ell}}(n_0))/r_i \ge q_\ell^m$, set $Z_{i\ell}^m = 0$ and $\mathrm{ON}_{i\ell} = \mathrm{ON}_{i\ell} \setminus \{m\}$.
    end for
    If $\mathrm{ON}_{i\ell} = \emptyset$, set $\mathrm{ON}_i = \mathrm{ON}_i \setminus \{\ell\}$.
  end for
  If $\exists$ minimum $\kappa \le \theta$ s.t. $\prod_{\ell=1}^s Z_{i\ell}^{I_\ell^{(\kappa)}} = 1$, and either $\kappa < \theta$ or $i \notin F$, then

- If $\kappa < \theta$, then set $F = \emptyset, \theta = \kappa$, and for all $j \in M$ delete q_ℓ^m from $\mathrm{ON}_{j\ell}$ if $\ell \in \mathrm{IP}$ and $m > I_\ell^{(\theta)}$ (if $\ell \notin \mathrm{IP}$, then q_ℓ^m can be removed from $\mathrm{ON}_{j\ell}$ if $I_\ell^{(\theta')} \neq m$ for all $\theta' \le \kappa$), and set $\mathrm{ON}_j = \mathrm{ON}_j \setminus \{\ell\}$ if $\mathrm{ON}_{j\ell} = \emptyset$.
- Add system i to F.

  If $\prod_{\ell=1}^s Z_{i\ell}^{I_\ell^{(\theta)}} = 0$ or 1 and either $\theta = 1$ or $\prod_{\ell=1}^s Z_{i\ell}^{I_\ell^{(\kappa)}} = 0$ for all $\kappa = 1, \ldots, \theta - 1$, then remove system i from M.

end for

[**Stopping Condition for Phase I**]:

If $M \neq \emptyset$, then for each system $i \in M$, set $r_i = r_i + 1$, take one additional observation $Y_{i\ell r_i}$, and update $\bar{Y}_{i\ell}(r_i)$ for $\ell \in \mathrm{ON}_i$, then go to [**Feasibility Check**]. Else, check the following conditions.

- If $|F| = 0$, stop and conclude no feasible systems;
- If $|F| = 1$, stop and return the system in F as the best; or
- If $|F| > 1$, go to [**Initialization for Phase II**].

[**Initialization for Phase II**:] Let η_c be a solution to $g(\eta_c) = \alpha_c'$, where

$$\alpha_c' = \begin{cases} 1 - (1-\alpha_c)^{1/(k-1)}, & \text{if systems are simulated independently;} \\ \alpha_c/(k-1), & \text{if systems are simulated under CRN.} \end{cases}$$

Let $M = F$ be the set of systems still in contention. For each system $i \in M$, perform an entirely new simulation and obtain $X_{i1}, \ldots, X_{in_0}$ independent of any $Y_{j\ell n}$ generated in Phase I. Compute $\bar{X}_i(n_0)$ and $S^2_{X_{ij}}(n_0)$ for $i, j \in M$ and $i \neq j$. Set $r = n_0$ and go to [**Comparison**].

[**Comparison**:] For $i, j \in M$ s.t. $i \neq j$ and

$$r\bar{X}_i(r) > r\bar{X}_j(r) + R(r; \delta, \eta_c, S^2_{X_{ij}}(n_0)),$$

eliminate j from M.

[**Stopping Condition for Phase II**:] If $|M| = 1$, then stop and select the system in M as the best. Otherwise, for each system $i \in M$, take one additional observation $X_{i,r+1}$ independent of any $Y_{j\ell n}$ generated in Phase I and compute $\bar{X}_i(r+1)$. Then, set $r = r + 1$ and go to [**Comparison**].

3.3.2 Statistical Validity of Procedure $\mathcal{ZAK}^{\mathcal{R}}$

In this section, we prove the statistical validity of the $\mathcal{ZAK}^{\mathcal{R}}$ procedure presented in Algorithm 3. Before presenting the main results, we need more definitions. Let θ^* be defined as in Section 3.2.2. We define the sets S_a and S_u as follows:

$$S_a = \text{ set of acceptable systems with respect to at least one of the threshold vectors } \boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta^*-1)};$$

$$S_u = \begin{cases} \text{set of unacceptable systems with respect to } \boldsymbol{q}^{(\theta^*)} \text{ among systems in } \Gamma \backslash S_a, & \text{if } \theta^* \leq d; \\ \Gamma \setminus S_a, & \text{if } \theta^* = d+1. \end{cases}$$

We further define

$$S_{a'} = \begin{cases} \text{set of acceptable systems with respect to } \boldsymbol{q}^{(\theta^*)} \text{ among systems in } \Gamma \backslash S_a, & \text{if } \theta^* \leq d; \\ \emptyset, & \text{if } \theta^* = d+1; \end{cases}$$

$$S_d = \begin{cases} \text{set of desirable systems with respect to } \boldsymbol{q}^{(\theta^*)} \text{ among systems in } \Gamma \backslash (S_a \cup [b]), & \text{if } \theta^* \leq d; \\ \emptyset, & \text{if } \theta^* = d+1. \end{cases}$$

We then let $j_a = |S_a|, j_{a'} = |S_{a'}|, j_d = |S_d|$, and $j_u = |S_u|$, and therefore $j_a + j_{a'} + j_d + j_u + \mathbb{1}(\theta^* \leq d) = k$. For correct selection, we must select a system in $S_a \cup \{[b]\}$ and eliminate the systems in $S_{a'} \cup S_d \cup S_u$ when $\theta^* \leq d$ (under Assumption 3); when $\theta^* = d+1$, CS involves eliminating all systems in S_u, and either declaring all systems infeasible or selecting a system in S_a.

To illustrate, consider a problem with two constraints where the first constraint has two thresholds and the second constraint has three thresholds. We consider all $d = 6$ possible threshold vectors $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(6)}$. Figure 8 shows possible (non-negative) secondary performance means and thresholds where the shaded areas represent acceptable regions with respect to one or more threshold vectors, and $A_1, \ldots, A_6$ are defined as in equation (5) and are separated by the solid lines. Suppose $\theta^* = 5$ in this example. Due to the definition of θ^*, there do not exist any desirable systems with respect to $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(4)}$, but there exists at least one system that falls in the desirable region of $\boldsymbol{q}^{(5)}$. It is possible that there exist acceptable systems with respect to $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(4)}$. Figure 8 shows systems a and b as two examples of acceptable systems with respect to preferred threshold vectors (i.e., $a, b \in S_a$). Note that system a is acceptable with respect to $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{q}^{(3)}$, and $\boldsymbol{q}^{(4)}$

and desirable with respect to $\boldsymbol{q}^{(5)}$, while system b is acceptable with respect to $\boldsymbol{q}^{(3)}$ but unacceptable to $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{q}^{(4)}$, and $\boldsymbol{q}^{(5)}$. System c is acceptable with respect to $\boldsymbol{q}^{(5)}$ (i.e., $c \in S_{a'}$) and unacceptable with respect to $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(4)}$.

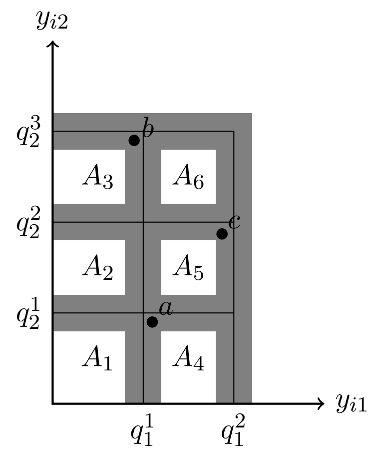


Figure 8: Regions for two secondary performance measures and six threshold vectors

To prove the statistical validity of $\mathcal{ZAK}^{\mathcal{R}}$, we start with the following lemma.

Lemma 2. *For system i and constraint ℓ with specific threshold value q_ℓ^m, the* [**Feasibility Check**] *steps in $\mathcal{ZAK}^{\mathcal{R}}$ that run to completion ensure* $\Pr(\mathrm{CD}_{i\ell}(q_\ell^m)) \geq 1 - \alpha'_f$.

Proof. When system i and constraint ℓ with specific threshold q_ℓ^m are considered separately, the [**Feasibility Check**] steps in $\mathcal{ZAK}^{\mathcal{R}}$ either conclude a feasibility decision or eliminate threshold q_ℓ^m for further consideration (when system i is declared feasible with respect to a threshold vector and all preferred threshold vectors do not involve threshold value q_ℓ^m on constraint ℓ). We see that when a feasibility decision is concluded, the [**Feasibility Check**] steps in $\mathcal{ZAK}^{\mathcal{R}}$ are essentially the same as for the statistically-valid feasibility check procedure $\mathcal{F}$ in [1] for a single system and a single constraint with one threshold value with confidence level $1 - \alpha'_f$. The result now follows from the special case of Theorem 1 in [1] with $k = 1$. □

We then introduce the following definitions for $i \in \Gamma$ and present a lemma that is essential in proving the statistical validity of $\mathcal{ZAK}^{\mathcal{R}}$:

$$\mathcal{A}_1^*(i) = \text{system } i \text{ is declared infeasible to } \boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\min\{\theta^*, d\})};$$

$$\mathcal{A}_2^*(i) = \text{system } i \text{ is declared infeasible to } \boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta^*-1)} \text{ if } 1 < \theta^* \leq d;$$

$$\mathcal{B}_1^* = \text{system } [b] \text{ is declared feasible to } \boldsymbol{q}^{(\theta^*)} \text{ if } \theta^* \leq d.$$

Lemma 3. *For a particular system i, the* [**Feasibility Check**] *steps in $\mathcal{ZAK}^{\mathcal{R}}$ ensure*

$$\Pr\left(\mathcal{A}_1^*(i)\right) \geq 1 - \min\{s, d\}\alpha_f', \ \textit{if } i \in S_u;$$

$$\Pr\left(\mathcal{A}_2^*(i)\right) \geq 1 - \min\{s, d-1\}\alpha_f', \ \textit{if } i \in S_d \cup S_{a'} \ \textit{and } 1 < \theta^* \leq d;$$

$$\Pr\left(\mathcal{B}_1^*\right) \geq 1 - s\alpha_f', \ \textit{if } \theta^* \leq d.$$

Proof. First, consider $i \in S_u$. Then system i must be unacceptable to $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta^*)}$ because it is unacceptable to $\boldsymbol{q}^{(\theta^*)}$, not in S_a, and there are no desirable systems with respect to $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta^*-1)}$.

When $\theta^* \leq d$, as system i is unacceptable with respect to $\boldsymbol{q}^1, \ldots, \boldsymbol{q}^{(\theta^*)}$, then for each $\kappa = 1, \ldots, \theta^*$, there exist at least one constraint ℓ_κ such that $y_{i\ell_\kappa} \geq q_{\ell_\kappa}^{(\kappa)} + \epsilon_{\ell_\kappa}$. Then we have

$$\Pr\left(\mathcal{A}_1^*(i)\right) \geq \Pr\left(\cap_{\kappa=1}^{\theta^*} \mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \geq 1 - \sum_{\kappa=1}^{\theta^*} \Pr\left(\mathrm{ICD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \geq 1 - d\alpha_f', \tag{6}$$

where we use $\mathrm{ICD}_{i\ell}(q_\ell^m)$ to denote the event of incorrect decision of system i with respect to constraint ℓ and threshold q_ℓ^m. The first inequality holds because declaring system i infeasible to constraint ℓ_κ is sufficient to declare system i infeasible to threshold vector $\boldsymbol{q}^{(\kappa)}$ and it is not possible to declare a system feasible with respect to a threshold vector without completing the comparison with all thresholds in that vector. The second inequality holds due to the Bonferroni inequality, and the last inequality holds due to Lemma 2 and the fact of $\theta^* \leq d$.

Observe that since there are only s constraints, the set $L = \{\ell_1, \ldots, \ell_{\theta^*}\}$ can have at most s distinct values. For $\ell \in L$, let $I_{i\ell}$ denote the largest threshold index on constraint ℓ that system i is unacceptable to, i.e.,

$$I_{i\ell} = \max_{1 \leq m \leq d_\ell} \left\{m : y_{i\ell} \geq q_\ell^m + \epsilon_\ell\right\}.$$

Thus, we know that $q_\ell^1 < q_\ell^2 < \cdots < q_\ell^{I_{i\ell}} \leq y_{i\ell} - \epsilon_\ell$ on constraint ℓ. Due to the discussion in Theorem 1 of Chapter 2, we know that $\mathrm{CD}_{i\ell}(q_\ell^{I_{i\ell}}) \subseteq \cdots \subseteq \mathrm{CD}_{i\ell}(q_\ell^2) \subseteq \mathrm{CD}_{i\ell}(q_\ell^1)$. Then $\mathrm{CD}_{i\ell}(q_\ell^{I_{i\ell}}) \subseteq \mathrm{CD}_{i\ell}(q_\ell^{(\kappa)})$ for $\kappa = 1, \ldots, \theta^*$ with $\ell_\kappa = \ell$. Thus, we also have

$$\begin{aligned}\Pr(\mathcal{A}_1^*(i)) &\geq \Pr\left(\cap_{\kappa=1}^{\theta^*}\mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \geq \Pr\left(\cap_{\ell\in L}\mathrm{CD}_{i\ell}(q_\ell^{I_{i\ell}})\right)\\ &\geq 1 - \sum_{\ell\in L}\Pr\left(\mathrm{ICD}_{i\ell}(q_\ell^{I_{i\ell}})\right) \geq 1 - |L|\alpha'_f \geq 1 - s\alpha'_f, \end{aligned} \tag{7}$$

where the third inequality is due to the Bonferroni inequality and the forth inequalty is due to Lemma 2. By comparing equations (6) and (7), we conclude that $\Pr(\mathcal{A}_1^*(i)) \geq 1 - \min\{s, d\}\alpha'_f$.

When $\theta^* = d + 1$, a similar argument yields

$$\Pr(\mathcal{A}_1^*(i)) \geq \Pr\left(\cap_{\kappa=1}^{d}\mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \geq 1 - \sum_{\kappa=1}^{d}\Pr\left(\mathrm{ICD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \geq 1 - d\alpha'_f,$$

and, defining $L = \{\ell_1, \ldots, \ell_d\}$,

$$\begin{aligned}\Pr(\mathcal{A}_1^*(i)) &\geq \Pr\left(\cap_{\kappa=1}^{d}\mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \geq \Pr\left(\cap_{\ell\in L}\mathrm{CD}_{i\ell}(q_\ell^{I_{i\ell}})\right)\\ &\geq 1 - \sum_{\ell\in L}\Pr\left(\mathrm{ICD}_{i\ell}(q_\ell^{I_{i\ell}})\right) \geq 1 - |L|\alpha'_f \geq 1 - s\alpha'_f. \end{aligned}$$

Therefore, $\Pr(\mathcal{A}_1^*(i)) \geq 1 - \min\{s, d\}\alpha'_f$.

Now, consider $i \in S_d \cup S_{a'}$ with $1 < \theta^* \leq d$. As system i is not in S_a and there are no desirable systems with respect to $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta^*-1)}$, system i must be unacceptable with respect to $\boldsymbol{q}^1, \ldots, \boldsymbol{q}^{(\theta^*-1)}$. Then for each $\kappa = 1, \ldots, \theta^*-1$, there exist at least one constraint ℓ_κ such that $y_{i\ell_\kappa} \geq q_{\ell_\kappa}^{(\kappa)} + \epsilon_{\ell_\kappa}$. Due to a similar argument as for $i \in S_u$, we have

$$\Pr(\mathcal{A}_2^*(i)) \geq \Pr\left(\cap_{\kappa=1}^{\theta^*-1}\mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \geq 1 - \sum_{\kappa=1}^{\theta^*-1}\Pr\left(\mathrm{ICD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \geq 1 - (d-1)\alpha'_f.$$

Based on a similar definition $L = \{\ell_1, \ldots, \ell_{\theta^*-1}\}$ and the discussion above, we have

$$\begin{aligned}\Pr(\mathcal{A}_2^*(i)) &\geq \Pr\left(\cap_{\kappa=1}^{\theta^*-1}\mathrm{CD}_{i\ell_\kappa}(q_{\ell_\kappa}^{(\kappa)})\right) \geq \Pr\left(\cap_{\ell\in L}\mathrm{CD}_{i\ell}(q_\ell^{I_{i\ell}})\right)\\ &\geq 1 - \sum_{\ell\in L}\Pr\left(\mathrm{ICD}_{i\ell}(q_\ell^{I_{i\ell}})\right) \geq 1 - |L|\alpha'_f \geq 1 - s\alpha'_f. \end{aligned}$$

Therefore, we have $\Pr(\mathcal{A}_2^*(i)) \geq 1 - \min\{s, d-1\}\alpha'_f$.

Finally, for $[b]$, when $\theta^* \leq d$, we have

$$\Pr\left(\mathcal{B}_1^*\right) = \Pr\left(\cap_{\ell=1}^{s} \mathrm{CD}_{i\ell}(q_\ell^{(\theta^*)})\right) \geq 1 - \sum_{\ell=1}^{s} \Pr\left(\mathrm{ICD}_{i\ell}(q_\ell^{(\theta^*)})\right) \geq 1 - s\alpha'_f,$$

where the last inequality is due to Lemma 2. □

For Lemma 3, one may notice that $d > s$ holds in most cases, and therefore $\Pr\left(\mathcal{A}_1^*(1)\right) \geq 1 - s\alpha'_f$ and $\Pr\left(\mathcal{A}_2^*(1)\right) \geq 1 - s\alpha'_f$ hold in most cases. Note that when $d \geq s$ and the systems are simulated independently, the inplementation parameter α'_f has a closed-form solution as

$$\alpha'_f = \frac{1}{s}\left[1 - (1 - \alpha_f)^{1/k}\right].$$

When $d < s$, one may need to find α'_f by numerically solving $(1-d\alpha'_f)^{k-1} \times (1-s\alpha'_f) = 1-\alpha_f$.

We then use CS_i to denote the correct selection between system $i \in S_{a'} \cup S_d$ and the best system $[b]$ and introduce the following lemma.

Lemma 4. *Given i such that $x_i \leq x_{[b]} - \delta$, the* **[Comparison]** *steps for system i and $[b]$ in $\mathcal{ZAK}^{\mathcal{R}}$ that run to completion ensure*

$$\Pr\left(\mathrm{CS}_i\right) \geq 1 - \alpha'_c.$$

Proof. When only system i and $[b]$ are considered, the **[Comparison]** steps in $\mathcal{ZAK}^{\mathcal{R}}$ are the same as the statistically-valid selection-of-the-best procedure provided in [10] when two systems are considered with confidence level $1 - \alpha'_c$. Therefore, the result follows from the special case of Theorem 1 of [10] with $k = 2$. □

We are now ready to prove the statistical validity of $\mathcal{ZAK}^{\mathcal{R}}$.

Theorem 4. *Under Assumptions 2 and 3, the $\mathcal{ZAK}^{\mathcal{R}}$ procedure guarantees*

$$\Pr\{\mathrm{CS}\} \geq 1 - \alpha.$$

Proof. We consider two cases, namely when $\theta^* \leq d$ and $\theta^* = d + 1$.

Case 1: $\theta^* \leq d$.

Note that any systems in $(S_{a'} \cup S_d)$ should not be declared feasible with respect to a more preferred threshold vector $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta^*-1)}$ as they could be selected as the best system otherwise. More specifically, we consider the following four events.

$\mathcal{A}_1^* =$ all systems in S_u are eliminated by infeasibility $= \cap_{i \in S_u} \mathcal{A}_1^*(i)$;

$\mathcal{A}_2^* =$ all systems in $(S_{a'} \cup S_d)$ are declared infeasible to thresholds $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta^*-1)}$

$= \cap_{i \in S_{a'} \cup S_d} \mathcal{A}_2^*(i)$ when $\theta^* > 1$;

$\mathcal{B}_2^* =$ system $[b]$ would be selected as the best system among the systems in $S_{a'} \cup S_d$;

$\mathcal{B}^* =$ system $[b]$ is declared feasible with respect to $\boldsymbol{q}^{(\theta^*)}$ and is selected as the best system among the surviving systems from Phase I.

Notice that $\mathcal{B}_1^* \cap \mathcal{B}_2^* \subseteq \mathcal{B}^*$ and $\mathcal{A}_2^*$ is not defined when $\theta^* = 1$. This means

$$\Pr\{\mathrm{CS}\} \geq \begin{cases} \Pr(\mathcal{A}_1^* \cap \mathcal{B}^*), & \text{if } \theta^* = 1; \\ \Pr(\mathcal{A}_1^* \cap \mathcal{A}_2^* \cap \mathcal{B}^*), & \text{if } \theta^* > 1. \end{cases}$$

We see that $\Pr\{\mathrm{CS}\}$ achieves its lower bound when $\theta^* > 1$ (because the bounds on $\Pr(\mathcal{A}_1^*), \Pr(\mathcal{B}_1^*)$, and $\Pr(\mathcal{B}_2^*)$ below do not depend on the value of θ^*), and thus we focus on the case when $\theta^* > 1$. We also see that $\mathcal{A}_1^*, \mathcal{A}_2^*$, and $\mathcal{B}_1^*$ are independent events when systems are simulated independently but are dependent events when systems are simulated under CRN. As we discard observations from Phase I and completely restart for Phase II, and as $\mathcal{B}_2^*$ involves making the correct selection from all systems in $S_{a'} \cup S_d$ (not only the ones surviving from Phase I), $\mathcal{B}_2^*$ is independent from $\mathcal{A}_1^*, \mathcal{A}_2^*$, and $\mathcal{B}_1^*$. We have

$$\begin{aligned} \Pr\{\mathrm{CS}\} &\geq \Pr(\mathcal{A}_1^* \cap \mathcal{A}_2^* \cap \mathcal{B}^*) \geq \Pr(\mathcal{A}_1^* \cap \mathcal{A}_2^* \cap \mathcal{B}_1^* \cap \mathcal{B}_2^*) \\ &= \begin{cases} \Pr(\mathcal{A}_1^*) \times \Pr(\mathcal{A}_2^*) \times \Pr(\mathcal{B}_1^*) \times \Pr(\mathcal{B}_2^*), & \text{if systems are simulated independently;} \\ [\Pr(\mathcal{A}_1^*) + \Pr(\mathcal{A}_2^*) + \Pr(\mathcal{B}_1^*) - 2] \times \Pr(\mathcal{B}_2^*), & \text{if systems are simulated under CRN.} \end{cases} \end{aligned}$$

We discuss the cases depending on whether systems are simulated independently or under CRN. When systems are simulated independently, by Lemma 3, we have

$$\Pr(\mathcal{A}_1^*) \geq (1 - \min\{s, d\}\alpha'_f)^{j_u};$$

$$\Pr(\mathcal{A}_2^*) \geq (1 - \min\{s, d-1\}\alpha'_f)^{j_{a'}+j_d} = (1 - \min\{s, d-1\}\alpha'_f)^{k-j_a-j_u-1};$$

$$\Pr(\mathcal{B}_1^*) \geq 1 - s\alpha'_f.$$

Let N_{ij} denote the number of observations taken for system i before a comparison decision is made between systems i and j, and let N_i denote the maximum number of observations that system i takes within Phase II. That is

$$N_{ij} = \left\lceil \frac{2c\eta_c(n_0-1)S^2_{X_{ij}}(n_0)}{\delta^2} \right\rceil, \text{ and } N_i = \max_{j \neq i} N_{ij}.$$

Then we have

$$\begin{aligned}
\Pr(\mathcal{B}_2^*) &\geq \Pr\left(\cap_{i \in S_{a'} \cup S_d} \mathrm{CS}_i\right) \\
&= \mathrm{E}\left[\Pr\left\{\cap_{i \in (S_d \cup S_{a'})} \mathrm{CS}_i \middle| X_{[b]1}, \ldots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right] \\
&= \mathrm{E}\left[\prod_{i \in (S_d \cup S_{a'})} \Pr\left\{\mathrm{CS}_i \middle| X_{[b]1}, \ldots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right] \\
&\geq \prod_{i \in (S_d \cup S_{a'})} \mathrm{E}\left[\Pr\left\{\mathrm{CS}_i \middle| X_{[b]1}, \ldots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right] \\
&= \prod_{i \in (S_d \cup S_{a'})} \Pr\left\{\mathrm{CS}_i\right\} \geq \prod_{i \in (S_d \cup S_{a'})} (1-\alpha_c') \\
&= (1-\alpha_c')^{j_d + j_{a'}} \geq (1-\alpha_c')^{k-j_u-j_a-1},
\end{aligned}$$

where the second inequality holds due to Lemma 2.4 in [17] and the third inequality follows from Lemma 4.

Thus, we know that

$$\begin{aligned}
\Pr\{\mathrm{CS}\} &\geq (1-\min\{s,d\}\alpha_f')^{j_u} \times (1-\min\{s,d-1\}\alpha_f')^{k-j_a-j_u-1} \times (1-s\alpha_f') \times (1-\alpha_c')^{k-j_u-j_a-1} \\
&\geq (1-\min\{s,d\}\alpha_f')^{j_u} \times (1-\min\{s,d\}\alpha_f')^{k-j_a-j_u-1} \times (1-s\alpha_f') \times (1-\alpha_c')^{k-j_u-j_a-1} \\
&= (1-\min\{s,d\}\alpha_f')^{k-j_a-1} \times (1-s\alpha_f') \times (1-\alpha_c')^{k-j_u-j_a-1} \\
&\geq (1-\min\{s,d\}\alpha_f')^{k-1} \times (1-s\alpha_f') \times (1-\alpha_c')^{k-1} \\
&= (1-\alpha_f) \times \left[(1-\alpha_c)^{1/(k-1)}\right]^{k-1} = (1-\alpha_f)(1-\alpha_c) = 1-\alpha,
\end{aligned}$$

where the third inequality holds since the lower bound of $(1-\min\{s,d\}\alpha_f')^{k-j_a-1}$ is achieved when $j_a = 0$ for $0 \leq 1-\min\{s,d\}\alpha_f' < 1$, and the lower bound of $(1-\alpha_c')^{k-j_u-j_a-1}$ is achieved when $j_a = j_u = 0$ for $0 \leq 1-\alpha_c' < 1$.

When systems are simulated under CRN, by Lemmas 3, 4, and the Bonferroni inequality, we have

$$
\begin{aligned}
&\Pr(\mathcal{A}_1^*) \geq 1 - j_u \min\{s, d\}\alpha'_f;\\
&\Pr(\mathcal{A}_2^*) \geq 1 - (j_{a'} + j_d)\min\{s, d-1\}\alpha'_f = 1 - (k - j_a - j_u - 1)\min\{s, d-1\}\alpha'_f;\\
&\Pr(\mathcal{B}_1^*) \geq 1 - s\alpha'_f;\\
&\Pr(\mathcal{B}_2^*) \geq \Pr\left(\cap_{i \in S_{a'} \cup S_d} \mathrm{CS}_i\right) \geq 1 - \sum_{i \in (S_d \cup S_{a'})} \Pr(\mathrm{ICS}_i) \geq 1 - (j_d + j_{a'})\alpha'_c\\
&\qquad = 1 - (k - j_u - j_a - 1)\alpha'_c,
\end{aligned}
$$

where ICS_i denotes the incorrect selection event between system $i \in S_d \cup S_{a'}$ and system $[b]$. Thus,

$$
\begin{aligned}
\Pr\{\mathrm{CS}\} &\geq \left[1 - j_u \min\{s, d\}\alpha'_f + 1 - (k - j_a - j_u - 1)\min\{s, d-1\}\alpha'_f + 1 - s\alpha'_f - 2\right]\\
&\quad \times [1 - (k - j_u - j_a - 1)\alpha'_c]\\
&\geq \left[1 - j_u \min\{s, d\}\alpha'_f + 1 - (k - j_a - j_u - 1)\min\{s, d\}\alpha'_f + 1 - s\alpha'_f - 2\right]\\
&\quad \times [1 - (k - j_u - j_a - 1)\alpha'_c]\\
&= \left[1 - (k - j_a - 1)\min\{s, d\}\alpha'_f - s\alpha'_f\right] \times [1 - (k - j_u - j_a - 1)\alpha'_c]\\
&\geq \left[1 - (k - 1)\min\{s, d\}\alpha'_f - s\alpha'_f\right] \times [1 - (k - 1)\alpha'_c] = (1 - \alpha_f)(1 - \alpha_c) = 1 - \alpha,
\end{aligned}
$$

where the third inequality holds since the lower bound of $(k - j_a - 1)\min\{s, d\}\alpha'_f$ is achieved when $j_a = 0$, and the lower bound of $1 - (k - j_u - j_a - 1)\alpha'_c$ is achieved when $j_a = j_u = 0$.

Case 2: $\theta^* = d + 1$.

If $\theta^* = d+1$, there are no desirable systems for any threshold vector. Based on the definition of CS, CS is to either declare all systems are infeasible or to select an acceptable system with respect to any of the threshold vectors $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(d)}$. Therefore, CS is ensured by correctly concluding feasibility decisions for all system $i \in S_u$. Then $\Pr(\mathrm{CS}) \geq \Pr(\mathcal{A}_1^*)$ and Lemma 3 and the Bonferroni inequality yield

$$
\begin{aligned}
\Pr\{\mathrm{CS}\} &\geq \begin{cases} (1 - \min\{s, d\}\alpha'_f)^{j_u}. & \text{if systems are simulated independently,}\\ 1 - j_u \min\{s, d\}\alpha'_f, & \text{if systems are simulated under CRN} \end{cases}\\
&\geq \begin{cases} (1 - \min\{s, d\}\alpha'_f)^{k}, & \text{if systems are simulated independently,}\\ 1 - k \min\{s, d\}\alpha'_f, & \text{if systems are simulated under CRN,} \end{cases}
\end{aligned}
$$

where the last inequality is due to the fact that $1 \leq j_u \leq k$ and $0 \leq \min\{s,d\}\alpha'_f \leq 1$. When systems are simulated independently, we have

$$\begin{aligned}\Pr\{\text{CS}\} &\geq (1-\min\{s,d\}\alpha'_f)^k \geq (1-\min\{s,d\}\alpha'_f)^{k-1}\cdot(1-s\alpha'_f)\\ &= 1-\alpha_f > 1-\alpha.\end{aligned}$$

When systems are simulated under CRN, we have

$$\begin{aligned}\Pr\{\text{CS}\} &\geq 1-k\min\{s,d\}\alpha'_f \geq 1-(k-1)\min\{s,d\}\alpha'_f - s\alpha'_f\\ &= 1-\alpha_f > 1-\alpha.\end{aligned}$$

□

The choices of α_f and α_c affect the performance of the $\mathcal{ZAK}^{\mathcal{R}}$ procedure. If Phase I is difficult (e.g., the secondary performance measures of many systems are close to some of the threshold values in threshold vectors $\boldsymbol{q}^{(1)},\ldots,\boldsymbol{q}^{(\theta^*)}$), one may want to choose a larger value for α_f than α_c to improve the efficiency. On the other hand, if Phase I is relatively easy compared with Phase II, then it is more efficient to assign a larger value of α_c than α_f. If the decision maker has knowledge on the relative difficulty of the feasibility checks and the comparison, she may first decide the choice of $e = \alpha_f/\alpha_c$, the ratio of the nominal error of Phase I to Phase II. One may then find α_c by solving

$$(1-e\times\alpha_c)(1-\alpha_c) = 1-\alpha,$$

and find the corresponding $\alpha_f = e\times\alpha_c$.

However, the decision maker usually does not have the information about the mean configurations of the primary and secondary performance measures of the systems. One possibility is to select $\alpha_f = \alpha_c = 1-(1-\alpha)^{1/2}$. If $s \leq d$, the formulas for selecting α'_f and α'_c in Algorithm 3 suggest one may first choose $e = s\alpha'_f/\alpha'_c$ (the ratio of the nominal error for feasibility checks across all the constraints for one system and the nominal error for the comparison between best system $[b]$ and one inferior system) and further find α'_f and α'_c depending on the value of e. Similarly, one may consider $e = d\alpha'_f/\alpha'_c$ if $d < s$.

We start with the case when $s \leq d$. When systems are simulated independently, we know that

$$1-\alpha = (1-\alpha_f)(1-\alpha_c) = (1-s\alpha'_f)^k \times (1-\alpha'_c)^{k-1} = (1-e\alpha'_c)^k(1-\alpha'_c)^{k-1},$$

where one can numerically solve for α'_c and $\alpha'_f = e\alpha'_c/s$. When systems are simulated under CRN, we know that

$$1 - \alpha = (1 - \alpha_f)(1 - \alpha_c) = (1 - ks\alpha'_f) \times (1 - (k-1)\alpha'_c) = (1 - ke\alpha'_c) \times (1 - (k-1)\alpha'_c),$$
$$= ek(k-1)(\alpha'_c)^2 - (ek + k - 1)\alpha'_c + 1.$$

As we also know that $0 < \alpha_c < 1$ and $\alpha_c = (k-1)\alpha'_c$, we may solve the above quadratic expression as $\alpha'_c = \frac{ek+k-1-\sqrt{(ek+k-1)^2-4ek(k-1)\alpha}}{2ek(k-1)}$ (the other root does not satisfy $\alpha'_c < \frac{1}{k-1}$).

We then discuss the case when $d < s$. One set $e = d\alpha'_f/\alpha'_c$ and find α'_c by solving

$$\begin{cases} (1 - e\alpha'_c)^{k-1} \times (1 - e\frac{s}{d}\alpha'_c) \times (1 - \alpha'_c)^{k-1} = 1 - \alpha, & \text{if systems are simulated independently;} \\ \left(1 - e\left(k - 1 + \frac{s}{d}\right)\alpha'_c\right) \times [1 - (k-1)\alpha'_c] = 1 - \alpha, & \text{if systems are simulated under CRN.} \end{cases}$$

where the former can be solved numerically and the latter can be found by solving the above quadratic expression as $\alpha'_c = \frac{e(k-1+\frac{s}{d})+k-1-\sqrt{[e(k-1+\frac{s}{d})+k-1]^2-4e(k-1+\frac{s}{d})(k-1)\alpha}}{2e(k-1+\frac{s}{d})(k-1)}$ as the other root does not satisfy $\alpha'_c < \frac{1}{k-1}$.

Note that $e = 1$ is the case when the decision maker sets the nominal error for feasibility checks across all the constraints for one system equals the nominal error for the comparison between best system $[b]$ and one inferior system, while $e = s$ is the case when the decision maker wants the same nominal error for the feasibility check for one system, one constraint, and one threshold value as for the comparison between the best system $[b]$ and one inferior system. Our experimental results (see Section 3.6) consider the cases when $e = 1$ and $e = s$.

3.3.3 Procedure $\mathcal{ZAK}$

In this section, we discuss a variation of the $\mathcal{ZAK}^{\mathcal{R}}$ procedure that collects observations on the primary performance measure in Phase I and recycles them in Phase II.

As $\mathcal{ZAK}^{\mathcal{R}}$ starts "from scratch" when performing the comparison, it discards all the information related to the primary performance measure obtained in Phase I, which can be quite inefficient in terms of the computation effort. One may consider collecting and storing all the observations of the primary performance measure in Phase I and then extracting information related to the primary performance measure when performing Phase

II. However, as Phase I may require a lot of observations, this approach requires significant memory for storing the observations from Phase I.

[16] propose the sequential selection with memory procedure ($\mathcal{SSM}$) that is specifically for use within an optimization-via-simulation algorithm when simulation is costly, and partial or complete information on solutions previously visited is maintained. When data storage is prohibitive, the procedure requires only summary statistics of the simulation output, which solves the memory space issue discussed above. We then present a sequentially-running procedure, namely $\mathcal{ZAK}$, that adopts the $\mathcal{SSM}$ procedure as its Phase II. The detailed description is shown in Algorithm 4.

Algorithm 4 Procedure $\mathcal{ZAK}$

[**Setup**:] Same as in $\mathcal{ZAK}^{\mathcal{R}}$ except for choosing $\alpha_f, \alpha_c > 0$ such that $\alpha_f + \alpha_c = \alpha$.
[**Initialization for Phase I**:] Same as in $\mathcal{ZAK}^{\mathcal{R}}$ except for the following additional steps for each system $i \in M$:

Obtain n_0 observations $X_{in}, n = 1, \ldots, n_0$.

For each system i, compute $\bar{X}_i(n_0)$.

For all systems $i \neq j$, compute $S^2_{X_{ij}}(n_0)$.

[**Feasibility Check**:] Same as in $\mathcal{ZAK}^{\mathcal{R}}$.
[**Stopping Condition for Phase I**:] Same as in $\mathcal{ZAK}^{\mathcal{R}}$ except that we also take one additional observation X_{i,r_i+1}, and update $\bar{X}_i(r_i)$ whenever we take one additional observation Y_{i,r_i+1} from $i \in M$.
[**Initialization for Phase II**:] Same as in $\mathcal{ZAK}^{\mathcal{R}}$ except we do not perform new simulations, do not compute $\bar{X}_i(n_0)$ and $S^2_{X_{ij}}(n_0)$, and set η_c as a solution to $g(\eta_c) = \alpha_c'$, where

$$\alpha_c' = \begin{cases} 1 - (1 - \alpha_c)^{1/(|F|-1)}, & \text{if systems are simulated independently;} \\ \alpha_c/(|F| - 1), & \text{if systems are simulated under CRN.} \end{cases}$$

Set $r = \min_{i \in F} r_i$ and go to [**Comparison**].
[**Comparison**:] For $i, j \in M$ s.t. $i \neq j$ and

$$r\bar{X}_i(r_i) > r\bar{X}_j(r_j) + R(r; \delta, \eta_c, S^2_{X_{ij}}(n_0)),$$

eliminate j from M.
[**Stopping Condition for Phase II**:] If $|M| = 1$, then stop and return the system in M as the best. Otherwise, for each system $i \in M$ with $r_i \leq r$, take one additional observation X_{i,r_i+1}, set $r_i = r_i + 1$ and compute $\bar{X}_i(r_i)$. Then, set $r = r + 1$ and go to [**Comparison**].

Similar to the discussion in [1], there are two difficulties in proving the statistical validity of $\mathcal{ZAK}$. First, as r_i, the number of observations X_{in} collected in Phase I, depends on $Y_{i\ell n}$ for system i, this dependency affects the comparison in Phase II. This dependency issue can be resolved by performing $\mathcal{ZAK}^{\mathcal{R}}$ instead as it restarts “from scratch” for the surviving systems of Phase I. Second, we use $g(\eta_c) = \alpha_c/(|F| - 1)$ instead of $g(\eta_c) = \alpha_c/(k - 1)$ to compute the implementation parameter η_c for Phase II. Thus we only allocate the nominal error for Phase II to the comparison between the best system $[b]$ and the surviving systems from Phase I, rather than all $k - 1$ other systems. As the comparison between $[b]$ and the other surviving systems is done with a larger nominal error, the resulting η_c is smaller, which helps improve the efficiency of our approach. However, the continuation region in Phase II now depends on the number of surviving systems from Phase I. We address the dependency between Phases I and II in $\mathcal{ZAK}$ by choosing the nominal errors α_f and α_c for Phases I and II as $\alpha_f + \alpha_c = \alpha$ to incorporate the correlation between the two phases. While $(1 - \alpha_f)(1 - \alpha_c)$ is always larger than $1 - (\alpha_f + \alpha_c)$, the difference is typically quite small. Although we have not proved the statistical validity of $\mathcal{ZAK}$, our experimental results (discussed in Section 3.6) do not show any violation of its validity.

3.4 Simultaneously-Running Procedure

In this section, we provide a procedure that implements Phases I and II simultaneously. Section 3.4.1 describes the simultaneously-running procedure and Section 3.4.2 proves its statistical validity. Section 3.4.3 discusses how to choose the implementation parameters for the proposed simultaneously-running procedure.

3.4.1 Procedure $\mathcal{ZAK}$+

In this section, we provide a procedure that runs Phases I and II simultaneously in Algorithm 5. Similar to the sequentially-running procedures $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$, we use the variable θ to keep track of the current most preferred threshold vector for which we are trying to determine feasibility. Initially, θ is set to d, which is the index of the least preferred threshold vector. We use sets M and F defined as in Section 4.3.1 and additionally define set SS_i as a set of systems found to be superior to system i in terms of the primary performance measure.

Rather than performing Phase II on the surviving systems from Phase I as $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$ do, we now perform both feasibility check and pairwise comparison for all systems that are still in consideration (i.e., $i \in M$) within each iteration. More specifically, for each system $i \in M$, we check whether there exists a minimum threshold vector that system i is feasible with respect to, use θ to keep track of this threshold index, and update set F if appropriate. When a feasible decision is made for system i, we perform an additional step in Phase I: eliminate system $j \in (M \cup F)$ if $i \in SS_j$ (system $i \in F$ is shown to be superior compared with system j) and system j is not feasible with respect to any of $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta-1)}$. In Phase II, once a system i is declared superior compared with system j in Phase II, we add system i to SS_j. Furthermore, if system $i \in (F \cap SS_j)$ and system j is infeasible with respect to all $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta-1)}$, then we eliminate system j from M and F. A detailed description of the simultaneously-running procedure $\mathcal{ZAK}+$ is shown in Algorithm 5.

3.4.2 Statistical Validity of the Simultaneously Running Procedure

In this section, we present the proof of the statistical validity of the simultaneously-running procedure $\mathcal{ZAK}+$.

Lemma 5. *For a particular system i, the* **[Feasibility Check]** *steps in $\mathcal{ZAK}+$ ensure*

$$\Pr\left(\mathcal{A}_1^*(i)\right) \geq 1 - \min\{s, d\}\beta_f, \ \ \text{if } i \in S_u;$$

$$\Pr\left(\mathcal{A}_2^*(i)\right) \geq 1 - \min\{s, d-1\}\beta_f, \ \ \text{if } i \in S_d \cup S_{a'} \text{ and } 1 < \theta^* \leq d;$$

$$\Pr\left(\mathcal{B}_1^*\right) \geq 1 - s\beta_f, \ \ \text{if } \theta^* \leq d.$$

Lemma 6. *Given i such that $x_i \leq x_{[b]} - \delta$, the* **[Comparison]** *steps in $\mathcal{ZAK}+$ run to completion ensure*

$$\Pr\left(\mathrm{CS}_i\right) \geq 1 - \beta_c$$

The proofs of Lemmas 5 and 6 are essentially same as those of Lemmas 3 and 4 because both α'_f of $\mathcal{ZAK}$ and β_f of $\mathcal{ZAK}+$ are the nominal error of feasibility check for one constraint of one single system with a fixed threshold, and both α'_c of $\mathcal{ZAK}$ and β_c of $\mathcal{ZAK}+$ are the nominal error of comparison between an inferior system and the best system $[b]$.

We are now ready to prove the statistical validity of $\mathcal{ZAK}+$.

Algorithm 5 $\mathcal{ZAK}+$

[**Setup**:] Choose confidence level $1-\alpha$, tolerance levels $\epsilon_1, \ldots \epsilon_s$, indifference-zone parameter δ, threshold vectors $\{\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(d)}\}$, and associated index vectors $\{\boldsymbol{I}^{(1)}, \boldsymbol{I}^{(2)}, \ldots, \boldsymbol{I}^{(d)}\}$. Set $M = \Gamma, SS_i = \emptyset$, and $Z_{i\ell}^m = 2$ for all $i \in M, \ell = 1, \ldots, s$, and $m = 1, \ldots, d_\ell$. Set $F = \emptyset$ and $\theta = d$. Choose $\beta_f, \beta_c > 0$ such that β_f and β_c satisfy

$$\min_{0 \le j \le k-1} \left[(1-\min\{s,d\}\beta_f)^j + (1-\min\{s,d-1\}\beta_f - \beta_c)^{k-j-1}\right] = 2 - \alpha + s\beta_f,$$

if systems are simulated independently; and

$$\max_{0 \le j \le k-1} \left[[j \min\{s,d\} + (k-j-1)\min\{s,d-1\} + s]\,\beta_f - (k-j-1)\beta_c\right] = \alpha,$$

if systems are simulated under CRN.

Set η_f and η_c such that $g(\eta_f) = \beta_f$ and $g(\eta_c) = \beta_c$. Add any constraint ℓ, where $\ell = 1, \ldots, s$, with increasing preference to set IP.

[**Initialization**:]

for each system $i \in M$ **do**

- Obtain n_0 observations from system i.
- Compute $\bar{X}_i(n_0), \bar{Y}_{i\ell}(n_0), S^2_{X_{ij}}(n_0)$, and $S^2_{Y_{i\ell}}(n_0)$ for all $i, j \in M$, where $i \neq j$, and $\ell = 1, \ldots, s$.
- Set $r = n_0, \text{ON}_i = \{1, \ldots, s\}$, and $\text{ON}_{i\ell} = \{1, \ldots, d_\ell\}$ for $\ell = 1, \ldots, s$.

end for

[**Feasibility Check**:]

for $i \in M$ **do**

  for $\ell \in \text{ON}_i$ **do**

    for $m \in \text{ON}_{i\ell}$ **do**

      If $\bar{Y}_{i\ell}(r) + R(r; \epsilon_\ell, \eta_f, S^2_{Y_{i\ell}}(n_0))/r \le q_\ell^m$, set $Z_{i\ell}^m = 1$ and $\text{ON}_{i\ell} = \text{ON}_{i\ell} \setminus \{m\}$;

      If $\bar{Y}_{i\ell}(r) - R(r; \epsilon_\ell, \eta_f, S^2_{Y_{i\ell}}(n_0))/r \ge q_\ell^m$, set $Z_{i\ell}^m = 0$ and $\text{ON}_{i\ell} = \text{ON}_{i\ell} \setminus \{m\}$.

    end for

    If $\text{ON}_{i\ell} = \emptyset$, set $\text{ON}_i = \text{ON}_i \setminus \{\ell\}$.

  end for

  If $\exists$ minimum $\kappa \le \theta$ s.t. $\prod_{\ell=1}^s Z_{i\ell}^{I_\ell^{(\kappa)}} = 1$, and either $\kappa < \theta$ or $i \notin F$, then

  - If $\kappa < \theta$, then set $F = \emptyset, \theta = \kappa$, and for all $j \in M$ delete q_ℓ^m from $\text{ON}_{j\ell}$ if $\ell \in \text{IP}$ and $m > I_\ell^{(\theta)}$ (if $\ell \notin \text{IP}$, then q_ℓ^m can be removed from $\text{ON}_{j\ell}$ if $I_\ell^{(\theta')} \neq m$ for all $\theta' \le \kappa$), and set $\text{ON}_j = \text{ON}_j \setminus \{\ell\}$ if $\text{ON}_{j\ell} = \emptyset$.
  - Add system i to F.
  - For all $j \in M$, if $i \in SS_j$ and either $\theta = 1$ or $\prod_{\ell=1}^s Z_{j\ell}^{I_\ell^{(\kappa)}} = 0$ for all $\kappa = 1, \ldots, \theta - 1$, then remove system j from M and F (if $j \in F$) and delete SS_j.

  If either $\prod_{\ell=1}^s Z_{i\ell}^{I_\ell^{(\kappa)}} = 0$ for all $1 \le \kappa \le \theta$, or $\theta > 1, \prod_{\ell=1}^s Z_{i\ell}^{I_\ell^{(\kappa)}} = 0$ for all $1 \le \kappa \le \theta - 1$, and there exists $j \in F \cap SS_i$, then remove i from M and delete SS_i.

end for

[**Comparison**:] For $i, j \in M$ s.t. $i \neq j, i \notin SS_j, j \notin SS_i$, and

$$r\bar{X}_i(r) > r\bar{X}_j(r) + R(r; \delta, \eta_c, S^2_{X_{ij}}(n_0)),$$

add system i to SS_j. If $i \in F$, then remove system j from M and F (if $j \in F$) if either $\theta = 1$ or $\prod_{\ell=1}^s Z_{j\ell}^{I_\ell^{(\kappa)}} = 0$ for all $\kappa = 1, \ldots, \theta - 1$, and delete SS_j.

[**Stopping Condition**:] If $M = F$ and $|F| = 1$, then stop and return the system in F as the best system. Else if $M = F$ and $|F| = 0$, then stop and return no feasible systems exist. Otherwise, for all $i \in M$, set $r = r + 1$, take one additional observation, update $\bar{X}_i(r)$ and $\bar{Y}_{i\ell}(r)$ for all $\ell \in \text{ON}_i$, and go to [**Feasibility Check**].

Theorem 5. *Under Assumptions 2 and 3, the $\mathcal{ZAK}+$ procedure guarantees*

$$\Pr\{\text{CS}\} \geq 1 - \alpha.$$

Proof. We consider two cases, namely when $\theta^* \leq d$ and $\theta^* = d + 1$.

Case 1: $\theta^* \leq d$.

Similar to the proof of statistical validity of $\mathcal{ZAK}^{\mathcal{R}}$, systems in $(S_{a'} \cup S_d)$ should not be declared feasible with respect to a more preferred threshold vector $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta^*-1)}$. We consider the events $\mathcal{A}_1^*, \mathcal{A}_2^*, \mathcal{B}_1^*$, and $\mathcal{B}_2^*$ defined in Section 4.3.2. As $\mathcal{ZAK}+$ performs Phases I and II simultaneously, events $\mathcal{A}_1^*, \mathcal{A}_2^*, \mathcal{B}_1^*$, and $\mathcal{B}_2^*$ are dependent. We then have

$$\Pr\{\text{CS}\} \geq \Pr\{\mathcal{A}_1^* \cap \mathcal{A}_2^* \cap \mathcal{B}_1^* \cap \mathcal{B}_2^*\} \geq \Pr(\mathcal{A}_1^*) + \Pr(\mathcal{A}_2^* \cap \mathcal{B}_2^*) + \Pr(\mathcal{B}_1^*) - 2.$$

Notice that $\mathcal{A}_2^* \cap \mathcal{B}_2^*$ is the event that all systems in $S_{a'} \cup S_d$ are declared infeasible to threshold vectors $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta^*-1)}$ and are eliminated by comparison with system $[b]$, i.e., $\mathcal{A}_2^* \cap \mathcal{B}_2^* = \cap_{i \in S_d \cup S_{a'}} \mathcal{A}_2^*(i) \cap \text{CS}_i$. Similarly, $\mathcal{A}_1^* = \cap_{i \in S_u} \mathcal{A}_1^*(i)$.

We discuss the cases depending on whether systems are simulated independently or under CRN. When systems are simulated independently, by Lemma 5, we have

$$\Pr(\mathcal{A}_1^*) \geq (1 - \min\{s, d\}\beta_f)^{j_u};$$
$$\Pr(\mathcal{B}_1^*) \geq 1 - s\beta_f.$$

We use the same notation N_{ij} from the proof of Theorem 4 and have

$$\begin{aligned}
\Pr(\mathcal{A}_2^* \cap \mathcal{B}_2^*) &= \Pr\left(\cap_{i \in (S_d \cup S_{a'})} (\mathcal{A}_2^*(i) \cap \text{CS}_i)\right) \\
&= \text{E}\left[\Pr\left\{\cap_{i \in (S_d \cup S_{a'})} (\mathcal{A}_2^*(i) \cap \text{CS}_i) \middle| X_{[b]1}, \ldots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right] \\
&= \text{E}\left[\prod_{i \in (S_d \cup S_{a'})} \Pr\left\{\mathcal{A}_2^*(i) \cap \text{CS}_i \middle| X_{[b]1}, \ldots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right] \\
&\geq \prod_{i \in (S_d \cup S_{a'})} \text{E}\left[\Pr\left\{\mathcal{A}_2^*(i) \cap \text{CS}_i \middle| X_{[b]1}, \ldots, X_{[b],N_{[b]}}, S^2_{X_{i[b]}}(n_0)\right\}\right] \\
&\geq \prod_{i \in (S_d \cup S_{a'})} \left[1 - \text{E}\left[\Pr\left\{(\mathcal{A}_2^*(i))^c \middle| S_{X^2_{i[b]}}(n_0)\right\}\right] - \text{E}\left[\Pr\left\{\text{ICS}_i \middle| S_{X^2_{i[b]}}(n_0)\right\}\right]\right] \\
&= \prod_{i \in (S_d \cup S_{a'})} \left[1 - \Pr\{(\mathcal{A}_2^*(i))^c\} - \Pr\{\text{ICS}_i\}\right]
\end{aligned}$$

$$\geq \prod_{i\in(S_d\cup S_{a'})} (1-\min\{s,d-1\}\beta_f-\beta_c) = (1-\min\{s,d-1\}\beta_f-\beta_c)^{j_d+j_{a'}}$$

$$= (1-\min\{s,d-1\}\beta_f-\beta_c)^{k-j_a-j_u-1},$$

where we use A^c to denote the complement event of A. The first inequality is from Lemma 2.4 of [17], the second inequality holds due to the Bonferroni inequality, and the last inequality is from Lemmas 5 and 6.

Thus, we know that

$$\begin{aligned}\Pr\{\text{CS}\} &\geq (1-\min\{s,d\}\beta_f)^{j_u} + (1-\min\{s,d-1\}\beta_f-\beta_c)^{k-j_a-j_u-1} + (1-s\beta_f) - 2\\ &\geq (1-\min\{s,d\}\beta_f)^{j_u} + (1-\min\{s,d-1\}\beta_f-\beta_c)^{k-j_u-1} - s\beta_f - 1,\end{aligned}$$

where the second inequality holds since the lower bound of $(1-\min\{s,d-1\}\beta_f-\beta_c)^{k-j_a-j_u-1}$ is achieved when $j_a = 0$ as $0 \leq 1-\min\{s,d-1\}\beta_f-\beta_c < 1$. As $0 \leq j_u \leq k-1$ (because $\theta^* \leq d$), we know that

$$\Pr\{\text{CS}\} \geq \min_{0\leq j\leq k-1}\left[(1-\min\{s,d\}\beta_f)^j + (1-\min\{s,d-1\}\beta_f-\beta_c)^{k-j-1} - s\beta_f - 1\right] = 1-\alpha.$$

When systems are simulated under CRN, by Lemmas 5 and 6, and the Bonferroni inequality, we have

$$\begin{aligned}\Pr(\mathcal{A}_1^*) &\geq 1 - j_u\min\{s,d\}\beta_f;\\ \Pr(\mathcal{B}_1^*) &\geq 1 - s\beta_f;\\ \Pr(\mathcal{A}_2^*\cap\mathcal{B}_2^*) &= \Pr\left(\cap_{i\in(S_d\cup S_{a'})}\left(\mathcal{A}_2^*(i)\cap \text{CS}_i\right)\right)\\ &\geq 1 - \sum_{i\in(S_d\cup S_{a'})}\left[\Pr\left(\mathcal{A}_2^*(i)\right)^c + \Pr(\text{ICS}_i)\right]\\ &\geq 1 - \sum_{i\in(S_d\cup S_{a'})}\left[\min\{s,d-1\}\beta_f+\beta_c\right]\\ &= 1 - (j_d+j_{a'})\left[\min\{s,d-1\}\beta_f+\beta_c\right]\\ &= 1 - (k-j_a-j_u-1)\left[\min\{s,d-1\}\beta_f+\beta_c\right],\end{aligned}$$

where the first inequality holds due to the Bonferroni inequality and the second inequality holds by Lemmas 5 and 6.

Thus, we know that

$$\begin{aligned}\Pr\{\text{CS}\} &\geq 1 - j_u \min\{s,d\}\beta_f + 1 - (k - j_a - j_u - 1)\left[\min\{s,d-1\}\beta_f + \beta_c\right] + 1 - s\beta_f - 2 \\ &\geq 1 - j_u \min\{s,d\}\beta_f - (k - j_u - 1)\left[\min\{s,d-1\}\beta_f + \beta_c\right] - s\beta_f, \\ &= 1 - \left[j_u \min\{s,d\} + (k - j_u - 1)\min\{s,d-1\} + s\right]\beta_f - (k - j_u - 1)\beta_c,\end{aligned}$$

where the second inequality holds since the lower bound of $1-(k-j_a-j_u-1)\left[\min\{s,d-1\}\beta_f + \beta_c\right]$ is achieved when $j_a = 0$. As $0 \leq j_u \leq k-1$, we know that

$$\Pr\{\text{CS}\} \geq \min_{0\leq j\leq k-1}\left[1 - \left[j\min\{s,d\} + (k-j-1)\min\{s,d-1\} + s\right]\beta_f - (k-j-1)\beta_c\right] = 1-\alpha.$$

Case 2: $\theta^* = d+1$.

If $\theta^* = d+1$, there are no desirable systems for any threshold vector. Similar to the discussion in the proof of Theorem 4, CS is ensured by correctly concluding feasibility decisions for all systems $i \in S_u$. Then $\Pr\{\text{CS}\} \geq \Pr(\mathcal{A}_1^*)$ and Lemma 5 and the Bonferroni inequality yield

$$\begin{aligned}\Pr\{\text{CS}\} &\geq \begin{cases}(1-\min\{s,d\}\beta_f)^{j_u}, & \text{if systems are simulated independently,}\\ 1 - j_u\min\{s,d\}\beta_f, & \text{if systems are simulated under CRN.}\end{cases}\\ &\geq \begin{cases}(1-\min\{s,d\}\beta_f)^{k}, & \text{if systems are simulated independently,}\\ 1 - k\min\{s,d\}\beta_f, & \text{if systems are simulated under CRN,}\end{cases}\end{aligned}$$

where the last inequality is due to the fact that $1 \leq j_u \leq k$ and $0 \leq \min\{s,d\}\beta_f \leq 1$. When systems are simulated independently, we have

$$\begin{aligned}\Pr\{\text{CS}\} &\geq (1-\min\{s,d\}\beta_f)^k \geq (1-\min\{s,d\}\beta_f)^k + \beta_f\left[(1-\min\{s,d\}\beta_f)^{k-1}\times\min\{s,d\} - s\right]\\ &= (1-\min\{s,d\}\beta_f)^{k-1}\left(1-\min\{s,d\}\beta_f + \min\{s,d\}\beta_f\right) - s\beta_f\\ &= (1-\min\{s,d\}\beta_f)^{k-1} - s\beta_f \geq (1-\min\{s,d\}\beta_f - \beta_c)^{k-1} - s\beta_f\\ &= (1-\min\{s,d\}\beta_f)^{0} + (1-\min\{s,d\}\beta_f - \beta_c)^{k-0-1} - s\beta_f - 1\\ &\geq \min_{0\leq j\leq k-1}(1-\min\{s,d\}\beta_f)^{j} + (1-\min\{s,d\}\beta_f - \beta_c)^{k-j-1} - s\beta_f - 1\\ &= 1-\alpha,\end{aligned}$$

where the second inequality holds as $\beta_f \left[(1 - \min\{s,d\}\beta_f)^{k-1} \cdot \min\{s,d\} - s\right] \leq 0$ since $(1 - \min\{s,d\}\beta_f)^{k-1} \leq 1$ and $\min\{s,d\} \leq s$, and the third equality holds because $(1 - \min\{s,d\}\beta_f)^0 = 1$.

When systems are simulated under CRN, we have

$$\begin{aligned}
\Pr\{\text{CS}\} &\geq 1 - k\min\{s,d\}\beta_f \geq 1 - (k-1)\min\{s,d\}\beta_f - s\beta_f \geq 1 - [(k-1)\min\{s,d\} + s]\beta_f \\
&= 1 - [(k-1)\min\{s,d\} + (k-(k-1)-1)\min\{s,d-1\} + s]\,\beta_f - (k-(k-1)-1)\beta_c \\
&\geq \min_{0\leq j\leq k-1} [1 - [j\min\{s,d\} + (k-j-1)\min\{s,d-1\} + s]\,\beta_f - (k-j-1)\beta_c] \\
&= 1-\alpha. \qquad \square
\end{aligned}$$

3.4.3 Implementation Parameters for Simultaneous Running Procedure

In this section, we discuss how to choose implementation parameters β_f and β_c in simultaneous-running procedure $\mathcal{ZAK}+$.

We start by considering the case when $s < d$, and the systems are simulated independently. In this case, we need to find β_f and β_c such that

$$\min_{0\leq j\leq k-1} \left[(1 - s\beta_f)^j + (1 - s\beta_f - \beta_c)^{k-j-1}\right] - s\beta_f = 2 - \alpha.$$

One approach is to first decide the choice of $e = s\beta_f/\beta_c$. Recall that this is the ratio of (i) the error for a feasibility check of one system for all constraints and all thresholds to (ii) the error of a comparison between two systems. The ratio should be decided based on the decision maker's idea on whether she wants to allocate more error to feasibility check or comparison.

Let $\beta = s\beta_f = e\beta_c$. Then we have

$$\Pr\{\text{CS}\} \geq (1-\beta)^j + (1 - (1+1/e)\beta)^{k-j-1} - \beta - 1.$$

Let $f(j)$ be a function of j such that $f(j) = (1-\beta)^j + (1-(1+1/e)\beta)^{k-j-1} - \beta - 1$. We need to find the lower bound of $f(j)$ given that $0 \leq j \leq k-1$. The first derivative of $f(j)$ is

$$\frac{\partial}{\partial j} f(j) = (1-\beta)^j \log(1-\beta) - (1-(1+1/e)\beta)^{k-j-1}\log(1-(1+1/e)\beta).$$

By setting $\frac{\partial}{\partial j} f(j) = 0$, we have

$$j^* = \frac{\log C + (k-1)\log(1-(1+1/e)\beta)}{\log(1-(1+1/e)\beta) + \log(1-\beta)}, \text{ where } C = \frac{\log(1-(1+1/e)\beta)}{\log(1-\beta)}. \quad (8)$$

By checking the second derivative of $f(j)$, we have

$$\frac{\partial^2}{\partial j^2} f(j) = [\log(1-\beta)]^2(1-\beta)^j + [\log(1-(1+1/e)\beta)]^2(1-(1+1/e)\beta)^{k-j-1} \geq 0,$$

which shows that j^* in (8) indeed gives us the lower bound of $f(j)$.

To compute the value of β, we solve the following equation for β:

$$(1-\beta)^{j^*} + (1-(1+1/e)\beta)^{k-j^*-1} + (1-\beta) - 2 = 1-\alpha,$$

where j^* is defined in (8). The resulting β is the value of $e\beta_c$ and $s\beta_f$.

We then consider the case when systems are simulated under CRN. We need to find α_f and α_c such that

$$\min_{0\leq j\leq k-1} [1 - ks\beta_f - (k-j-1)\beta_c] = 1 - ks\beta_f - (k-1)\beta_c = 1-\alpha.$$

By setting $\beta = s\beta_f = e\beta_c$, we have $\alpha = \left[k + \frac{k-1}{e}\right]\beta$ and, therefore, the value of $e\beta_c$ and $s\beta_f$ can be found as $s\beta_f = e\beta_c = \alpha / \left[k + \frac{k-1}{e}\right]$.

[5] choose $e = 1$ to balance the errors assigned to feasibility check and comparison. We use $e = 1$ and $e = s$ for the experimental results (discussed in Section 3.6) to demonstrate the performance of our proposed procedure.

When $s \geq d$, setting implemetation parameters β_f and β_c follows a similar approach as above. We therefore omit the discussion for the sake of space.

3.5 Different Preference Orders of Input Threshold Vectors

As discussed in Section 2.1, our procedures $\mathcal{ZAK}^{\mathcal{R}}$, $\mathcal{ZAK}$, and $\mathcal{ZAK}+$ require lists of threshold vectors $\{\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(d)}\}$ and index vectors $\{\boldsymbol{I}^{(1)}, \boldsymbol{I}^{(2)}, \ldots, \boldsymbol{I}^{(d)}\}$. Rather than having to input all the threshold and index vectors, it would be easier for the decision maker to give threshold values on each constraint and how she would like to prioritize each constraint. In this section, we discuss three preference orders for formulating the input threshold vectors, namely ranked constraints, equally important constraints, and total violation with ranked constraints. The experimental results for multiple constraints shown in Section 3.6 are based on these three preference orders.

Ranked constraints: The constraints are ranked with respect to their importance and the decision maker wants to relax the least important constraint first while keeping the rest of the constraints fixed at the current threshold values, and then move to the second least important constraint, etc. Figure 9(a) shows A_θ for $\theta = 1, \ldots, 9$ when $s = 2$ and $d_1 = d_2 = 3$, the secondary performance measures are non-negative, and constraint 1 is more important than constraint 2. The inventory example discussed in Sections 2.1 and 2.2 also has ranked constraints with constraint 1 being more important than constraint 2. Algorithm 6 constructs the threshold vectors for ranked constraints.

Algorithm 6 Constructing threshold vectors for ranked constraints

---

Input q_ℓ^m for all $\ell = 1, \ldots, s$ and $m = 1, \ldots, d_\ell$. Let $\boldsymbol{Q}$ be an empty list of threshold vectors and let `threshold` be a vector of length s.
for $m_1 = 1, \ldots, d_1$ **do**
    for $m_2 = 1, \ldots, d_2$ **do**
        ...
        for $m_s = 1, \ldots, d_s$ **do**
            for $\ell = 1, \ldots, s$ **do**
                Set `threshold`$[\ell] = q_\ell^{m_\ell}$.
            end for
            Add `threshold` to $\boldsymbol{Q}$.
        end for
    end for
end for
return $\boldsymbol{Q}$

Equally important constraints: All constraints are equally important and the decision maker wants to relax all constraints by one threshold at the same time. If the constraints do not all have the same number of thresholds, when the threshold vectors are formulated, then constraints that have gone through all their thresholds keep the "loosest" threshold (i.e., $q_\ell^{d_\ell}$ for constraint ℓ) while the other constraints relax. Figure 9(b) shows this case when there are two constraints and three thresholds on each constraint. Algorithm 7 formulates the threshold vectors for this option.

Algorithm 7 Constructing threshold vectors for equally important constraints

Input q_ℓ^m for all $\ell = 1, \ldots, s$ and $m = 1, \ldots, d_\ell$. Let $\boldsymbol{Q}$ be an empty list of threshold vectors and let `threshold` be a vector of length s. Set $L = \max_{\ell=1,\ldots,s} d_\ell$.
for $m = 1, \ldots, L$ **do**
  for $\ell = 1, \ldots, s$ **do**
    if $m \leq d_\ell$ **then**
      Set `threshold`$[\ell] = q_\ell^m$.
    else
      Set `threshold`$[\ell] = q_\ell^{d_\ell}$.
    end if
  end for
  Add `threshold` to $\boldsymbol{Q}$.
end for
return $\boldsymbol{Q}$

Total violation with ranked constraints: The decision maker wants to minimize the number of total violations on ranked constraints. For constraint ℓ with threshold q_ℓ^m, its violation is defined as $m - 1$. Then the total violation is defined as the sum of the violations for all constraints. The decision maker always prefers threshold vectors that have fewer total violations, and among threshold vectors that have the same total violation, her preference order is based the priority of the constraints. In Figure 9(c), constraint 1 more important than constraint 2. Region A_1 is defined with respect to (q_1^1, q_2^1) and has total violation 0. Regions A_2 and A_3 are defined with respect to (q_1^1, q_2^2) and (q_1^2, q_2^1), respectively, and have total violation 1, with A_2 preferred to A_3 due to the ranking of constraints 1 and 2. In this preference order, we start with a threshold vector with total violation equal to 0 and then relax the total violation by relaxing the less importants constraint first. The largest total violation is $\sum_{\ell=1}^{s}(d_\ell - 1)$. Algorithm 8 formulates threshold vectors according to this option.

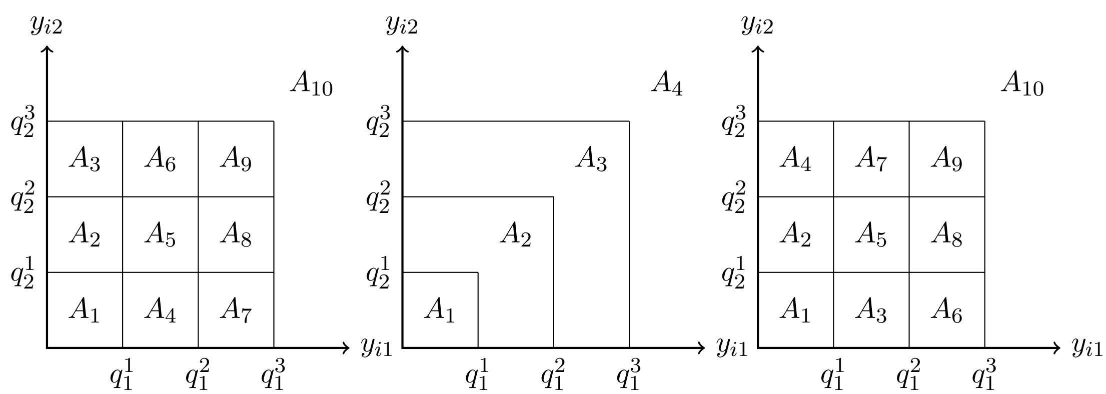
3.6 Experimental Results

In this section, we present experimental results to demonstrate the performances of our proposed procedures $\mathcal{ZAK}^{\mathcal{R}}, \mathcal{ZAK}$, and $\mathcal{ZAK}+$. We compare the performance of each proposed procedure with an alternative procedure that iteratively applies sequential or

Algorithm 8 Constructing threshold vectors for total violation with ranked constraints

Input q_ℓ^m for all $\ell = 1,\dots,s$ and $m = 1,\dots,d_\ell$. Let $\boldsymbol{Q}$ be an empty list of threshold vectors and let `threshold` be a vector of length s.

```
for v = 0, ..., Σ_{ℓ=1}^{s} (d_ℓ − 1) do
    for v_1 = 0, ..., v do
        for v_2 = 0, ..., v − v_1 do
            for v_3 = 0, ..., v − (v_1 + v_2) do
                ...
                for v_s = 0, ..., v − Σ_{ℓ'=1}^{s−1} v_ℓ' do
                    for ℓ = 1, ..., s do
                        Set threshold[ℓ] = q_ℓ^{v_ℓ+1}.
                    end for
                end for
            end for
            Add threshold to Q.
        end for
    end for
end for
return Q
```



(a) Ranked constraints (b) Equally important constraints (c) Total violation with ranked constraints

Figure 9: Three preference orders

simultaneous procedures to threshold vector $\boldsymbol{q}^{(1)},\dots,\boldsymbol{q}^{(d)}$. If a single constraint is considered, we use $\mathcal{AK}$ or $\mathcal{AK}+$ due to [1] for each threshold value. If multiple constraints are considered, we use $\mathcal{HAK}$ or $\mathcal{HAK}+$ due to [5] for each threshold vector. We name the procedures that iteratively implement $\mathcal{AK}$ and $\mathcal{AK}+$ as Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$, respectively. Similarly, we name the procedures that iteratively implement $\mathcal{HAK}$ and $\mathcal{HAK}+$ as Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$, respectively. Notice that Restart$^{\mathcal{AK}}$ (Restart$^{\mathcal{AK}+}$) is a special case of Restart$^{\mathcal{HAK}}$ (Restart$^{\mathcal{HAK}+}$) when there is a single constraint. We provide the algorithm statements and discussions of the statistical validity of procedures Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ in Appendices B.1 and B.2, respectively.

All the experimental results are based on 10,000 macro replications with $\alpha = 0.05$ and $n_0 = 20$ and we report average numbers of observations (OBS) and observed probability of correct selection (PCS). We set $\delta = \epsilon_\ell = 1/\sqrt{n_0}$, where $\ell = 1, \ldots, s$. We discuss the experimental configurations in Section 3.6.1 and show the experimental results for a single constraint in Section 3.6.2. Section 3.6.3 discusses how the number of unacceptable systems with respect to $\boldsymbol{q}^{(\theta^*)}$ effects the value of PCS of $\mathcal{ZAK}$+. Experimental results for two (four) constraints are provided in Section 3.6.4 (3.6.5). Section 3.6.6 discusses a case when all systems are infeasible. Experimental results for the inventory example discussed in Sections 3.1 and 3.2 are provided in Section 3.6.7.

3.6.1 Experimental Configurations

In this section, we discuss the mean and variance configurations for primary and secondary performance measures. We consider two mean configurations of systems, namely difficult means (DM) and monotone increasing means (MIM). Let b be the number of systems that are desirable with respect to threshold vector $\boldsymbol{q}^{(\theta^*)}$. As the existence of acceptable systems will not lower the PCS (because declaring acceptable systems feasible or infeasible with respect to a specific threshold value are both considered as correct feasibility decisions) and as [1] show by experiments that the presence of acceptable systems does not significantly affect the overall performance of procedures $\mathcal{AK}$ and $\mathcal{AK}$+, we do not include acceptable systems in our two configurations.

The DM configuration sets the mean of all secondary performance measures to the boundary of the desirable region of $\boldsymbol{q}^{(\theta^*)}$ for b systems (i.e., the mean of the secondary performances measure ℓ for b systems is $q_\ell^{(\theta^*)} - \epsilon_\ell$) and to the boundary of the unacceptable region of $\boldsymbol{q}^{(\theta^*)}$ for the other $(k-b)$ systems (i.e., the mean of the secondary performances measure ℓ for $(k-b)$ systems is $q_\ell^{(\theta^*)} + \epsilon_\ell$) , which makes the feasibility check difficult. Moreover, the DM configuration has one system whose mean performance of the primary performance is δ, the other systems that are feasible with respect to $\boldsymbol{q}^{(\theta^*)}$ have primary performances equal to 0, and all infeasible systems with respect to $\boldsymbol{q}^{(\theta^*)}$ have 2δ as their primary performance measures. This means that all the infeasible systems are superior compared with the best system while all other feasible systems are only δ inferior compared with the best system, which makes the comparison also difficult. More specifically, in the DM configuration,

$$x_i = \mathrm{E}\left[X_{in}\right] = \begin{cases} 0, & i = 1, 2, \ldots, b-1, \\ \delta, & i = b, \\ 2\delta, & i = b+1, \ldots, k, \end{cases}$$

$$y_{i\ell} = \mathrm{E}[Y_{i\ell n}] = \begin{cases} q_\ell^{(\theta^*)} - \epsilon_\ell, & i = 1, 2, \ldots, b, \\ q_\ell^{(\theta^*)} + \epsilon_\ell, & i = b+1, \ldots, k. \end{cases}$$

In the DM configuration, we consider the case when the decision maker prefers threshold $q_\ell^1 = 0$ for constraint ℓ and would relax the constraint threshold by $2\epsilon_\ell$ every time when she wants to consider a "looser" threshold value on that constraint. In other words, we choose thresholds $\{0, 2\epsilon_\ell\}$ and $\{0, 2\epsilon_\ell, 4\epsilon_\ell, 6\epsilon_\ell\}$ on constraint ℓ when there are two or four thresholds are in consideration, respectively.

The MIM configuration sets the mean of all secondary performance measures to the boundary of the desirable region of $\boldsymbol{q}^{(\theta^*)}$ for b systems and the other $(k-b)$ systems are evenly distributed over $A_{(\theta^*+1)}, \ldots, A_{(d+1)}$ with respect to their secondary performance measures. The mean of the primary performance is monotonically increasing from 0 with an increment of δ. More specifically, in the MIM configuration, $x_i = \mathrm{E}\left[X_{in}\right] = (i-1)\delta, i = 1, \ldots, k$, and

$$y_{i\ell} = \mathrm{E}[Y_{i\ell n}] = \begin{cases} q_\ell^{(\theta^*)} - 2\epsilon_\ell, & i = 1, 2, \ldots, b, \\ q_\ell^{(\theta^*+1)} - 2\epsilon_\ell, & i = b+1, \ldots, \lceil b + \frac{k-b}{d+1-\theta^*} \rceil, \\ q_\ell^{(\theta^*+2)} - 2\epsilon_\ell, & i = \lceil b + \frac{k-b}{d+1-\theta^*} \rceil + 1, \ldots, \lceil b + 2\frac{k-b}{d+1-\theta^*} \rceil, \\ \cdots & \\ q_\ell^{(d)} - 2\epsilon_\ell, & i = \lceil b + (d-\theta^*-1)\frac{k-b}{d+1-\theta^*} \rceil + 1, \ldots, \lceil b + (d-\theta^*)\frac{k-b}{d+1-\theta^*} \rceil, \\ q_\ell^{(d)} + 2\epsilon_\ell, & i = \lceil b + (d-\theta^*)\frac{k-b}{d+1-\theta^*} \rceil + 1, \ldots, k. \end{cases}$$

In the MIM configuration, we consider the case when the decision maker prefers $q_\ell^1 = 0$ and would like to relax by $4\epsilon_\ell$ when she wants to consider "looser" threshold values. That is, the distance between two successive thresholds on constraint ℓ equals $4\epsilon_\ell$.

We consider three variance configurations to test different levels of relative difficulty of the feasibility check and the comparison. We use $\sigma_{x_i}^2$ to denote the variance of the primary performance from system i, $\sigma_{y_{i\ell}}^2$ to denote the variance of the secondary performance ℓ from system i, and consider both low variance (L) and high variance (H). When the difficulty between feasibility checks and comparison are similar, we set $\sigma_{x_i}^2 = 1$ and $\sigma_{y_{i\ell}}^2 = 1$ (L/L); when the comparison is relatively more difficult compared with the feasibility checks, we set $\sigma_{x_i}^2 = 1$ and $\sigma_{y_{i\ell}}^2 = 5$ (L/H); and when the feasibility checks are relatively more difficult compared with comparison, we set $\sigma_{x_i}^2 = 5$ and $\sigma_{y_{i\ell}}^2 = 1$ (H/L).

Due to the discussion in [1], the correlation between the primary and secondary performance measures does not have a significant impact on the experimental results. [5] and the experimental results in Chapter 2 also show that the correlation between secondary performance measures do not show siginificant impact on the result. Therefore, we assume the observations from each system are standard normal random vectors. With 10,000 macro replications, the first four digits of the OBS showed in the tables are meaningful, and the PCS are meaningful up to the 0.001th digit.

3.6.2 Single Constraint

In this section, we consider a single constraint with two and four thresholds. We consider $b \in \{25, 50, 75\}$ under all three variance configurations and choose $k = 100$. As discussed in Section 4.3.2, we introduced two approaches of setting the implementation parameters, namely setting $\alpha_f = \alpha_c = 1 - (1-\alpha)^{1/2}$ and setting $s\alpha'_f = \alpha'_c$. We let $\mathcal{ZAK}_1^{\mathcal{R}}$ to denote the procedure $\mathcal{ZAK}^R$ that follows $\alpha_f = \alpha_c = 1 - (1-\alpha)^{1/2}$ and let $\mathcal{ZAK}_2^{\mathcal{R}}$ to denote the procedure that follows $s\alpha'_f = \alpha'_c$.

We first consider the case of two thresholds under the DM configuration (so that $q_1^1 = 0$ and $q_1^2 = 2\epsilon_1$). Table 13 shows the result for $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$, and Table 14 provides comparisons between $\mathcal{ZAK}^{\mathcal{R}}$, $\mathcal{ZAK}$, and Restart$^{\mathcal{AK}}$ and between $\mathcal{ZAK}+$ and Restart$^{\mathcal{AK}+}$, respectively, for $\theta^* \in \{1, 2\}$. We see that the performance of $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$ are very similar and their performance is worse compared with that of $\mathcal{ZAK}$ in terms of the required number of observations. This is expected as $\mathcal{ZAK}^{\mathcal{R}}$ restarts in Phase II and also set implementation parameters in a conservative way.

We see that $\mathcal{ZAK}$ outperforms Restart$^{\mathcal{AK}}$ under all cases and $\mathcal{ZAK}+$ outperforms Restart$^{\mathcal{AK}+}$ in most cases. If there exist at least one feasible system with respect to $\boldsymbol{q}^{(1)}$ (i.e., $\theta^* = 1$), then it is likely that $\mathcal{AK}$ and $\mathcal{AK}+$ are executed only once. As Restart$^{\mathcal{AK}}$ allocates the nominal error for the two thresholds, the resulting continuation regions used for feasibility check and for comparison are larger than those of $\mathcal{ZAK}$, and therefore its performance is worse than that of $\mathcal{ZAK}$. For single execution, Restart$^{\mathcal{AK}+}$ has a smaller continuation region than that of $\mathcal{ZAK}+$ as Restart$^{\mathcal{AK}+}$ does not need to allocate nominal error in the feasibility check to avoid the inferior systems being declared feasible to a preferred threshold vector (since Restart$^{\mathcal{AK}+}$ only considers one threshold vector). As Restart$^{\mathcal{AK}+}$ also allocates the nominal error for the two thresholds, the difference between the continuation regions used for the feasibility checks and the comparison are similar compared with that of $\mathcal{ZAK}+$, and therefore, $\mathcal{ZAK}+$ and Restart$^{\mathcal{AK}+}$ have similar overall performance when $\theta^* = 1$. Notice that $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$ has similar continuation region used for feasibility check compared with that of $\mathcal{ZAK}$ but larger continuation region used for comparison compared with that of $\mathcal{ZAK}$. For single execution, $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$ has smaller continuation region compared with that of Restart$^{\mathcal{AK}+}$ for the feasibility check but much larger continuation region for the comparison. This is due to the fact that Restart$^{\mathcal{AK}+}$ allocates nominal error for two thresholds (so that the continuation region for the feasibility check is larger) and does not restart while also set parameter less conservative for the comparison (so that the continuation region for the comparison is smaller). Therefore, it is expected that the performance of $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$ is worse compared with that of Restart$^{\mathcal{AK}}$.

When there are no feasible systems with respect to $\boldsymbol{q}^{(1)}$ (i.e., $\theta^* \geq 2$), $\mathcal{AK}$ and $\mathcal{AK}+$ need to be implemented multiple times and become conservative. Notice that the average number of observations for $\mathcal{ZAK}^{\mathcal{R}}, \mathcal{ZAK}$, and $\mathcal{ZAK}+$ does not increase significantly as θ^* increases, while the average number of observations for Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ increase dramatically. It is also expected because the feasibility check of $\mathcal{ZAK}, \mathcal{ZAK}^{\mathcal{R}}$, and $\mathcal{ZAK}+$ are designed for one critical threshold vector regardless of the number of threshold values. One may also notice that the required number of observations increase as b increases for all procedures. As the difficulty of performing feasibility check is similar under different values of b (each system requires current feasibility decision with respect of one threshold vector), larger b implies requiring more correct selections between comparisons and therefore requires larger number of observations.

Table 13: Average number of observations and observed PCS (reported in parentheses) of $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$ for $k = 100$ system and $s = 1$ constraint with two thresholds under the DM configuration

	θ^*	$\mathcal{ZAK}_1^{\mathcal{R}}$ $b=25$	$\mathcal{ZAK}_1^{\mathcal{R}}$ $b=50$	$\mathcal{ZAK}_1^{\mathcal{R}}$ $b=75$	$\mathcal{ZAK}_2^{\mathcal{R}}$ $b=25$	$\mathcal{ZAK}_2^{\mathcal{R}}$ $b=50$	$\mathcal{ZAK}_2^{\mathcal{R}}$ $b=75$
L/L	1	22806	28896	34536	22812	28910	34482
		(0.977)	(0.980)	(0.982)	(0.978)	(0.980)	(0.980)
	2	23445	30120	36336	23429	30145	36346
		(0.974)	(0.968)	(0.964)	(0.971)	(0.968)	(0.967)
L/H	1	84927	90907	96549	84909	90862	96532
		(0.976)	(0.978)	(0.981)	(0.977)	(0.975)	(0.981)
	2	88250	97620	106655	88162	97660	106699
		(0.971)	(0.967)	(0.964)	(0.974)	(0.967)	(0.964)
H/L	1	51261	81416	109293	51196	81237	109484
		(0.979)	(0.978)	(0.983)	(0.977)	(0.978)	(0.982)
	2	51622	81814	110128	51691	81828	109954
		(0.973)	(0.967)	(0.964)	(0.975)	(0.968)	(0.963)

Table 14: Average number of observations and observed PCS (reported in parentheses) of $\mathcal{ZAK}, \mathcal{ZAK}+$, Restart$^{\mathcal{AK}}$, and Restart$^{\mathcal{AK}+}$ for $k = 100$ system and $s = 1$ constraint with two thresholds under the DM configuration

	θ^*	$\mathcal{ZAK}$ $b=25$	$\mathcal{ZAK}$ $b=50$	$\mathcal{ZAK}$ $b=75$	Restart$^{\mathcal{AK}}$ $b=25$	Restart$^{\mathcal{AK}}$ $b=50$	Restart$^{\mathcal{AK}}$ $b=75$	$\mathcal{ZAK}+$ $b=25$	$\mathcal{ZAK}+$ $b=50$	$\mathcal{ZAK}+$ $b=75$	Restart$^{\mathcal{AK}+}$ $b=25$	Restart$^{\mathcal{AK}+}$ $b=50$	Restart$^{\mathcal{AK}+}$ $b=75$
L/L	1	17667	20270	23203	20107	23258	26629	19101	21499	23511	19084	21538	23599
		(0.970)	(0.972)	(0.980)	(0.984)	(0.990)	(0.990)	(0.977)	(0.977)	(0.979)	(0.977)	(0.981)	(0.981)
	2	17879	20697	23813	30286	35850	41745	19211	21889	24210	27968	32562	36617
		(0.967)	(0.965)	(0.966)	(0.984)	(0.982)	(0.982)	(0.972)	(0.970)	(0.968)	(0.973)	(0.969)	(0.963)
L/H	1	78078	78059	78178	88475	88612	88728	72485	63744	54254	72485	63897	54464
		(0.981)	(0.987)	(0.994)	(0.990)	(0.993)	(0.997)	(0.982)	(0.981)	(0.985)	(0.980)	(0.980)	(0.985)
	2	81276	84859	88405	138753	151277	163560	79368	79721	80028	116566	118678	120211
		(0.975)	(0.978)	(0.975)	(0.989)	(0.989)	(0.987)	(0.973)	(0.971)	(0.965)	(0.975)	(0.968)	(0.969)
H/L	1	37475	63883	91651	43652	74449	106928	47199	73615	97505	47187	73624	97958
		(0.968)	(0.971)	(0.975)	(0.982)	(0.986)	(0.987)	(0.979)	(0.979)	(0.980)	(0.977)	(0.977)	(0.982)
	2	37370	63339	90343	53787	86921	121584	46939	72943	96577	55892	84079	109815
		(0.962)	(0.962)	(0.961)	(0.978)	(0.980)	(0.982)	(0.970)	(0.967)	(0.964)	(0.972)	(0.968)	(0.964)

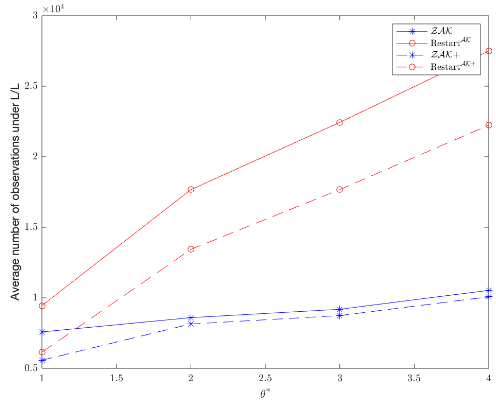
We then consider the case of four thresholds under the DM configuration ($q_1^1 = 0, q_1^2 = 2\epsilon_1, q_1^3 = 4\epsilon_1$, and $q_1^4 = 6\epsilon_1$). We show the results $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$ and the results for comparison between $\mathcal{ZAK}, \text{Restart}^{\mathcal{AK}}$ and $\mathcal{ZAK}+, \text{Restart}^{\mathcal{AK}+}$ in Tables 15 and 16, respectively, for $\theta^* \in \{1, 2, 3, 4\}$. We see that the performance between $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$ are still similar under all cases.

As the number of thresholds is larger than before, the implementation parameters for Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}}$+ take more conservative values while $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ use the same implementation parameters as the case of two thresholds. Thus $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ perform better than Restart$^{\mathcal{AK}}$ and $\mathcal{ZAK}+$ even for $\theta^* = 1$ and their efficiency becomes more profound as θ^* gets larger. We also notice that $\mathcal{ZAK}$ performs slightly better compared with $\mathcal{ZAK}+$ under L/L and H/L in most cases but performs slightly worse under L/H in terms of the required number of observations.

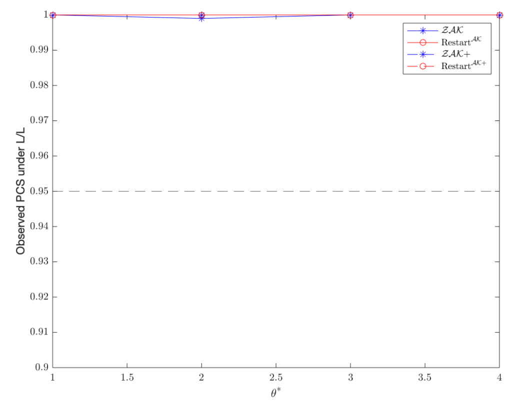
We also test the performance of $\mathcal{ZAK}^{\mathcal{R}}, \mathcal{ZAK}$, and $\mathcal{ZAK}+$ under the MIM configuration for both two and four thresholds. We show the experimental results for $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$ for two thresholds in Table 17 and the comparison between $\mathcal{ZAK}, \text{Restart}^{\mathcal{AK}}$ and $\mathcal{ZAK}+, \text{Restart}^{\mathcal{AK}+}$ in Table 18. The results under the MIM configuration show a similar pattern to the results under the DM configuration. In general, $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ perform similarly to Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ when $\theta^* = 1$ but perform better when $\theta^* = 2$.

As $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$ still have similar performance, we omit the results for future experiments for simplicity. Since the results for different b have similar pattern, we only show the results for $b = 50$ when $d = 4$ in Figure 10. We see that $\mathcal{ZAK}$ performs slightly worse than Restart$^{\mathcal{AK}}$ under LL and LH when $\theta^* = 1$ but better under all other cases. $\mathcal{ZAK}+$ performs better than Restart$^{\mathcal{AK}+}$ under all cases.

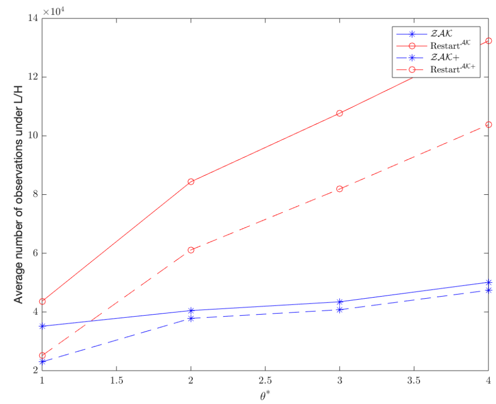
To further illustrate that $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ perform even better under a larger θ^*, we test ten thresholds (and also $b = 50$) and show experimental results in Figure 11. We see that average numbers of observations of $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ remain similarly as θ^* increases under both DM and MIM, while those of Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ increase dramatically. Again, this shows that our proposed procedures scale well with respect to the number of threshold vectors in consideration, and it is expected that the saving of $\mathcal{ZAK}$ and $\mathcal{ZAK}+$



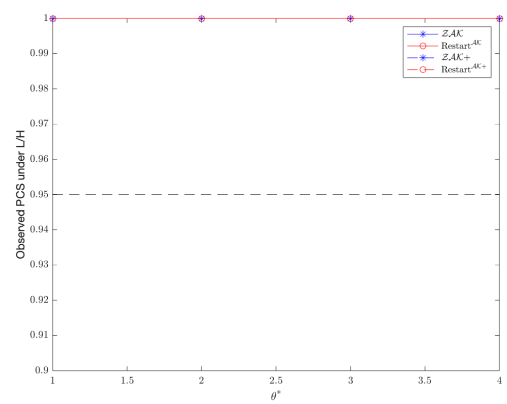
(a) OBS with LL



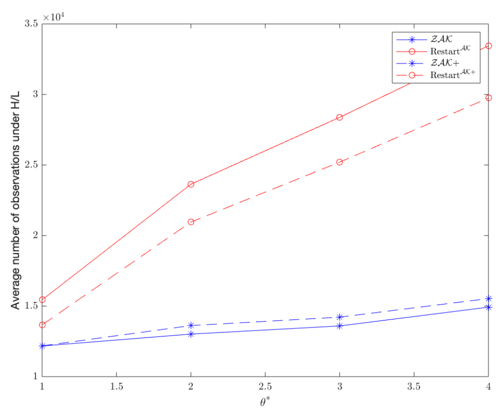
(b) PCS under LL



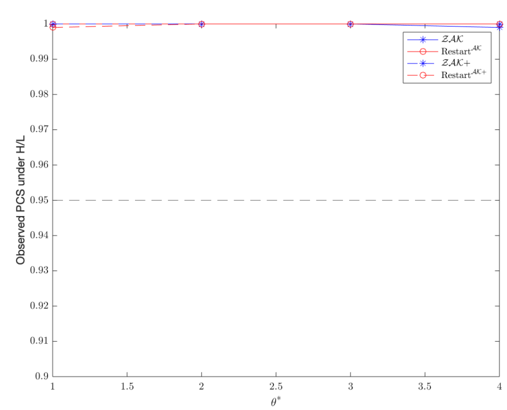
(c) OBS under LH



(d) PCS under LH



(e) OBS under HL



(f) PCS under HL

Figure 10: Average number of observations and observed PCS of $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}+$ and Restart$^{\mathcal{HAK}+}$ for $k = 100$ systems and $s = 1$ constraint with four thresholds and $b = 50$ under the MIM configuration and different variance configuration

Table 15: Average number of observations and observed PCS (reported in parentheses) of $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$ for $k = 100$ system and $s = 1$ constraint with four thresholds under the DM configuration

		$\mathcal{ZAK}_1^{\mathcal{R}}$			$\mathcal{ZAK}_2^{\mathcal{R}}$		
	θ^*	$b = 25$	$b = 50$	$b = 75$	$b = 25$	$b = 50$	$b = 75$
L/L	1	22810	28896	34536	22794	28890	34482
		(0.973)	(0.980)	(0.982)	(0.978)	(0.979)	(0.980)
	2	23439	30144	36434	23466	30100	36248
		(0.971)	(0.965)	(0.967)	(0.975)	(0.967)	(0.963)
	3	23497	30111	36402	23463	30129	36327
		(0.974)	(0.968)	(0.968)	(0.975)	(0.972)	(0.964)
	4	23448	30099	36335	23396	30137	36376
		(0.975)	(0.966)	(0.963)	(0.970)	(0.969)	(0.965)
L/H	1	84959	90988	96549	84911	90871	96534
		(0.978)	(0.978)	(0.981)	(0.975)	(0.978)	(0.982)
	2	88392	97708	106771	88413	97582	106733
		(0.972)	(0.970)	(0.966)	(0.971)	(0.964)	(0.965)
	3	88322	97671	106704	88341	97657	106779
		(0.973)	(0.969)	(0.964)	(0.969)	(0.965)	(0.967)
	4	88195	97602	106650	88170	97621	106699
		(0.971)	(0.965)	(0.961)	(0.974)	(0.968)	(0.964)
H/L	1	51271	81257	109319	51229	81247	109325
		(0.978)	(0.980)	(0.981)	(0.977)	(0.977)	(0.982)
	2	51551	81831	109641	51568	82231	109684
		(0.972)	(0.969)	(0.962)	(0.972)	(0.967)	(0.965)
	3	51437	82135	110345	51619	82140	109729
		(0.970)	(0.970)	(0.966)	(0.974)	(0.967)	(0.962)
	4	51690	81977	110093	51664	82021	109887
		(0.972)	(0.969)	(0.966)	(0.972)	(0.967)	(0.964)

become more when θ^* gets larger.

3.6.3 PCS of $\mathcal{ZAK}+$ when Different Number of Unacceptable Systems are in Presence

In this section, we discuss how the value of PCS of procedure $\mathcal{ZAK}+$ changes as the value of j, the number of unacceptable systems with respect to threshold vector $\boldsymbol{q}^{(\theta^*)}$, changes.

As discussed in Section 3.4.3, the value of the lower bound of Pr{CS} of the $\mathcal{ZAK}+$ procedure depends on the value of j and one can see that the lower bound of Pr{CS} is achieved when $j = 0$ when there is single or two constraints and $s\beta_f = \beta_c$. Tables 19 and 20 report the observed PCS of $\mathcal{ZAK}+$ when the number of constraints is one and two (we choose the ranked constraint configuration when there are two constraints), respectively.

Table 16: Average number of observations and observed PCD (reported in parentheses) of $\mathcal{ZAK}, \mathcal{ZAK}+, \text{Restart}^{\mathcal{AK}}$, and $\text{Restart}^{\mathcal{AK}+}$ for $k = 100$ system and $s = 1$ constraint with four thresholds under the DM configuration

		$\mathcal{ZAK}$			$\text{Restart}^{\mathcal{AK}}$			$\mathcal{ZAK}+$			$\text{Restart}^{\mathcal{AK}+}$		
	θ^*	$b=25$	$b=50$	$b=75$	$b=25$	$b=50$	$b=75$	$b=25$	$b=50$	$b=75$	$b=25$	$b=50$	$b=75$
L/L	1	17672	20270	23198	22810	26449	30482	19109	21497	23511	21772	24694	27208
		(0.971)	(0.972)	(0.980)	(0.992)	(0.993)	(0.996)	(0.977)	(0.979)	(0.979)	(0.990)	(0.990)	(0.990)
	2	17917	20702	23789	34307	40748	47564	19257	21894	24218	31866	37354	42245
		(0.963)	(0.964)	(0.970)	(0.991)	(0.990)	(0.992)	(0.972)	(0.970)	(0.966)	(0.987)	(0.985)	(0.983)
	3	17922	20720	23772	40559	47819	55396	19263	21909	24223	37466	43528	49122
		(0.965)	(0.967)	(0.966)	(0.990)	(0.989)	(0.992)	(0.973)	(0.969)	(0.965)	(0.987)	(0.982)	(0.982)
	4	17882	20710	23794	44969	52589	60495	19198	21889	24234	41372	47791	53768
		(0.965)	(0.966)	(0.966)	(0.993)	(0.992)	(0.989)	(0.972)	(0.970)	(0.967)	(0.984)	(0.985)	(0.983)
L/H	1	78091	78059	78178	99842	99974	100101	72474	63807	54325	82655	73066	62719
		(0.980)	(0.987)	(0.994)	(0.995)	(0.977)	(0.998)	(0.977)	(0.981)	(0.984)	(0.989)	(0.990)	(0.990)
	2	81467	84886	88383	156723	170774	185127	79485	79786	80016	132812	135543	137747
		(0.976)	(0.976)	(0.976)	(0.993)	(0.993)	(0.993)	(0.972)	(0.970)	(0.968)	(0.988)	(0.983)	(0.983)
	3	81475	84895	88409	187901	205988	224225	79458	79753	80070	160206	166670	172151
		(0.978)	(0.975)	(0.977)	(0.993)	(0.993)	(0.994)	(0.973)	(0.970)	(0.970)	(0.984)	(0.984)	(0.983)
	4	81325	84851	88416	209685	229585	249572	79279	79733	80046	179703	187617	194643
		(0.959)	(0.962)	(0.959)	(0.989)	(0.990)	(0.989)	(0.972)	(0.972)	(0.969)	(0.986)	(0.986)	(0.984)
H/L	1	37505	63878	91651	50442	86231	123963	47247	73549	97505	54270	85117	113641
		(0.966)	(0.971)	(0.975)	(0.992)	(0.991)	(0.994)	(0.981)	(0.980)	(0.980)	(0.988)	(0.989)	(0.990)
	2	37338	63341	90530	61892	100541	140325	47107	72730	96473	64284	97475	128203
		(0.962)	(0.962)	(0.964)	(0.989)	(0.990)	(0.989)	(0.974)	(0.969)	(0.965)	(0.986)	(0.984)	(0.981)
	3	37366	63404	90394	68165	107118	147997	47120	72658	96473	69919	103666	134915
		(0.960)	(0.962)	(0.958)	(0.990)	(0.989)	(0.989)	(0.975)	(0.966)	(0.965)	(0.986)	(0.985)	(0.980)
	4	37348	63258	90474	72662	112284	153429	46939	72943	96577	73776	108173	139815
		(0.959)	(0.962)	(0.959)	(0.989)	(0.990)	(0.989)	(0.970)	(0.967)	(0.964)	(0.985)	(0.985)	(0.982)

Table 17: Average number of observations and observed PCS (reported in parentheses) of $\mathcal{ZAK}_1^{\mathcal{R}}$ and $\mathcal{ZAK}_2^{\mathcal{R}}$ for $k = 100$ system and $s = 1$ constraint with two thresholds under the MIM configuration

		$\mathcal{ZAK}_1^{\mathcal{R}}$			$\mathcal{ZAK}_2^{\mathcal{R}}$		
	θ^*	$b=25$	$b=50$	$b=75$	$b=25$	$b=50$	$b=75$
L/L	1	9222	10244	11429	9205	10243	11412
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
	2	11417	12250	13079	11410	12243	13068
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
L/H	1	39044	42151	46014	39017	42108	45990
		(1.000)	(1.000)	(1.000)	(0.999)	(1.000)	(1.000)
	2	50001	52147	54276	49928	52092	54247
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
H/L	1	15658	17294	18557	15689	17277	18642
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
	2	17843	19272	20264	17846	19282	20247
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)

Both experimental results are under the DM configuration.

Although the actual observed PCS does not always achieve its minimum when $j = 0$, one may find that the value of observered PCS tends to become higher when j increases.

Table 18: Average number of observations and observed PCS (reported in parentheses) of $\mathcal{ZAK}, \mathcal{ZAK}+$, Restart$^{\mathcal{AK}}$, and Restart$^{\mathcal{AK}+}$ for $k = 100$ system and $s = 1$ constraint with two thresholds under the MIM configuration

		$\mathcal{ZAK}$			Restart$^{\mathcal{AK}}$			$\mathcal{ZAK}+$			Restart$^{\mathcal{AK}+}$		
	θ^*	$b=25$	$b=50$	$b=75$	$b=25$	$b=50$	$b=75$	$b=25$	$b=50$	$b=75$	$b=25$	$b=50$	$b=75$
L/L	1	7936	8571	9320	8701	9607	10540	7402	6556	5732	7144	6466	5705
		(0.999)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
	2	10087	10532	10932	16903	18662	20384	10070	10074	10079	14389	14434	14367
		(0.999)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
L/H	1	37434	40092	43509	40989	44933	49167	32700	28030	23702	31376	27579	23500
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
	2	48445	50114	51801	81637	89794	98059	47349	47408	47426	67261	67158	66533
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
H/L	1	11690	13145	14389	13094	14846	16334	13418	13180	12616	13208	13081	12623
		(0.999)	(1.000)	(1.000)	(0.999)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
	2	13712	14924	15810	21269	23922	26169	15455	15533	15550	20389	21068	21283
		(0.999)	(0.999)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)

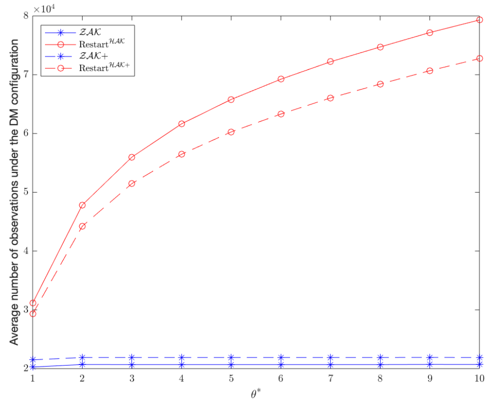
Table 19: Observed PCS of $\mathcal{ZAK}+$ for $k = 100$ system and $s = 1$ constraint with ten thresholds under the DM configuration and different number of unacceptable systems with respect to $\boldsymbol{q}^{(\theta^*)}$

j \ θ^*	1	2	3	5	7	10
0	0.983	0.966	0.966	0.966	0.966	0.965
10	0.983	0.964	0.964	0.964	0.964	0.962
20	0.981	0.966	0.960	0.960	0.960	0.962
30	0.981	0.968	0.968	0.968	0.968	0.965
50	0.979	0.970	0.970	0.970	0.970	0.970
70	0.978	0.971	0.971	0.971	0.971	0.973
90	0.977	0.976	0.976	0.976	0.976	0.979

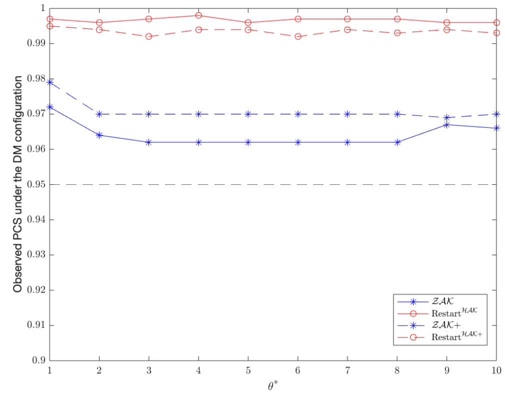
Table 20: Observed PCS (reported in parentheses) of $\mathcal{ZAK}+$ for $k = 100$ system and $s = 2$ constraint under the DM configuration and different number of unacceptable systems with respect to $\boldsymbol{q}^{(\theta^*)}$

j \ θ^*	1	2	3	4	5	9
0	0.983	0.970	0.972	0.974	0.964	0.963
10	0.985	0.976	0.973	0.975	0.962	0.965
20	0.985	0.974	0.975	0.973	0.965	0.969
30	0.984	0.976	0.978	0.976	0.966	0.973
50	0.984	0.981	0.979	0.975	0.969	0.981
70	0.986	0.982	0.984	0.975	0.971	0.988
90	0.989	0.988	0.987	0.979	0.976	0.995

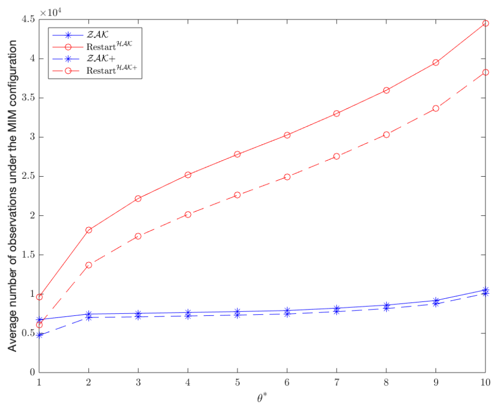
For later sections, we test the cases when $j = 0$ (i.e., all the systems are desirable with respect to threshold vector $\boldsymbol{q}^{(\theta^*)}$). This is also equivalent to $b = 100$.



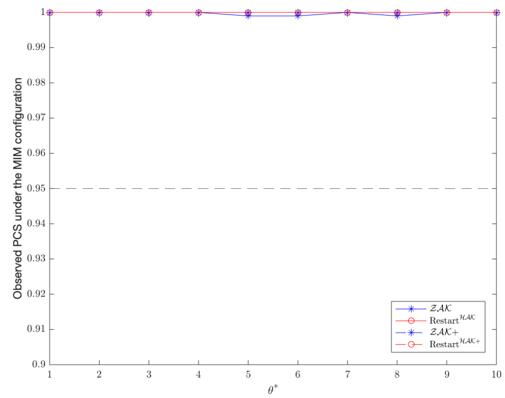
(a) OBS under DM



(b) PCS under DM



(c) OBS under MIM



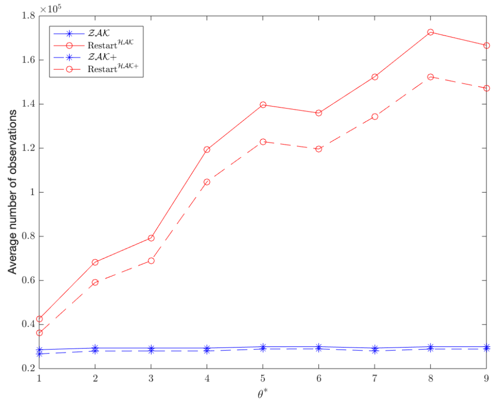
(d) PCS under MIM

Figure 11: Average number of observations and observed PCS of $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}+$ and Restart$^{\mathcal{HAK}+}$ for $k = 100$ systems and $s = 1$ constraint with ten thresholds and $b = 50$ under the DM and the MIM configuration

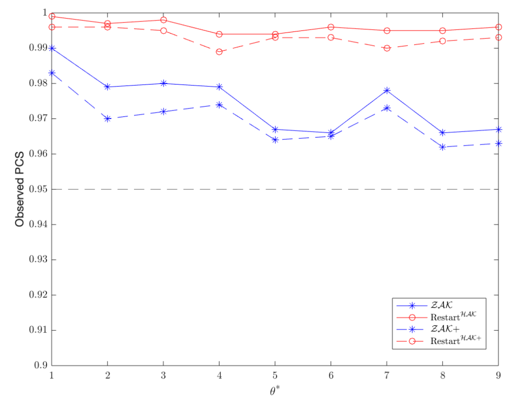
3.6.4 Two Constraints

In this section, we consider two constraints with the three different formulation of the input threhold vectors discussed in Section 3.5. We set $k = 100$, and $d_\ell = 3$ for $\ell = 1, 2$. We also set $q_\ell = \{0, 2\epsilon_\ell, 4\epsilon_\ell\}$ under the DM configuration and $q_\ell = \{0, 4\epsilon_\ell, 8\epsilon_\ell\}$ under the MIM configuration for $\ell = 1, 2$, and perform experiments under the L/L variance configuration.

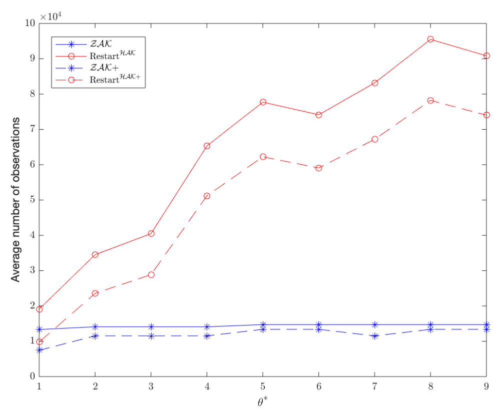
We first show the experimental results for the ranked constraint formulation under the DM and the MIM configurations in Figure 12. As there are 9 threshold vectors in total (i.e., $d = 9$), the implementation parameters for both Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ take conservative values. We see that $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ perform better than their competing counterpart procedures under all cases and the efficiency of our proposed procedures becomes more obvious as θ^* increases. Notice that similar to the single constraint case, $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ scale well in terms of OBS as θ^* increases, however, the average numbers of observations of Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ increase dramatically as θ^* increases.



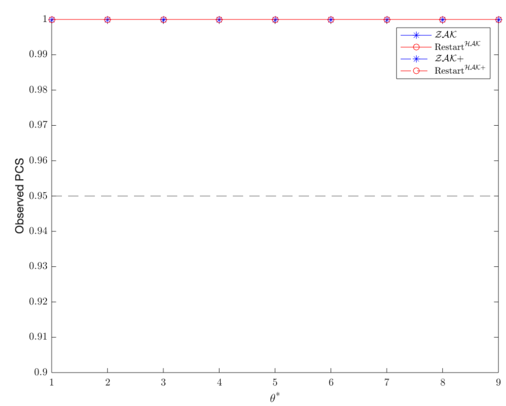
(a) OBS under DM



(b) PCS under DM



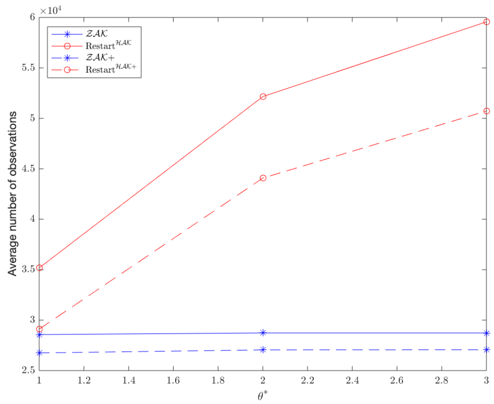
(c) OBS under MIM



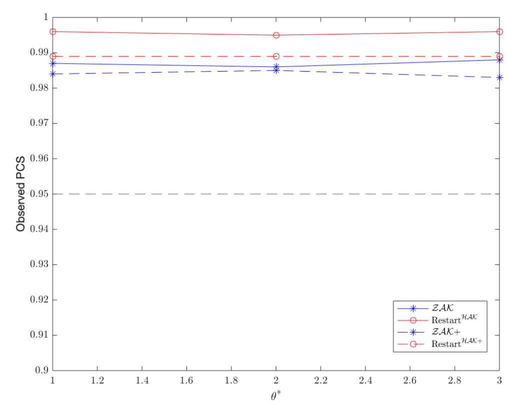
(d) PCS under MIM

Figure 12: Average number of observations and observed PCS of $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}+$ and Restart$^{\mathcal{HAK}+}$ for $k = 100$ systems and $s = 2$ constraint under the DM configuration and Option 1

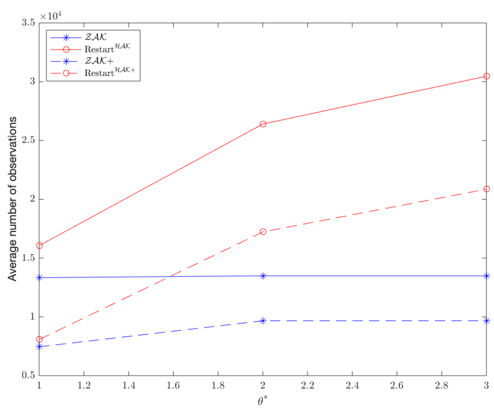
We then show the experimental results for the formulation of equally important constraints under the DM configuration and the MIM configuration in Figure 13. Since the possible number of threshold vectors $d = 3$ is relatively small, setting the implementation parameter for Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ is less conservative compared with that in the formulation of ranked constraints. We see that both $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ perform better, but the difference between the proposed procedures and their alternative procedures is less compared with that in the ranked constraints formulation.



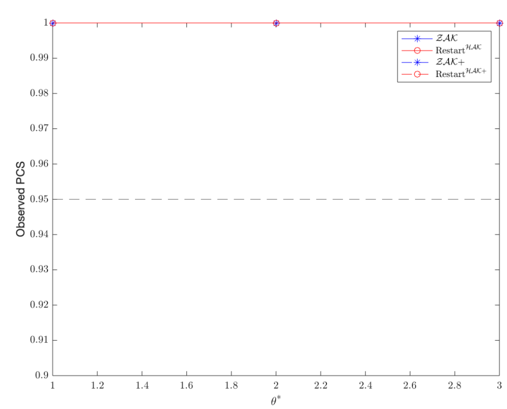
(a) OBS under DM



(b) PCS under DM



(c) OBS under MIM



(d) PCS under MIM

Figure 13: Average number of observations and observed PCS of $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}+$ and Restart$^{\mathcal{HAK}+}$ for $k = 100$ systems and $s = 2$ constraint under the DM configuration and Option 2

Total violation with ranked constraints formulation has the same number of threshold vectors as that in ranked constraints formulation and all procedures show very similar OBS and PCS compared to that of ranked constraints formulation. Thus we omit the experimental results for this formulation for the sake of brevity.

3.6.5 Four Constraints

In this section, we consider four constraints case in order to demonstrate the performance of our proposed procedures in handling multiple constraints.

We choose the formulation of ranked constraints. The number of thresholds is set to four on the first and the fourth constraints ($d_1 = d_4 = 4$), two on the second constraint ($d_2 = 2$), and three on the third constraint ($d_3 = 3$). The configuration of thresholds on each constraint is as follows.

- Constraint 1: $q_1 \in \{0, 2\epsilon_1, 4\epsilon_1, 6\epsilon_1\}$ for the DM configuration, and $q_1 \in \{0, 4\epsilon_1, 8\epsilon_1, 12\epsilon_1\}$ for the MIM configuration.

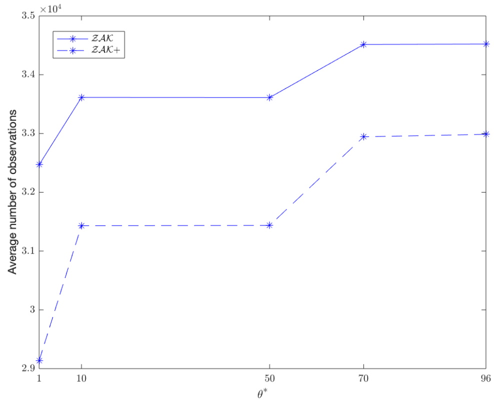
- Constraint 2: $q_2 \in \{0, 2\epsilon_2\}$ for the DM configuration, and $q_2 \in \{0, 4\epsilon_2\}$ for the MIM configuration.

- Constraint 3: $q_3 \in \{0, 2\epsilon_3, 4\epsilon_3\}$ for the DM configuration, and $q_3 \in \{0, 4\epsilon_3, 8\epsilon_3\}$ for the MIM configuration.

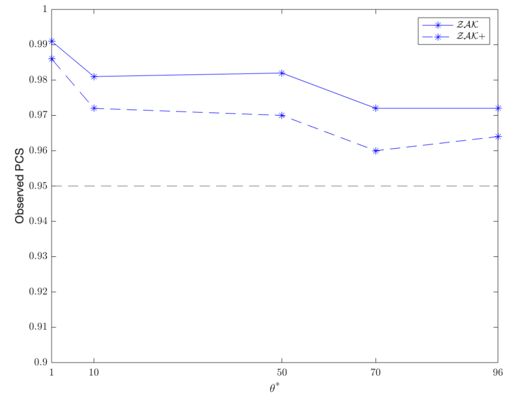
- Constraint 4: $q_4 \in \{0, 2\epsilon_4, 4\epsilon_4, 6\epsilon_4\}$ for the DM configuration, and $q_4 \in \{0, 4\epsilon_4, 8\epsilon_4, 12\epsilon_4\}$ for the MIM configuration.

Ranked constraints formulation considers 96 ($= 4 \times 2 \times 3 \times 4$) threshold vectors in total. Due to the large number of the threshold vectors, Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ take extreme conservative values for the implemetation parameters and we indeed observe extremely long run-time in order to conclude the selection for both procedures. On the other hand, $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ scale well to the number of threshold vectors. Due to the extremely long run time of Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$, we only provide the experimental results for $\mathcal{ZAK}$ and $\mathcal{ZAK}+$. We consider $\theta^* \in \{1, 10, 50, 70, 96\}$ and show the results in Figure 14.

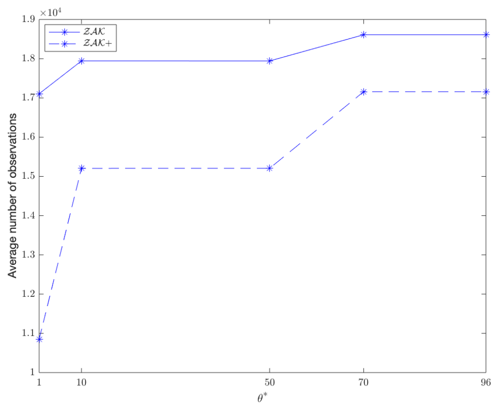
We see that both $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ have observed PCS greater than 0.95 and the average numbers of observations remain similar to different values of θ^*. It is also expected that the number of observations would remain similarly if more thresholds are considered.



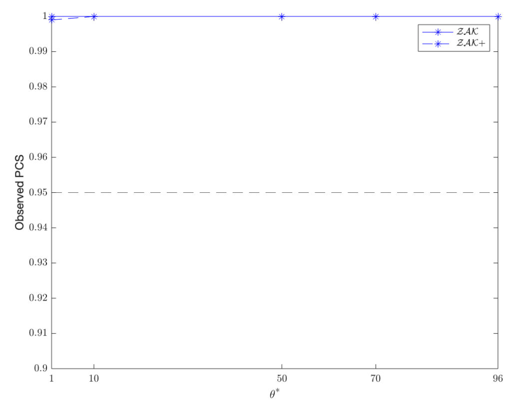
(a) OBS under the DM



(b) PCS under the DM



(c) OBS under MIM



(d) PCS under MIM

Figure 14: Average number of observations and observed PCS of $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ for $k = 100$ systems and $s = 4$ constraint

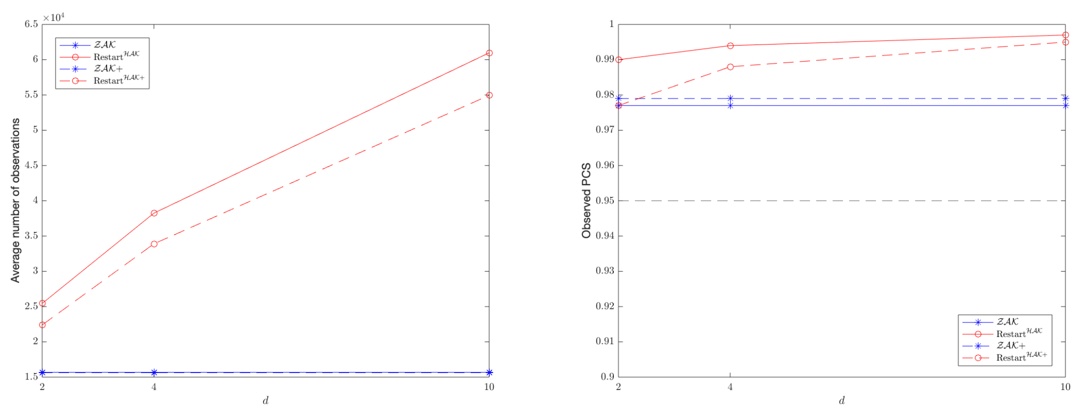
3.6.6 All Systems Infeasible

In this section, we discuss a case when all systems are infeasible with respect to all threshold vectors $\boldsymbol{q}^{(\theta)}$ for $\theta = 1, \ldots, d$. In this case, Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ need to be implemented to all threshold vectors and declare no feasible systems eventually. Since both $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ scale well with respect to the number of threshold vectors, we expect to observe the maximum possible efficiency of our proposed procedures compared to alternative procedures.

We first consider single constraint case with 2, 4, and 10 thresholds (i.e., $d \in \{2, 4, 10\}$).

We follow the same setting of the secondary performance measure as that in the DM configuration (i.e., each pair of consecutive thresholds are $2\epsilon_1$ away) and all the systems have the mean of the secondary performance measure $2\epsilon_1$ larger than the largest threshold value on the constraint. We also set the mean of the primary performance measure of all systems to 0. The experimental results are shown in Figure 15.

We see that the number of observations of Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ are much larger than that of $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ and the difference increases as the number of thresholds increases. This is expected as Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ need to implement more times and therefore the implementation parameters take more conservative values as d increases. One may also notice that the required number of $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ remains stable when d increases.



(a) OBS (b) PCS

Figure 15: Average number of observations and observed PCS of $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}+$ and Restart$^{\mathcal{HAK}+}$ for $k = 100$ systems and $s = 1$ constraint under the DM configuration with $d \in \{2, 4, 10\}$

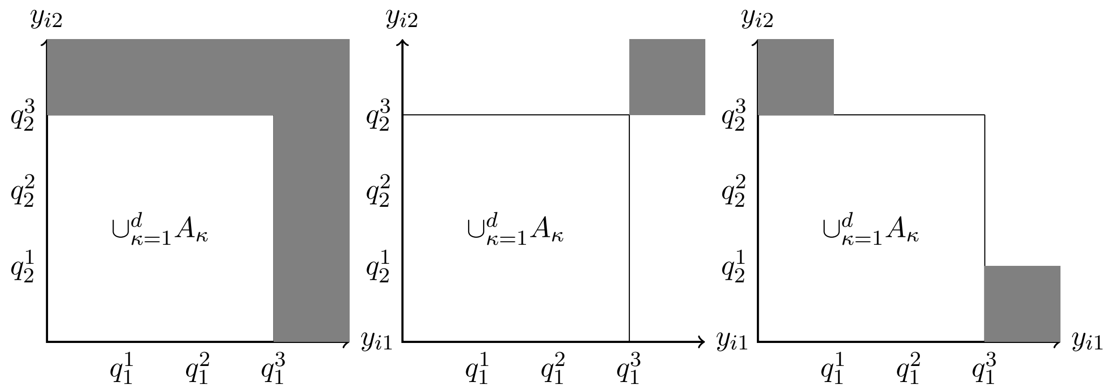
We then consider the case when there are two constraints with three thresholds on each constraint ($d_\ell = 3$, where $\ell = 1, 2$). Although there can be many different ways of setting the mean performance of the infeasible systems, we consider the following three cases in particular.

- Case 1: Infeasible systems are evenly distributed to the regions when at least one constraint is infeasible to threshold q_ℓ^3, where $\ell = 1, 2$.

- Case 2: Infeasible systems all fall in the region that are infeasible to both constraint with thresholds q_ℓ^3, where $\ell = 1, 2$.

- Case 3: Infeasible systems are evenly distributed to the region that one constraint is infeasible to threshold q_ℓ^3, where $\ell = 1, 2$, while the other constraint is feasible to the most preferred threshold value on that constraints.

We use Figure 16 to show these three cases where we use shaded region to denote the region that infeasible systems evenly fall into. We choose the ranked constraints formulation



(a) Case 1 (b) Case 2 (c) Case 3

Figure 16: Three distribution of infeasible systems

and show the experimental results in Table 21. As there are $d = 9$ threshold vectors in consideration, Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ need to implement nine times (once with respect to each threshold vector) to conclude that there do not exist any feasible systems, and therefore they take much more conservative values for the implemetation parameters compared with that of $\mathcal{ZAK}$ and $\mathcal{ZAK}+$. Thus, it is expected that the performance of $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ is better. One may also notice that Case 3 requires more observations while Case 2 requires least observations for all four procedures. This is because declaring a system infeasible with respect to q_1^3 on the first constraint or q_2^3 on the second constraint is suffient to declare a system infeasible with respect to all threshold vectors in Case 2, while an infeasible system in Case 3 has only one constraint that the system is infeasible with respect to all the threshold values on it. Moreover, an infeasible system in Case 3 is feasible to the most preferred threshold value on one of the constraints, which makes the feasibility check more difficult. As Case 1 is a combination of Case 2 and 3, it is expected that it requires more observations than that of Case 2 but less observations than that of Case 3.

Table 21: Average number of observations and observed PCS (reported in parentheses) of $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}$+ and Restart$^{\mathcal{HAK}+}$ for $k = 100$ system and $s = 2$ constraint when no feasible systems exists to all thresholds under the DM configuration and the ranked constraints formulation

	$\mathcal{ZAK}$	Restart$^{\mathcal{HAK}}$	$\mathcal{ZAK}$+	Restart$^{\mathcal{HAK}+}$
Case 1	17266	116475	17204	104267
	(0.990)	(0.998)	(0.993)	(0.996)
Case 2	13962	83888	13915	75059
	(1.000)	(1.000)	(1.000)	(1.000)
Case 3	17807	130889	17743	117243
	(0.990)	(0.996)	(0.987)	(0.993)

3.6.7 Inventory Policy Example

In this section, we study the performance of $\mathcal{ZAK}$ and $\mathcal{ZAK}$+, as well as their competeing procedures Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$, based on an (s, S) inventory policy example from [11].

A decision maker controls inventory using (s, S) policy, and the costs are given as (i) ordering cost at 3 per item; (ii) fixed ordering cost at 32 per order; (iii) holding cost at 1 per item per review period; and (iv) penalty cost at 5 per item of unsatisfied demand. She considers the primary performance measure as maximizing the fill rate per review period and two secondary performance measures as (1) the failure probability ($\ell = 1$), which is the probability that a shortage occurs between two successive review periods; and (2) the expected cost per review period ($\ell = 2$), which is the average total cost for each review period.

Systems in consideration are given as

$$\Gamma = \left\{(s, S)|s = 20 + 2m', S = 40 + 10n', \text{ where } m' = 0, 1, 2, \ldots, 10, \text{ and } n' = 0, 1, 2, \ldots, 6\right\},$$

which contains 77 systems in total. Demand during each review period is assumed indenpendelty and follows Poisson distribution with mean 25. The primary and secondary performance measures are computed based on 100 review periods. The correlation between two constraints is estimated analytically using a Markov chain model. The estimated correlation ranges from -0.23 to 0.52.

We test procedures $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}+$ and Restart$^{\mathcal{HAK}+}$ with three thresholds on the first constraint ($q_1 \in \{0.01, 0.05, 0.1\}$) and eight thresholds on the second constraint ($q_2 \in \{100, 105, 110, 115, 120, 125,$
$130, 135\}$). We formulate the input threshold vectors based on ranked constraints, where we prioritize the first constraint over the second constraint (relax the second constraint first). An analytical analysis based on a Markov chain model shows the best system is $(28, 60)$ whose analytical values of failure probability is 0.0211 and expected cost per review period is 113.9701. The value of θ^* is 11, which correspond to the threshold vector $(q_1, q_2) = (0.05, 115)$. The experimental result is shown in Table 22.

Table 22: Average number of observations and observed PCS (reported in parentheses) of $\mathcal{ZAK}$, Restart$^{\mathcal{HAK}}$, $\mathcal{ZAK}+$ and Restart$^{\mathcal{HAK}+}$ for the inventory policy example

	$\mathcal{ZAK}$	Restart$^{\mathcal{HAK}}$	$\mathcal{ZAK}+$	Restart$^{\mathcal{HAK}+}$
Without CRN	5117 (1.000)	27045 (1.000)	3478 (1.000)	22758 (1.000)
With CRN	5186 (1.000)	27064 (1.000)	3528 (1.000)	22781 (1.000)

We see that $\mathcal{ZAK}$ and $\mathcal{ZAK}+$ spend less than one fifth of the observations compared tothat of Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$, respectively. Both proposed procedures perform much better than their alternative procedures while also remaining statistically valid at the same time. One may also notice that the required number of observations when systems are simulated with CRN is slightly more than that when systems are simulated indepedently. This is because the implemetation parameters takes a conservative value when CRN is applied to incorporate the depedency between systems.

3.7 Conclusion

We consider the selection-of-the-best problem when subjective stochastic constraints are present. When a decision maker has flexibility with thresholds, she may be willing to consider multiple threshold values for each constraint. We discuss how to combine thresholds on constraints into threshold vectors based on how a decision maker prioritizes each constraint. We propose two procedures that select the best system with respect to a primary performance measure while also satisfying secondary performance measures with respect to the most preferred threshold possible, namely one that runs feasibility check and comparison sequentially and another that runs them simultaneously. We discuss the statistical validity of the proposed procedures and show that the required number of observations remains steady when the number of threshold vectors grows. Our experimental results also show that the proposed procedures perform well in reducing the average number of needed observations as compared with procedures that repeatedly solve the problem for each threshold vector.

CHAPTER IV

FINDING A PORTFOLIO OF BEST SYSTEMS FOR SUBJECTIVE CONSTRAINTS

4.1 Introduction

We consider the problem of finding a portfolio of the best systems with respect to a primary performance measure among a finite number of simulated systems as stochastic constraints on secondary performance measures are relaxed. Thus the constraints are subjective in that their thresholds can be relaxed. When multiple thresholds are considered, there can be a system that is infeasible to more preferred thresholds but has a significantly better primary performance compared with those systems that are feasible to the more preferred thresholds. Consequently, the information of how the identity and performance of the best system depends on the thresholds can be helpful in decision making.

For example, suppose that a decision maker uses an (s, S) inventory policy (namely ordering products to increase the inventory level up to S when the inventory level at a review period is below s, and placing no order otherwise) to manage the inventory level. The decision maker hopes to identify the values of S and s (among finitely many possible values) such that the expected fill rate within each review period is maximized. In addition, the chosen values of S and s should satisfy two constraints: (i) the probability that a shortage occurs between two successive review periods is small (less than or equal to q_1 percent); (ii) the expected cost per review period is small (less than or equal to q_2 thousand dollars). The decision maker may consider multiple values for both constraint thresholds (e.g., $q_1 \in \{1, 5, 10\}$ and $q_2 \in \{100, 105, 110\}$). Rather than selecting the values of S and s that achieve the highest fill rate and are also feasible with respect to both constraints with fixed thresholds, the decision maker is interested in how the values of S and s depend on the different combinations of thresholds on both constraints (e.g., $(q_1, q_2) = (1, 100), (1, 105), (1, 110)$, etc.).

The information of the portfolio of best solutions (values of S and s) with respect to each combination of thresholds can be beneficial in decision making as: (1) the decision maker may identify a robust solution with respect to different levels of feasibility on constraints; and (2) she can consider the trade-off between the primary performance measure and the level of feasibility on constraints. For example, if a specific set of (s, S) keeps appearing as the best feasible choice for different values of the thresholds, then this choice is robust to the level of feasibility. Or if quite different solutions are best when thresholds are relaxed, one can estimate the difference between their primary performance values. If the difference is large, it means that a large sacrifice is needed to achieve a tighter level of feasibility, and the decision maker can make her choice with this information in mind.

In this chapter, we propose fully sequential procedures that identify a portfolio of best systems with respect to each set of relaxed thresholds in consideration so that the decision maker can consider the robustness of the solutions and the trade-off between performance measures. We prove the statistical validity of our proposed procedures and also document their efficiency in terms of reducing the required number of simulation observations until the decision is made, as compared with straight-forward repeating procedures such as applying the procedures of [1] or [5] (depending on whether the problem has one or more constraints) repeatedly to each possible set of threshold values.

The rest of this chapter is organized as follows: Section 4.2 provides the background for our problem. Sections 4.3 and 4.4 discuss our proposed sequential and simultaneously running procedures, respectively. Section 4.5 shows experimental results to demonstrate the performance of our proposed procedures. Concluding remarks are provided in Section 4.6. A discussion of two alternative procedures that we compare with our proposed procedures is included in Appendices C.1 and C.2.

4.2 Background

In this section, we provide the problem formulation in Section 4.2.1, and discuss how we define our correct selection event in Section 4.2.2. The necessary assumptions that guarantee the statistical validity of our proposed procedures are provided in Section 4.2.3.

4.2.1 Problem Formulation

We consider $k \geq 2$ systems whose primary performance measure and s secondary performance measures can be estimated using stochastic simulation, and let Γ denote the index set of all possible systems (i.e., $\Gamma = \{1, \ldots, k\}$). We use X_{in} and $Y_{i\ell n}$ to denote the observation associated with the primary performance measure and the ℓth secondary performance measure, where $\ell = 1, \ldots, s$, of system i from replication n, respectively. The expected values of the primary and secondary performance measure for system i, where $i = 1, \ldots, k$, and constraint ℓ, where $\ell = 1, \ldots, s$, are defined as $x_i = \mathrm{E}[X_{in}]$ and $y_{i\ell} = \mathrm{E}[Y_{i\ell n}]$, respectively. Constrained R&S is to select

$$\begin{aligned} &\arg\max_{i \in \Gamma} \quad x_i \\ &\text{s.t.} \qquad\qquad y_{i\ell} \leq q_\ell \quad \text{for all } \ell = 1, \ldots, s, \end{aligned}$$

where q_ℓ denotes the constraint threshold for constraint ℓ. For a given threshold vector $\boldsymbol{q} = (q_1, \ldots, q_s)$, procedures due to [1] and [5] can be used to select the best feasible system. In this paper, we assume that the decision maker considers multiple constraint thresholds on some or all constraints and aim to find a portfolio of best systems with respect to each combination of thresholds. We let d_ℓ denote the number of the distinct threshold values on constraint ℓ that the decision maker is interested in and let q_ℓ^m denote the mth distinct threshold value on constraint ℓ, where $\ell = 1, \ldots, s$ and $m = 1, \ldots, d_\ell$. We also assume $q_\ell^1 < \ldots < q_\ell^{d_\ell}$ where $\ell = 1, \ldots, s$.

We let $\{\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(d)}\}$ be an ordered list of threshold vectors that are formulated based on the threshold values of each constraint, where d denotes the total number of threshold vectors that the decision maker is interested in testing, and $\boldsymbol{q}^{(\theta)}$ is the θth threshold vector in consideration, where $\theta = 1, \ldots, d$. We use $\boldsymbol{q}^{(d+1)}$ to denote a threshold vector that contains "$+\infty$" as the threshold value for each constraint so that the constrained R&S problem becomes an unconstrained problem. We assume that $\boldsymbol{q}^{(1)}$ is preferred to $\boldsymbol{q}^{(2)}$, $\boldsymbol{q}^{(2)}$ is preferred to $\boldsymbol{q}^{(3)}$, and so on. We let $q_\ell^{(\theta)}$ be the threshold value on constraint ℓ in $\boldsymbol{q}^{(\theta)}$, where $\theta = 1, \ldots, d+1$ and $\ell = 1, \ldots, s$. We then introduce the threshold index vector $\boldsymbol{I}^{(\theta)}$ to include the indices of the threshold values that form $\boldsymbol{q}^{(\theta)}$ where $\theta = 1, \ldots, d$. Similar to the definition of $q_\ell^{(\theta)}$, $I_\ell^{(\theta)}$ represents the threshold index on constraint ℓ. A decision maker can input the ordered list of threshold vectors, or the decision maker can input an ordered list of threshold values for each constraint and a mechanism for constructing an ordered list of threshold vectors from the inputed threshold values (see Chapter 3).

Consider the example of an inventory control policy discussed in Section 2.1. We have $s = 2$, $d_1 = 2$ (two threshold values for the first constraint), $d_2 = 3$ (three threshold values for the second constraint), $q_1^1 = 1, q_1^2 = 5$, and $q_2^1 = 100000, q_2^2 = 105000$, and $q_2^3 = 110000$. Moreover, we consider the following $d = 6$ ordered threshold vectors:

$$\boldsymbol{q}^{(1)} = \begin{bmatrix} 1 \\ 100000 \end{bmatrix}, \quad \boldsymbol{q}^{(2)} = \begin{bmatrix} 1 \\ 105000 \end{bmatrix}, \quad \boldsymbol{q}^{(3)} = \begin{bmatrix} 1 \\ 110000 \end{bmatrix}, \boldsymbol{q}^{(4)} = \begin{bmatrix} 5 \\ 100000 \end{bmatrix},$$

$$\boldsymbol{q}^{(5)} = \begin{bmatrix} 5 \\ 105000 \end{bmatrix}, \quad \boldsymbol{q}^{(6)} = \begin{bmatrix} 5 \\ 110000 \end{bmatrix}, \text{ and } \boldsymbol{q}^{(7)} = \begin{bmatrix} +\infty \\ +\infty \end{bmatrix}.$$

Note that $q_1^{(1)} = q_1^{(2)} = q_1^{(3)} = 1$ and $q_1^{(4)} = q_1^{(5)} = q_1^{(6)} = 5$ while $q_2^{(1)} = q_2^{(4)} = 100000$, $q_2^{(2)} = q_2^{(5)} = 105000$, and $q_2^{(3)} = q_2^{(6)} = 110000$. The threshold index vectors are

$$\boldsymbol{I}^{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{I}^{(2)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{I}^{(3)} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad \boldsymbol{I}^{(4)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{I}^{(5)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \text{ and } \boldsymbol{I}^{(6)} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Hence $I_1^{(1)} = I_1^{(2)} = I_1^{(3)} = 1$ and $I_1^{(4)} = I_1^{(5)} = I_1^{(6)} = 2$, while $I_2^{(1)} = I_2^{(4)} = 1$, $I_2^{(2)} = I_2^{(5)} = 2$, and $I_2^{(3)} = I_2^{(6)} = 3$.

Let $\mathbf{P}(\theta)$ for $\theta = 1, 2, \ldots, d+1$ represent the following problem:

$$\begin{array}{lll} \arg\max_{i \in \Gamma} & x_i & \\ \text{s.t.} & y_{i\ell} \leq q_\ell^{(\theta)} & \text{for all } \ell = 1, \ldots, s. \end{array}$$

Then our problem is to solve a series of problems $\mathbf{P}(\theta)$, $\theta = 1, 2, \ldots, d+1$, and return a set of solutions to them. Note that $\theta = d + 1$ is the case when the decision maker is willing to completely relax all the constraints and identify the best system of the unconstrained optimization problem.

4.2.2 Correct Selection

To solve a constrained R&S problem, we consider two phases, namely finding the feasible systems with respect to the secondary performance measure of $\boldsymbol{q}^{(\theta)}$ (Phase I) and selecting a system with the best primary performance measure based on comparisons among feasible systems (Phase II) where $\theta = 1, \ldots, d+1$.

To perform the feasibility check with respect to constraint ℓ, [1] introduce a tolerance level ϵ_ℓ. This is a positive real value predefined by the decision maker and is often interpreted as the amount that she is willing to be off from a given threshold value. Consider a threshold value q_ℓ^m for $m = 1, \ldots, d_\ell$. Any systems with $y_{i\ell} \leq q_\ell^m - \epsilon_\ell$ are considered as desirable systems with respect to constraint ℓ and threshold q_ℓ^m. Systems with $y_{i\ell} \geq q_\ell^m + \epsilon_\ell$ are considered as unacceptable systems for constraint ℓ and threshold q_ℓ^m. Systems that satisfy $q_\ell^m - \epsilon_\ell < y_{i\ell} < q_\ell^m + \epsilon_\ell$, which means their mean performances fall within the tolerance level of q_ℓ^m, are considered as acceptable systems. We use $D_\ell(q_\ell^m), A_\ell(q_\ell^m)$ and $U_\ell(q_\ell^m)$ to denote the sets of desirable systems, acceptable systems, and unacceptable systems, respectively. More specifically,

$$D_\ell(q_\ell^m) = \{i \in \Gamma | y_{i\ell} \leq q_\ell^m - \epsilon_\ell\};$$
$$U_\ell(q_\ell^m) = \{i \in \Gamma | y_{i\ell} \geq q_\ell^m - \epsilon_\ell\}; \text{ and}$$
$$A_\ell(q_\ell^m) = \{i \in \Gamma | q_\ell^m - \epsilon_\ell < y_{i\ell} < q_\ell^m + \epsilon_\ell\}.$$

When feasibility check is performed to completion (until a decision is made), we let $\mathrm{CD}_{i\ell}(q_\ell^m)$ denote the correct decision event of system i with respect to constraint ℓ and threshold q_ℓ^m from the feasibility check. This is defined as declaring system i feasible if $i \in D_\ell(q_\ell^m)$ and infeasible if $i \in U_\ell(q_\ell^m)$. Any feasibility decision is considered correct if $i \in A_\ell(q_\ell^m)$. For any threshold vector $\boldsymbol{q}^{(\theta)}$, we say that system i is desirable with respect to $\boldsymbol{q}^{(\theta)}$ when it is desirable with respect to all the constraints, i.e., $i \in D_\ell(q_\ell^{(\theta)})$ for all $\ell = 1, \ldots, s$. System i is unacceptable with respect to $\boldsymbol{q}^{(\theta)}$ if it is unacceptable with respect to at least one constraint, i.e., there exists ℓ such that $i \in U_\ell(q_\ell^{(\theta)})$. When system i is acceptable with respect to some (or all) the constraints and desirable with respect to the other constraints, system i is called acceptable with respect to $\boldsymbol{q}^{(\theta)}$.

To perform comparisons pairwisely between systems so that a system with the best primary performance measure can be identified, a decision maker needs to first specify an indifference-zone parameter δ. This is the smallest absolute difference that the decision maker considers significant, and any systems whose primary performance measure is at least δ smaller (larger) than system i is considered as inferior (superior) to system i.

For each threshold vector $\boldsymbol{q}^{(\theta)}$, where $\theta = 1, \ldots, d+1$, we let $S_{de}^{(\theta)}$ and $S_a^{(\theta)}$ denote the set of desirable and acceptable systems with respect to $\boldsymbol{q}^{(\theta)}$, respectively. That is,

$$S_{de}^{(\theta)} = \cap_{\ell=1}^{s} D_\ell(q_\ell^{(\theta)}) \quad \text{and} \quad S_a^{(\theta)} = \left(\cap_{\ell=1}^{s} \left(D_\ell(q_\ell^{(\theta)}) \cup A_\ell(q_\ell^{(\theta)})\right)\right) \setminus S_{de}^{(\theta)}.$$

Then we use $[b_\theta]$ to denote the index of the best system among the systems in $S_{de}^{(\theta)}$ so that $x_{[b_\theta]} \geq x_i$ for $i, [b_\theta] \in S_{de}^{(\theta)}$. If $S_{de}^{(\theta)} = \emptyset$, then define $[b_\theta]$ as "nothing" and $|\{[b_\theta]\}| = 0$ (where we use $|S|$ to denote the cardinality of a set S). Let $I_b = \left\{\theta \mid |S_{de}^{(\theta)} \cup S_a^{(\theta)}| \geq 2\right\}$ be the set of threshold vector indices that may require comparison among systems (as there can be more than one system declared feasible with respect to threshold vector $\boldsymbol{q}^{(\theta)}$). We use $\mathrm{CS}^{(\theta)}$ to denote the correct selection event with respect to $\boldsymbol{q}^{(\theta)}$, and define $\mathrm{CS}^{(\theta)}$ as either selecting a desirable or acceptable system with respect to $\boldsymbol{q}^{(\theta)}$ whose primary performance is not inferior to the best system $[b_\theta]$ when $[b_\theta]$ exists, or an acceptable system with respect to θ (if one exists) when $[b_\theta]$ does not exist (i.e., nothing). More specifically,

$$\mathrm{CS}^{(\theta)} = \Big\{\text{select } i \text{ such that } i \in S_{de}^{(\theta)} \cup S_a^{(\theta)} \text{ and } x_i > x_{[b_\theta]} - \delta \text{ if } |\{[b_\theta]\}| \neq 0$$
$$\text{or any } i \in S_a^{(\theta)} \text{ if } |\{[b_\theta]\}| = 0\Big\}.$$

We define the correct selection (CS) of our problem as selecting a portfolio of best systems $[b_\theta]$ such that $\mathrm{CS}^{(\theta)}$ is satisfied for all $\theta = 1, \ldots, d+1$. That is $\mathrm{CS} = \cap_{\theta=1}^{d+1} \mathrm{CS}^{(\theta)}$.

4.2.3 Notation and Assumptions

Throughout the paper, we use $\mathbb{1}(\cdot)$ to denote the indicator function and consider additional notation as follows:

$$n_0 \equiv \text{initial sample size for each system } (n_0 \geq 2);$$
$$r_i \equiv \text{number of observations so far for system } i \ (r_i \geq n_0);$$
$$\bar{X}_i(r_i) \equiv \text{average value of } X_{i1}, \ldots, X_{ir_i} \text{ for system } i;$$
$$\bar{Y}_{i\ell}(r_i) \equiv \text{average value of } Y_{i\ell 1}, \ldots, Y_{i\ell r_i} \text{ for system } i \text{ and constraint } \ell;$$
$$S_{X_{ij}}^2(n_0) \equiv \text{sample variance of } X_{i1} - X_{j1}, \ldots, X_{in_0} - X_{jn_0} \text{ between system } i \text{ and } j;$$

$$S^2_{Y_{i\ell}}(n_0) \equiv \text{sample variance of } Y_{i\ell 1}, \ldots, Y_{i\ell n_0} \text{ for system } i \text{ and constraint } \ell;$$

$$R(r_i; v, w, z) \equiv \max\left\{0, \frac{(n_0-1)wz}{v} - \frac{v}{2c}r_i\right\} \text{ for } v, w, z \in \mathbb{R}^+ \text{ and } c \in \{1, 2, \ldots\};$$

$$g(\eta) \equiv \sum_{j=1}^{c}(-1)^{j+1}\left(1 - \frac{1}{2}\mathbb{1}(j=c)\right) \times \left(1 + \frac{2\eta(2c-j)j}{c}\right)^{-(n_0-1)/2};$$

$$\alpha \equiv \text{overall nominal error for a procedure under consideration.}$$

Note that a parameter c is required for both $R(r_i; v, w, z)$ and $g(\eta)$. This is a user-defined parameter that impacts the shape of the continuation region defined by $(-R(r_i; v, w, z), R(r_i; v, w, z))$ (it becomes a longer and narrower triangle as c increases). The choice $c = 1$ is recommended in $g(\eta)$ function as it guarantees a unique and easy solution when computing the implementation parameter η from $g(\eta)$. [10] also suggest that $c = 1$ is a good choice when the decision maker does not have information about the systems' mean configuration. The experimental results for our proposed procedure are based on $c = 1$.

We then discuss the necessary assumptions to guarantee the statistical validity of our proposed procedures.

Assumption 4. *For each system i, where $i = 1, \ldots, k$, we have*

$$\begin{bmatrix} X_{in} \\ Y_{i1n} \\ \vdots \\ Y_{isn} \end{bmatrix} \overset{iid}{\sim} N_{s+1}\left(\begin{bmatrix} x_i \\ y_{i1} \\ \vdots \\ y_{is} \end{bmatrix}, \Sigma_i\right), \qquad n = 1, 2, \ldots$$

where $\overset{iid}{\sim}$ denotes independent and identically distributed, N_{s+1} denotes $(s+1)$-dimensional multivariate normal, and Σ_i is the $(s+1) \times (s+1)$ covariance matrix of the vector $(X_{in}, Y_{i1n}, \ldots, Y_{isn})$.

Normally distributed data is a common assumption used in many R&S procedures due to the fact that it can be justified by the Central Limit Theorem when observations are either within-replication averages or batch means [12]. Moreover, primary and secondary performance measures are usually correlated. When common random numbers (CRN) are introduced in simulating observations from each system, observations between systems are correlated. Our formulation allows correlations between both performance measures and systems.

Assumption 5. *For any* $\theta = 1, \ldots, d+1$ *and system* $i \in S_{de}^{(\theta)} \cup S_{a}^{(\theta)}$ *and* $i \neq [b_\theta]$, *we assume* $x_i \leq x_{[b_\theta]} - \delta$.

Assumption 5 implies that there exists only one best system $[b_\theta]$ with respect to $\boldsymbol{q}^{(\theta)}$, and any systems that are desirable or acceptable to $q_\ell^{(\theta)}$ for all constraints $\ell = 1, \ldots, s$ are inferior to system $[b_\theta]$, where $\theta = 1, \ldots, d+1$. This assumption is a standard assumption for proving the statistical validity of IZ approaches in the R&S literature.

4.3 Sequentially-Running Procedures

In this section, we first propose a procedure ($\mathcal{FAP}^{\mathcal{R}}$) for finding a portfolio $\{[b_\theta] : \theta = 1, \ldots, d+1\}$ in Section 4.3.1. $\mathcal{FAP}^{\mathcal{R}}$ incorporates restart for different threshold vectors and runs Phases I and II sequentially. We prove the statistical validity of $\mathcal{FAP}^{\mathcal{R}}$ in Section 4.3.2. A more efficient sequentially-running procedure $\mathcal{FAP}$ is provided in Section 4.3.3.

4.3.1 Procedure $\mathcal{FAP}^{\mathcal{R}}$

In this section, we discuss the sequentially-running procedure $\mathcal{FAP}^{\mathcal{R}}$ and also provide a detailed algorithm. We first introduce some sets that we use in the proposed procedure:

- M is a set of systems whose feasibility are not yet determined with respect to all threshold vector $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(d+1)}$, i.e., there exist at least one threshold vector $\boldsymbol{q}^{(\theta)}$ and one constraint ℓ such that the systems are not decided feasible or infeasible with respect to $q_\ell^{(\theta)}$.

- $\mathcal{F}_\theta$ is a set of systems that are feasible with respect to threshold vector $\boldsymbol{q}^{(\theta)}$, where $\theta = 1, \ldots, d+1$. Initially, $\mathcal{F}_\theta$ is an empty set for $\theta = 1, 2, \ldots, d$ and $\mathcal{F}_{d+1}$ contains all systems in consideration (i.e., $\mathcal{F}_{d+1} = \Gamma$).

The sequentially-running procedure $\mathcal{FAP}^{\mathcal{R}}$ performs Phase I first by declaring feasibility decisions for each system with respect to all thresholds on all constraints. Our approach builds on the work discussed in Chapter 2, where an efficient fully-sequential procedure for checking the feasibility of all systems with respect to all constraints and all thresholds simultaneously is proposed and analyzed. Specifically, our proposed procedure incorporates procedure $\mathcal{RF}$ discussed in Chapter 2 by returning $Z_{i\ell}^m = 1 (Z_{i\ell}^m = 0)$ if system i is declared feasible (infeasible) with respect to constraint ℓ and threshold q_ℓ^m. If a system's feasibility check decisions are completed with respect to all threshold values, it is added to $\mathcal{F}_\theta$ only when it is feasible to $\boldsymbol{q}^{(\theta)}$ for $\theta = 1, \ldots, d$. Note that there is no elimination of inferior systems in Phase I. Theorem 1 in Chapter 2 shows that once a system i is declared feasible with respect to a threshold q_ℓ^m such that $q_\ell^m \geq y_{i\ell} + \epsilon_\ell$, this system will be declared feasible with respect to all thresholds $q_\ell^{m+1}, \ldots, q_\ell^{d_\ell}$ on constraint ℓ. Similarly, if a system i is declared infeasible with respect to a threshold q_ℓ^m such that $q_\ell^m \leq y_{i\ell} - \epsilon_\ell$, then this system will be declared infeasible with respect to all the thresholds $q_\ell^1, \ldots, q_\ell^{m-1}$. Our proposed procedures utilize the same idea.

During Phase II, we compare systems pairwisely and find a best system among the systems in each set $\mathcal{F}_\theta$ where $\theta = 1, \ldots, d+1$. Note that we do not collect observations from systems in $\mathcal{F}_\theta$ if $|\mathcal{F}_\theta| = 1$ as this system would already be selected as the best system with respect to $\boldsymbol{q}^{(\theta)}$ (i.e., we only perform pairwise comparison among systems in $\cup_{\theta=1}^{d+1} \mathcal{F}_\theta$ such that $|\mathcal{F}_\theta| > 1$). Elimination occurs only between systems in the same set $\mathcal{F}_\theta$ and continues until each $\mathcal{F}_\theta$ contains at most one system. Then we report those surviving systems as a portfolio of best systems. In order to prove the statistical validity of our proposed sequentially-running procedure, we avoid the correlation between the primary and the secondary performance measures by not recycling observations collected from Phase I and starting "from scratch" when performing Phase II. A detailed description of the algorithm $\mathcal{FAP}^{\mathcal{R}}$ is provided in Algorithm 9.

4.3.2 Statistical Validity of Procedure $\mathcal{FAP}^{\mathcal{R}}$

In this section, we prove the statistical validity of $\mathcal{FAP}^{\mathcal{R}}$ presented in Algorithm 9. Before presenting the main results, we first introduce the following sets.

$$B^{(\theta)} = \text{ set of best systems with respect to } \boldsymbol{q}^{(\theta)} = \{[b_\theta]\}\,;$$

$$S_u^{(\theta)} = \text{ set of unacceptable systems with respect to } \boldsymbol{q}^{(\theta)} \text{ for } \theta = 1, \ldots, d+1.$$

Algorithm 9 $\mathcal{FAP}^{\mathcal{R}}$

[**Setup**:] Select the overall nominal confidence level $1-\alpha$ and choose $\alpha_f, \alpha_c > 0$ such that $(1-\alpha_f)(1-\alpha_c) = 1-\alpha$. Choose tolerance levels $\epsilon_1, \ldots, \epsilon_s$, indifference-zone parameter δ, threshold vectors $\{\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(d)}\}$, and associated index vectors $\{\boldsymbol{I}^{(1)}, \boldsymbol{I}^{(2)}, \ldots, \boldsymbol{I}^{(d)}\}$. Set $M = \Gamma$ and $Z_{i\ell}^m = 2$ for all $i \in M, \ell = 1, \ldots, s$, and $m = 1, \ldots, d_\ell$. Set $\mathcal{F}_\theta = \emptyset$ for $\theta = 1, \ldots, d$ and $\mathcal{F}_{d+1} = \Gamma$. Set η_f such that $g(\eta_f) = \alpha'_f$, where α'_f is set as the solution to

$$(1-\min\{s,d\}\alpha'_f)^{k-\min\{d+1,k\}}(1-2s\alpha'_f)^{\min\{d+1,k\}} = 1-\alpha_f, \text{ if systems are simulated independently;}$$

and set as

$$\alpha'_f = \alpha_f / \left[k \times \min\{s,d\} + (2s - \min\{s,d\}) \times \min\{d+1,k\}\right], \text{ if systems are simulated under CRN.}$$

[**Initialization for Phase I**:]
for each system $i \in M$ **do**

- Obtain n_0 observations $Y_{i\ell 1}, Y_{i\ell 2}, \ldots, Y_{i\ell n_0}$ for $\ell = 1, 2, \ldots, s$.
- Compute $\bar{Y}_{i\ell}(n_0)$ and $S^2_{Y_{i\ell}}(n_0)$.
- Set $r_i = n_0, \mathrm{ON}_i = \{1, 2, \ldots, s\}$, and $\mathrm{ON}_{i\ell} = \{1, 2, \ldots, d_\ell\}$ for $\ell = 1, 2, \ldots, s$.
- Set $v_{i\ell}^{\mathrm{UB}} = \infty$ and $v_{i\ell}^{\mathrm{LB}} = -\infty$ for $\ell = 1, \ldots, s$.

end for
[**Feasibility Check**:]
for each system $i \in M$ **do**
  for $\ell \in \mathrm{ON}_i$ **do**
    $v_{i\ell}^{\mathrm{UB}} = \min(v_{i\ell}^{\mathrm{UB}}, \bar{Y}_{i\ell}(r_i) + R(r_i; \epsilon_\ell, \eta_f, S^2_{Y_{i\ell}}(n_0))/r_i)$.
    $v_{i\ell}^{\mathrm{LB}} = \max(v_{i\ell}^{\mathrm{LB}}, \bar{Y}_{i\ell}(r_i) - R(r_i; \epsilon_\ell, \eta_f, S^2_{Y_{i\ell}}(n_0))/r_i)$.
    for $m \in \mathrm{ON}_{i\ell}$ **do**,
      If $v_{i\ell}^{\mathrm{UB}} \leq q_\ell^m$, set $Z_{i\ell}^m = 1$ and $\mathrm{ON}_{i\ell} = \mathrm{ON}_{i\ell} \setminus \{m\}$.
      If $v_{i\ell}^{\mathrm{LB}} \geq q_\ell^m$, set $Z_{i\ell}^m = 0$ and $\mathrm{ON}_{i\ell} = \mathrm{ON}_{i\ell} \setminus \{m\}$.
    end for
    If $\mathrm{ON}_{i\ell} = \emptyset$, set $\mathrm{ON}_i = \mathrm{ON}_i \setminus \{\ell\}$.
  end for
  If $\mathrm{ON}_i = \emptyset$, then remove system i from M and, for $\theta = 1, \ldots, d$, add system i to $\mathcal{F}_\theta$ if $\prod_{\ell=1}^s Z_{i\ell}^{I_\ell^{(\theta)}} = 1$.
end for
[**Stopping Condition for Phase I**:] If $|M| = 0$, then go to [**Initialization for Phase II**]. Otherwise, for each $i \in M$, set $r_i = r_i + 1$, take one additional observation $Y_{i\ell r_i}$, update $\bar{Y}_{i\ell}(r_i)$ for $\ell \in \mathrm{ON}$ and go to [**Feasibility Check**].
[**Initialization for Phase II**:] Let η_c be a solution to $g(\eta_c) = \alpha'_c$, where

$$\alpha'_c = \begin{cases} 1 - (1 - \alpha_c/(d+1))^{1/(k-1)}, & \text{if systems are simulated independently;} \\ \alpha_c/[(d+1)(k-1)], & \text{if systems are simulated under CRN.} \end{cases}$$

Let $T = \{\theta = 1, \ldots, d+1 : |\mathcal{F}_\theta| \geq 2|\}$ and $M = \cup_{\theta \in T} \mathcal{F}_\theta$ be the set of systems that still need comparison. For each system $i \in M$, perform an entirely new simulation and obtain $X_{i1}, \ldots, X_{in_0}$ independent of any $Y_{i\ell n}$ generated in Phase I. Compute $\bar{X}_i(n_0)$ and $S^2_{X_{ij}}(n_0)$ for $i, j \in M$ and $i \neq j$. Set $r = n_0$ and go to [**Comparison**].
[**Comparison**:]
for $i, j \in M$ s.t. $i \neq j$ and $i, j \in \mathcal{F}_\theta$ for any $\theta \in T$, if

$$r\bar{X}_i(r) > r\bar{X}_j(r) + R(r; \delta, \eta_c, S^2_{X_{ij}}(n_0)), \tag{9}$$

do
  Eliminate j from $\mathcal{F}_{d+1}$, and
  for $\theta \in T \setminus \{d+1\}$ **do**
    If $i \in \mathcal{F}_\theta$ and $j \in \mathcal{F}_\theta$, then eliminate j from $\mathcal{F}_\theta$. If $|\mathcal{F}_\theta| = 1$, eliminate θ from T.
  end for
  If $|T| \geq 1$ and $j \notin \mathcal{F}_\theta$ for all $\theta \in T$, then eliminate j from M. Also, eliminate i from M if $i \notin \mathcal{F}_\theta$ for all $\theta \in T$.
end for
[**Stopping Condition for Phase II**:]
If $|T| = 0$, stop and return the set of systems in $\mathcal{F}_\theta, \theta = 1, \ldots, d+1$ as the portfolio of best systems. Otherwise, for all $i \in M$, take one additional observation $X_{i,r+1}$ and compute $\bar{X}_i(r+1)$. Then set $r = r+1$ and go to [**Comparison**].

Recall that $[b_\theta]$ can be 'nothing' and thus $B^{(\theta)}$ can be an empty set. Also, we define $S_{de}^{(\theta)}$ and $S_a^{(\theta)}$ in Section 4.2.2 as

$$\begin{aligned} S_{de}^{(\theta)} &= \text{ set of desirable systems with respect to } \boldsymbol{q}^{(\theta)} \text{ for } \theta = 1, \ldots, d+1; \\ S_a^{(\theta)} &= \text{ set of acceptable systems with respect to } \boldsymbol{q}^{(\theta)} \text{ for } \theta = 1, \ldots, d+1; \\ I_b &= \text{ set of threshold vector indices such that } |S_{de}^{(\theta)} \cup S_a^{(\theta)}| \geq 2. \end{aligned}$$

Notice that $S_{de}^{(d+1)} = \Gamma$ while $S_u^{(d+1)} = S_a^{(d+1)} = \emptyset$, and $|S_{de}^{(\theta)}| + |S_a^{(\theta)}| + |S_u^{(\theta)}| = k$ for any $\theta = 1, \ldots, d+1$. We then let B denote the set of all the best systems with respect to $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(d+1)}$, i.e., $B = \cup_{\theta=1}^{d+1} B^{(\theta)}$. We let S_u denote the set of systems that are unacceptable with respect to at least one $\boldsymbol{q}^{(\theta)}$ among $\Gamma \setminus B$, i.e., $S_u = \cup_{\theta=1}^{d} S_u^{(\theta)} \setminus B$, let S_{de} denote the set of desirable systems with respect to all $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(d)}$ among $\Gamma \setminus B$, i.e., $S_{de} = \cap_{\theta=1}^{d} S_{de}^{(\theta)} \setminus B$, and let S_a denote the set of systems in $\Gamma \setminus B$ that are acceptable with respect to at least one $\boldsymbol{q}^{(\theta)}$ but desirable with respect to the other threshold vectors, i.e., $S_a = \cup_{\theta=1}^{d} S_a^{(\theta)} \setminus (B \cup S_u)$. We further let j_b denote the number of best systems, i.e., $j_b = |B|$, and let $j_u = |S_u|, j_d = |S_{de}|$, and $j_a = |S_a|$. Note that $j_b + j_u + j_d + j_a = k$ and $1 \leq j_b \leq \min\{k, d+1\}$.

To further illustrate the above notation, we consider an example when two constraints are in presence, where the first constraint has two thresholds and the second constraint has three thresholds. Suppose that the threshold vectors are ordered based on ranked constraints. That is, there are $d = 7$ possible threshold vectors:

$$\boldsymbol{q}^{(1)} = \begin{bmatrix} q_1^1 \\ q_2^1 \end{bmatrix}, \boldsymbol{q}^{(2)} = \begin{bmatrix} q_1^1 \\ q_2^2 \end{bmatrix}, \boldsymbol{q}^{(3)} = \begin{bmatrix} q_1^1 \\ q_2^3 \end{bmatrix}, \boldsymbol{q}^{(4)} = \begin{bmatrix} q_1^2 \\ q_2^1 \end{bmatrix}, \boldsymbol{q}^{(5)} = \begin{bmatrix} q_1^2 \\ q_2^2 \end{bmatrix}, \boldsymbol{q}^{(6)} = \begin{bmatrix} q_1^2 \\ q_2^3 \end{bmatrix}, \boldsymbol{q}^{(7)} = \begin{bmatrix} +\infty \\ +\infty \end{bmatrix}.$$

Figure 17 shows (non-negative) secondary performance means where the shaded areas represent acceptable regions with respect to one or more threshold vectors.

We see that system a is desirable with respect to all threshold vectors, i.e., $a \in S_{de}^{(\theta)}$ for all $\theta = 1, \ldots, 7$. System b is acceptable with respect to $\boldsymbol{q}^{(2)}, \boldsymbol{q}^{(3)}$, and $\boldsymbol{q}^{(5)}$, desirable with respect to $\boldsymbol{q}^{(6)}$ and $\boldsymbol{q}^{(7)}$, and unacceptable with respect to $\boldsymbol{q}^{(1)}$ and $\boldsymbol{q}^{(4)}$. Therefore, we know that $b \in S_a^{(\theta)}$ where $\theta = 2, 3, 5$, $b \in S_{de}^{(\theta)}$ where $\theta = 6, 7$, and $b \in S_u^{(\theta)}$ where $\theta = 1, 4$. Similarly, system c is acceptable with respect to $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{q}^{(3)}$, and $\boldsymbol{q}^{(4)}$, meaning that $c \in S_a^{(\theta)}$ where

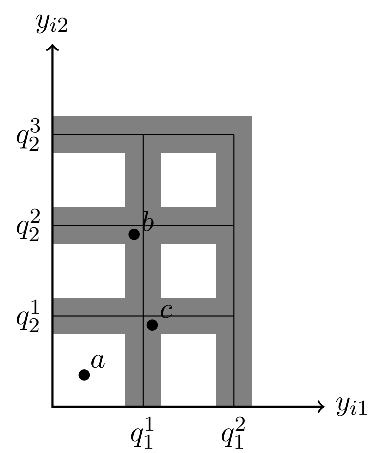


Figure 17: Two secondary performance measures and seven threshold vectors

$\theta = 1, 2, 3, 4$, and desirable with respect to all the other threshold vectors, meaning that $c \in S_{de}^{(\theta)}$ where $\theta = 5, 6, 7$. We also know that $a \in S_{de}$ (if $a \neq [b_1]$) as it is desirable with respect to all threshold vectors, $b \in S_u$ (if $b \neq [b_6]$) as it is unacceptable with respect to at least one threshold vectors, and $c \in S_a$ (if $c \neq [b_5]$) as it is acceptable with respect to some threshold vectors and desirable with respect to the other threshold vectors. As system a is desirable with respect to all threshold vectors, system b is acceptable to $\boldsymbol{q}^{(2)}, \boldsymbol{q}^{(3)}, \boldsymbol{q}^{(5)}$ and desirable with respect to $\boldsymbol{q}^{(6)}, \boldsymbol{q}^{(7)}$, and system c is acceptable to $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{q}^{(3)}.\boldsymbol{q}^{(4)}$ and desirable to $\boldsymbol{q}^{(5)}, \boldsymbol{q}^{(6)}, \boldsymbol{q}^{(7)}$, we know that all threshold vectors have at least two desirable or acceptable systems. Therefore, we have $I_b = \{1, 2, 3, 4, 5, 6, 7\}$.

To prove the statistical validity of $\mathcal{FAP}^{\mathcal{R}}$, we start with the following lemma.

Lemma 7. *For system i and constraint ℓ with specific threshold value q_ℓ^m, the* [**Feasibility Check**] *steps in $\mathcal{FAP}^{\mathcal{R}}$ ensure* $\Pr(\mathrm{CD}_{i\ell}(q_\ell^m)) \geq 1 - \alpha_f'$.

Proof. When system i and constraint ℓ with specific threshold q_ℓ^m are considered separately, the [**Feasibility Check**] steps in $\mathcal{FAP}^{\mathcal{R}}$ are essentially the same as for the statistically-valid feasibility check procedure $\mathcal{F}$ in [1] for a single system and a single constraint with one threshold value with confidence level $1 - \alpha_f'$. The result now follows from the special case of Theorem 1 in [1] with $k = 1$. □

We then let $\Theta_u(i)$ and $\Theta_d(i)$ denote the set of indices of the threshold vectors that system i is unacceptable and desirable with respect to, respectively. More specifically, we let

$$\Theta_u(i) = \left\{\theta = 1, \dots d \,\middle|\, \text{there exists a constraint } \ell \text{ such that } i \in U_\ell(q_\ell^{(\theta)})\right\};$$
$$\Theta_d(i) = \left\{\theta = 1, \dots d \,\middle|\, i \in \cap_{\ell=1}^{s} D_\ell(q_\ell^{(\theta)})\right\}.$$

We introduce two events.

$$\mathcal{A}^*(i) = \text{system } i \text{ is declared infeasible to all } \boldsymbol{q}^{(\theta)} \text{ such that } \theta \in \Theta_u(i);$$
$$\mathcal{B}_1(i) = \text{system } i \text{ is declared feasible to all } \boldsymbol{q}^{(\theta)} \text{ such that } \theta \in \Theta_d(i) \text{ and}$$
$$\text{declared infeasible to all } \boldsymbol{q}^{(\theta)} \text{ such that } \theta \in \Theta_u(i).$$

Lemma 8. *For any system $i \in S_u$, the* [**Feasibility Check**] *steps in $\mathcal{FAP}^{\mathcal{R}}$ ensure*

$$\Pr\left(\mathcal{A}^*(i)\right) \geq 1 - \min\{s, d\}\alpha'_f.$$

Proof. For a particular system i, we know that for each $\theta \in \Theta_u(i)$, there exists at least one constraint ℓ_θ such that $y_{i\ell_\theta} \geq q_{\ell_\theta}^{(\theta)} + \epsilon_{\ell_\theta}$. Then we have

$$\Pr\left(\mathcal{A}^*(i)\right) \geq \Pr\left(\cap_{\theta=1}^{d} \mathrm{CD}_{i\ell_\theta}(q_{\ell_\theta}^{(\theta)})\right) \geq 1 - \sum_{\theta=1}^{d} \Pr\left(\mathrm{ICD}_{i\ell_\theta}(q_{\ell_\theta}^{(\theta)})\right) \geq 1 - d\alpha'_f, \tag{10}$$

where we use $\mathrm{ICD}_{i\ell}(q_\ell^m)$ to denote the incorrect decision event of system i with respect to constraint ℓ and threshold q_ℓ^m. The first inequality holds because declaring system i infeasible to constraint ℓ_θ is sufficient to declare system i infeasible to threshold vector $\boldsymbol{q}^{(\theta)}$ and it is not possible to declare a system feasible with respect to a threshold vector without completing the feasibility check with all thresholds in that vector. The second inequality holds due to the Bonferroni inequality, and the last inequality holds due to Lemma 7.

Observe that since there are only s constraints, the set $L = \{\ell_\theta | \theta \in \Theta_u(i)\}$ can have at most s distinct values. For $\ell \in L$, let $I_{i\ell}$ denote the largest threshold index on constraint ℓ that system i is unacceptable to, i.e.,

$$I_{i\ell} = \max_{1 \leq m \leq d_\ell} \{m : y_{i\ell} \geq q_\ell^m + \epsilon_\ell\}.$$

Thus, we know that $q_\ell^1 < q_\ell^2 < \cdots < q_\ell^{I_{i\ell}} \leq y_{i\ell} - \epsilon_\ell$ on constraint ℓ. Due to the discussion in Chapter 2, we know that $\mathrm{CD}_{i\ell}(q_\ell^{I_{i\ell}}) \subseteq \cdots \subseteq \mathrm{CD}_{i\ell}(q_\ell^2) \subseteq \mathrm{CD}_{i\ell}(q_\ell^1)$. Then

$\mathrm{CD}_{i\ell}(q_\ell^{I_{i\ell}}) \subseteq \mathrm{CD}_{i\ell}(q_\ell^{(\theta)})$ for $\theta \in \Theta_u(i)$ with $\ell = \ell_\theta$. Thus, we also have

$$\begin{aligned}\Pr\left(\mathcal{A}_1^*(i)\right) &\geq \Pr\left(\cap_{\theta=1}^d \mathrm{CD}_{i\ell_\theta}(q_{\ell_\theta}^{(\theta)})\right) \geq \Pr\left(\cap_{\ell\in L}\mathrm{CD}_{i\ell}(q_\ell^{I_{i\ell}})\right)\\ &\geq 1 - \sum_{\ell\in L}\Pr\left(\mathrm{ICD}_{i\ell}(q_\ell^{I_{i\ell}})\right) \geq 1 - |L|\alpha_f' \geq 1 - s\alpha_f', \end{aligned} \tag{11}$$

where the third inequality is due to the Bonferroni inequality and the forth inequalty is due to Lemma 7. By comparing equations (10) and (11), we conclude that $\Pr\left(\mathcal{A}_1^*(i)\right) \geq 1 - \min\{s,d\}\alpha_f'$. □

Lemma 9. *For any system $i \in B$, the* [**Feasibility Check**] *steps in $\mathcal{FAP}^{\mathcal{R}}$ ensure*

$$\Pr\left(\mathcal{B}_1(i)\right) \geq 1 - 2s\alpha_f'.$$

Proof. For a particular system $i \in B$, to ensure $\mathcal{B}_1(i)$, we need to ensure correct decisions with respect to all threshold vectors $\theta \in \Theta_d(i) \cup \Theta_u(i)$. Correct decisions with respect to all constraints and all the threshold values (i.e., $\cap_{\ell=1}^s \cap_{m=1}^{d_\ell} \mathrm{CD}_{i\ell}(q_\ell^m)$) will ensure that this event occurs. As the [**Feasibility Check**] steps in $\mathcal{FAP}^{\mathcal{R}}$ are the same as the satistical-valid feasibility check procedure $\mathcal{RF}$ provided in Algorithm 1, we know that

$$\begin{aligned}\Pr\left(\cap_{m=1}^{d_\ell}\mathrm{CD}_{i\ell}(q_\ell^m)\right) &\geq \Pr\left(\mathrm{CD}_{i\ell}(y_{i\ell}-\epsilon_\ell), \mathrm{CD}_{i\ell}(y_{i\ell}+\epsilon_\ell)\right)\\ &\geq \Pr\left(\mathrm{CD}_{i\ell}(y_{i\ell}-\epsilon_\ell)\right) + \Pr\left(\mathrm{CD}_{i\ell}(y_{i\ell}+\epsilon_\ell)\right) - 1\\ &\geq 1 - \alpha_f' + 1 - \alpha_f' - 1 = 1 - 2\alpha_f',\end{aligned}$$

where $\Pr\left(\mathrm{CD}_{i\ell}(y_{i\ell}-\epsilon_\ell), \mathrm{CD}_{i\ell}(y_{i\ell}+\epsilon_\ell)\right)$ denotes the joint probability of the events $\mathrm{CD}_{i\ell}(y_{i\ell}-\epsilon_\ell)$ and $\mathrm{CD}_{i\ell}(y_{i\ell}+\epsilon_\ell)$ and the first inequality holds due to Theorem 1 in Chapter 2. The last inequality holds due to Lemma 7. Therefore,

$$\Pr\left(\mathcal{B}_1(i)\right) \geq \Pr\left(\cap_{\ell=1}^s \cap_{m=1}^{d_\ell} \mathrm{CD}_{i\ell}(q_\ell^m)\right) \geq 1 - \sum_{\ell=1}^s \Pr\left(\left(\cap_{m=1}^{d_\ell}\mathrm{CD}_{i\ell}(q_\ell^m)\right)^c\right) \geq 1 - s(2\alpha_f'),$$

where we use A^c to denote the complement event of evnet A. □

We let $\mathrm{CS}_i^{(\theta)}$ denote the correct selection between system $i \in (S_a^{(\theta)} \cup S_{de}^{(\theta)}) \setminus \{[b_\theta]\}$ and the best system $[b_\theta]$ and introduce the following lemma.

Lemma 10. *Given an existing best desirable system $[b_\theta]$ and i such that $x_i \le x_{[b_\theta]} - \delta$, the* **[Comparison]** *steps for system i and $[b_\theta]$ in $\mathcal{FAP}^{\mathcal{R}}$ run to completion ensure*

$$\Pr\left(\mathrm{CS}_i^{(\theta)}\right) \ge 1 - \alpha'_c.$$

Proof. Given an existing best system $[b_\theta]$ (i.e., $[b_\theta] \neq$ 'nothing'), when only system i and $[b_\theta]$ are considered, the **[Comparison]** steps in Restart$^{\mathcal{S}}$ are the same as the statistically-valid selection-of-the-best procedure provided in [10] when two systems are considered with confidence level $1 - \alpha'_c$. Therefore, the result follows from the special case of Theorem 1 of [10] with $k = 2$. □

Theorem 6. *Under Assumptions 4 and 5, $\mathcal{FAP}^{\mathcal{R}}$ guarantees*

$$\Pr\{\mathrm{CS}\} \ge 1 - \alpha.$$

Proof. For correct selection, we need correct feasibility decision for any systems in S_u and B in Phase I and each $[b_\theta]$ for $\theta \in I_b$ needs to be selected as the best in Phase II. More specifically, we consider the following events.

$$\begin{aligned}
\mathcal{A}^* &= \text{each system } i \in S_u \text{ is declared infeasible with respect to any } \boldsymbol{q}^{(\theta)} \text{ such that } \theta \in \Theta_u(i)\\
&= \cap_{i \in S_u} \mathcal{A}^*(i);\\
\mathcal{B}_1^* &= \text{each system } i \in B \text{ is declared feasible with respect to } \boldsymbol{q}^{(\theta)} \text{ such that } \theta \in \Theta_d(i) \text{ and}\\
&\qquad \text{infeasible with respect to } \boldsymbol{q}^{(\theta)} \text{ such that } \theta \in \Theta_u(i)\\
&= \cap_{i \in B} \mathcal{B}_1(i);\\
\mathcal{B}_2^+ &= \text{ for each } \theta \text{ such that } \theta \in T,\ [b_\theta] \text{ is declared as the best system among the systems in } \mathcal{F}_\theta\\
&\subseteq \cap_{\theta \in T} \cap_{i \in \mathcal{F}_\theta, i \neq [b_\theta]} \mathrm{CS}_i^{(\theta)};\\
\mathcal{B}_2^* &= \text{ for each } \theta \in I_b,\ [b_\theta] \text{ is declared as the best system among the systems in } (S_a^{(\theta)} \cup S_{de}^{(\theta)})\\
&\subseteq \cap_{\theta \in I_b} \cap_{i \in (S_a^{(\theta)} \cup S_{de}^{(\theta)}), i \neq [b_\theta]} \mathrm{CS}_i^{(\theta)}.
\end{aligned}$$

Recall $T = \{\theta : |\mathcal{F}_\theta| \ge 2\}$. Then

$$\begin{aligned}
\Pr\{\mathrm{CS}\} &\ge \Pr(\mathcal{A}^* \cap \mathcal{B}_1^* \cap \mathcal{B}_2^+)\\
&\ge \Pr(\mathcal{A}^* \cap \mathcal{B}_1^* \cap \mathcal{B}_2^*)\\
&= \Pr\left(\mathcal{A}^* \cap \mathcal{B}_1^*\right) \Pr\left(\mathcal{B}_2^*\right),
\end{aligned}$$

where the second inequality holds because when event $\mathcal{A}^* \cap \mathcal{B}_1^*$ occurs, we have $T \subset I_b$ and $\mathcal{F}_\theta \subset (S_a^{(\theta)} \cup S_{de}^{(\theta)})$ for any θ, and the final inequality holds due to the independence of Phases I and II.

We also see that $\mathcal{A}^*$ and $\mathcal{B}_1^*$ are independent events when systems are simulated independently but are dependent events when systems are simulated under CRN. We have

$$\Pr\{\mathrm{CS}\} \geq \Pr(\mathcal{A}^* \cap \mathcal{B}_1^*)\Pr(\mathcal{B}_2^*)$$
$$= \begin{cases} \Pr(\mathcal{A}^*) \times \Pr(\mathcal{B}_1^*) \times \Pr(\mathcal{B}_2^*), & \text{if systems are simulated independently;} \\ [\Pr(\mathcal{A}^*) + \Pr(\mathcal{B}_1^*) - 1] \times \Pr(\mathcal{B}_2^*), & \text{if systems are simulated under CRN.} \end{cases}$$

We discuss the cases depending on whether the systems are simulated independently or under CRN.

When systems are simulated independently, due to Lemmas 8 and 9, we have

$$\Pr(\mathcal{A}^*) \geq (1 - \min\{s,d\}\alpha'_f)^{j_u} = (1 - \min\{s,d\}\alpha'_f)^{k-j_b-j_d-j_a};$$
$$\Pr(\mathcal{B}_1^*) \geq (1 - 2s\alpha'_f)^{j_b}.$$

Let N_{ij} denote the number of observations taken for system i before a comparison decision is made between systems i and j, and let N_i denote the maximum number of observations that system i takes within Phase II. That is

$$N_{ij} = \left\lceil \frac{2c\eta_c(n_0-1)S^2_{X_{ij}}(n_0)}{\delta^2} \right\rceil, \text{ and } N_i = \max_{j \neq i} N_{ij}.$$

Notice that as we use the same observations from each system to perform comparison among systems within each set $\mathcal{F}_\theta$ where $\theta = 1, \ldots, d+1$. As one system may appear in multiple sets $\mathcal{F}_\theta$, the comparison between systems in Phase II is depedent. Then we have

$$\Pr(\mathcal{B}_2^*) \geq \Pr\left(\cap_{\theta\in I_b} \cap_{i\in(S_a^{(\theta)}\cup S_{de}^{(\theta)}), i\neq[b_\theta]} \mathrm{CS}_i^{(\theta)}\right) \geq \sum_{\theta\in I_b} \Pr\left(\cap_{i\in(S_a^{(\theta)}\cup S_{de}^{(\theta)}), i\neq[b_\theta]} \mathrm{CS}_i^{(\theta)}\right) - (|I_b|-1)$$
$$= \sum_{\theta\in I_b} \mathrm{E}\left[\Pr\left\{\cap_{i\in(S_a^{(\theta)}\cup S_{de}^{(\theta)}), i\neq[b_\theta]} \mathrm{CS}_i^{(\theta)} \middle| X_{[b_\theta]1}, \ldots, X_{[b_\theta],N_{[b_\theta]}}, S^2_{X_{i[b_\theta]}}(n_0)\right\}\right] - (|I_b|-1)$$
$$= \sum_{\theta\in I_b} \mathrm{E}\left[\prod_{i\in(S_a^{(\theta)}\cup S_{de}^{(\theta)}), i\neq[b_\theta]} \Pr\left\{\mathrm{CS}_i^{(\theta)} \middle| X_{[b_\theta]1}, \ldots, X_{[b_\theta],N_{[b_\theta]}}, S^2_{X_{i[b_\theta]}}(n_0)\right\}\right] - (|I_b|-1)$$

$$
\begin{aligned}
&\geq \sum_{\theta\in I_b} \prod_{i\in(S_a^{(\theta)}\cup S_{de}^{(\theta)}), i\neq[b_\theta]} \mathrm{E}\left[\Pr\left\{\mathrm{CS}_i^{(\theta)} \middle| X_{[b_\theta]1},\ldots,X_{[b_\theta],N_{[b_\theta]}}, S^2_{X_{i[b_\theta]}}(n_0)\right\}\right] - (|I_b|-1) \\
&\geq \sum_{\theta\in I_b} \prod_{i\in(S_a^{(\theta)}\cup S_{de}^{(\theta)}), i\neq[b_\theta]} \Pr\left(\mathrm{CS}_i^{(\theta)}\right) - (|I_b|-1) \\
&\geq \sum_{\theta\in I_b} (1-\alpha_c')^{|S_a^{(\theta)}|+|S_{de}^{(\theta)}|-1} - (|I_b|-1),
\end{aligned}
$$

where the second inequality holds due to the Bonferroni inequality, the third inequality holds due to Lemma 2.4 in [17] and the last inequality holdes due to Lemma 10. As we know that $|S_a^{(\theta)}| + |S_{de}^{(\theta)}| = k - |S_u^{(\theta)}| \leq k$, we have

$$
\begin{aligned}
\Pr(\mathcal{B}_2^*) &\geq \sum_{\theta\in I_b} (1-\alpha_c')^{k-1} - (|I_b|-1) = |I_b|(1-\alpha_c')^{k-1} - (|I_b|-1) \\
&= |I_b|\left[(1-\alpha_c')^{k-1} - 1\right] + 1 \geq (d+1)\left[(1-\alpha_c')^{k-1} - 1\right] + 1 \\
&= (d+1)(1-\alpha_c')^{k-1} - d,
\end{aligned}
$$

where the last inequality holds as $1 \leq |I_b| \leq d+1$ and $(1-\alpha_c')^{k-1} - 1 \leq 0$.

Thus, we know that

$$
\begin{aligned}
\Pr\{\mathrm{CS}\} &\geq (1-\min\{s,d\}\alpha_f')^{k-j_b-j_d-j_a} \times (1-2s\alpha_f')^{j_b} \times \left[(d+1)(1-\alpha_c')^{k-1} - d\right] \\
&\geq (1-\min\{s,d\}\alpha_f')^{k-j_b} \times (1-2s\alpha_f')^{j_b} \times \left[(d+1)(1-\alpha_c')^{k-1} - d\right] \\
&\geq (1-\min\{s,d\}\alpha_f')^{k-\min\{d+1,k\}} \times (1-2s\alpha_f')^{\min\{d+1,k\}} \times \left[(d+1)(1-\alpha_c')^{k-1} - d\right] \\
&= (1-\alpha_f)(1-\alpha_c) = 1-\alpha,
\end{aligned}
$$

where the second inequality holds since the lower bound of $(1-\min\{s,d\}\alpha_f')^{k-j_b-j_d-j_a}$ is achieved when $j_d = j_a = 0$, and the third inequality holds since the lower bound of $(1-\min\{s,d\}\alpha_f')^{k-j_b} \times (1-2s\alpha_f')^{j_b}$ is achieved when $j_b = \min\{d+1,k\}$.

When systems are simulated under CRN, due to Lemmas 8, 9, 10, and the Bonferroni inequality, we have

$$
\begin{aligned}
&\Pr(\mathcal{A}^*) \geq 1 - j_u(\min\{s,d\}\alpha_f') = 1 - (k-j_b-j_d-j_a)(\min\{s,d\}\alpha_f') \\
&\Pr(\mathcal{B}_1^*) \geq 1 - j_b(2s\alpha_f') \\
&\Pr(\mathcal{B}_2^*) \geq \Pr\left(\cap_{\theta\in I_b} \cap_{i\in(S_a^{(\theta)}\cup S_{de}^{(\theta)}), i\neq[b_\theta]} \mathrm{CS}_i^{(\theta)}\right) \geq 1 - \sum_{\theta\in I_b} \sum_{i\in(S_a^{(\theta)}\cup S_{de}^{(\theta)}), i\neq[b_\theta]} \Pr\left(\mathrm{CS}_i^{(\theta)}\right)
\end{aligned}
$$

$$\geq 1 - |I_b| \left(|S_a^{(\theta)}| + |S_{de}^{(\theta)}| - 1 \right) \alpha'_c \geq 1 - (d+1)(k-1)\alpha'_c,$$

where the last inequality holds as $1 \leq |I_b| \leq d+1$ and $|S_a^{(\theta)}| + |S_{de}^{(\theta)}| \leq k$.

Thus, we know that

$$\begin{aligned}
\Pr\{\text{CS}\} &\geq \left[1 - (k - j_b - j_d - j_a)(\min\{s,d\}\alpha'_f) - j_b(2s\alpha'_f)\right] \left[1 - (d+1)(k-1)\alpha'_c\right] \\
&\geq \left[1 - (k - j_b)(\min\{s,d\}\alpha'_f) - j_b(2s\alpha'_f)\right] \left[1 - (d+1)(k-1)\alpha'_c\right] \\
&= \left[1 - (k\min\{s,d\} + (2s - \min\{s,d\})j_b)\alpha'_f\right] \left[1 - (d+1)(k-1)\alpha'_c\right] \\
&\geq \left[1 - (k\min\{s,d\} + (2s - \min\{s,d\}) \cdot \min\{d+1,k\})\alpha'_f\right] \left[1 - (d+1)(k-1)\alpha'_c\right] \\
&= (1-\alpha_f)(1-\alpha_c) = 1 - \alpha,
\end{aligned}$$

where the second inequality holds as the lower bound of $\left[1 - (k - j_b - j_d - j_a)(\min\{s,d\}\alpha'_f) - j_b(2s\alpha'_f)\right]$ is achieved when $j_d = j_a = 0$, and the last inequality holds as the lower bound of $\left[1 - (k\min\{s,d\} + (2s - \min\{s,d\})j_b)\alpha'_f\right]$ is achieved when $j_b = \min\{d+1,k\}$. □

The chosen values of α_f and α_c affect the performance of the $\mathcal{FAP}^{\mathcal{R}}$ procedure. If Phase I is relatively more difficult compared with Phase II (e.g., the secondary performance measures of many systems are close to some threshold vectors in consideration), choosing a larger value of α_f compared with α_c can improve the efficiency. On the other hand, if Phase II is relatively more difficult, choosing a larger α_c is plausible. If the decision maker has knowledge on the relative difficulty of Phases I and II, she may first decide the choice of $e = \alpha_f/\alpha_c$, the ratio of the nominal error of Phase I to that of Phase II. The value of α_c can be found by solving

$$(1 - e \times \alpha_c)(1 - \alpha_c) = 1 - \alpha,$$

and then $\alpha_f = e \cdot \alpha_c$. However, the decision maker usually does not have information about the mean configurations of the primary and secondary performance measures of the systems in advance. One possibility is to select $\alpha_f = \alpha_c = 1 - (1-\alpha)^{1/2}$. One may also consider similar approaches as in Chapter 3 such as setting $s\alpha'_f = \alpha'_c$ or $d\alpha'_f = \alpha'_c$ depending on the relationship between s and d.

4.3.3 Procedure $\mathcal{FAP}$

In this section, we discuss a variation of the $\mathcal{FAP}^{\mathcal{R}}$ procedure that collects observations on the primary performance measure in Phase I and recycles them in Phase II.

As $\mathcal{FAP}^{\mathcal{R}}$ starts "from scratch" when performing the comparison, it discards all the information related to the primary performance measure obtained in Phase I, which can be quite inefficient in terms of the computation effort. One may consider restoring all the observations of the primary performance measure collected from Phase I and then extracting information related to the primary performance measure when performing Phase II. However, as Phase I may require a lot of observations, this approach requires a significant amount of memory for restoring the observations from Phase I.

[16] propose the sequential selection with memory procedure ($\mathcal{SSM}$) that is specifically for use within an optimization-via-simulation algorithm when simulation is costly, and partial or complete information on solutions previously visited is maintained. When data storage is prohibitive, the procedure requires only summary statistics of the simulation output, which solves the memory space issue discussed above. We then present a sequentially running procedure, namely $\mathcal{FAP}$, that adopts the $\mathcal{SSM}$ procedure as its Phase II. The detailed description is shown in Algorithm 10.

There is a major difficulty in proving the statistical validity of $\mathcal{FAP}$. As the number of observations X_{in} collected in Phase I depends on $Y_{i\ell n}$ for system i, this dependency affects the comparison in Phase II. This dependency issue can be resolved by performing $\mathcal{FAP}^{\mathcal{R}}$ instead as it restarts "from scratch" for the surviving systems of Phase I. We address the dependency between Phases I and II in $\mathcal{FAP}$ by choosing the nominal errors α_f and α_c such that $\alpha_f + \alpha_c = \alpha$. Although we have not proved the statistical validity of $\mathcal{FAP}$, our experimental results (discussed in Section 4.5) do not show any violation of its validity.

4.4 Simultaneously-Running Procedure

In this section, we first propose a simultaneously-running procedure $\mathcal{FAP}+$ in Section 4.4.1, and then prove its statistical validity in Section 4.4.2.

Algorithm 10 $\mathcal{FAP}$

[**Setup**:] Same as in $\mathcal{FAP}^{\mathcal{R}}$ except for choosing $\alpha_f, \alpha_c > 0$ such that $\alpha_f + \alpha_c = \alpha$.
[**Initialization for Phase I**:] Same as in $\mathcal{FAP}^{\mathcal{R}}$ except for the following additional steps for each system:

- Obtain n_0 observations $X_{in}, n = 1, \ldots, n_0$.
- For each system i, compute $\bar{X}_i(n_0)$.
- For all systems $i \neq j$, compute $S^2_{X_{ij}}(n_0)$.

[**Feasibility Check**:] Same as in $\mathcal{FAP}^{\mathcal{R}}$.
[**Stopping Condition for Phase I**:] Same as in $\mathcal{FAP}^{\mathcal{R}}$ except that we also take one additional observation X_{i,r_i+1} and update $\bar{X}_i(r_i)$ whenever we take one additional observation $Y_{i,r_i+1,\ell}$ for $\ell = 1, \ldots, s$, from $i \in M$.
[**Initialization for Phase II**:] Same as in $\mathcal{FAP}^{\mathcal{R}}$ except we do not perform new simulation and do not compute $\bar{X}_i(n_0)$ and $S^2_{X_{ij}}(n_0)$. Set $r = \min_{i\in\Gamma} r_i$ and go to [**Comparison**].
[**Comparison**:] Same as in $\mathcal{FAP}^{\mathcal{R}}$ except we replace equation (9) by

$$r\bar{X}_i(r_i) > r\bar{X}_j(r_j) + R(r; \delta, \eta_c, S^2_{X_{ij}}(n_0)).$$

[**Stopping Condition for Phase II**:]
If $|T| = 0$, stop and return the set of systems in $\mathcal{F}_\theta$, where $\theta = 1, \ldots, d+1$, as the portfolio of best systems. Otherwise, for all $i \in M$ with $r_i = r$, take one additional observation X_{i,r_i+1}, compute $\bar{X}_i(r_i+1)$, and set $r_i = r_i + 1$. Then set $r = r + 1$ and go to [**Comparison**].

4.4.1 Procedure $\mathcal{FAP}+$

In this section, we discuss a procedure that runs Phases I and II simultaneously.

We use M and $\mathcal{F}_\theta$ defined as in Section 4.3.1 and introduce the following additional sets:

- SS_i is a set of systems that are superior to system i in terms of the primary performance measure, where $i \in \Gamma$.
- M_θ is a set of systems that can potentially be the best feasible system with respect to $\boldsymbol{q}^{(\theta)}$ but need more observations for either feasibility check or comparison against other systems in M_θ or $\mathcal{F}_\theta$, where $\theta = 1, \ldots, d+1$.

Procedure $\mathcal{FAP}+$ performs feasibility check and comparison among systems that are still in consideration (i.e., $i \in M$ where $M = \cup_{\theta=1}^{d+1} M_\theta$) within each iteration. Similar to the sequentially-running procedures, we utilize procedure $\mathcal{RF}$ in Algorithm 1 for the feasibility checks. We remove system i from set M_θ if it is declared infeasible with respect to $\boldsymbol{q}^{(\theta)}$ (as it can no longer be considered as a potential best system with respect to $\boldsymbol{q}^{(\theta)}$). If system i is declared feasible with respect to $\boldsymbol{q}^{(\theta)}$, we first add system i to set $\mathcal{F}_\theta$, then eliminate any inferior system j (i.e., with $i \in SS_j$) within set $\mathcal{F}_\theta$ or M_θ. System i can also be eliminated from $\mathcal{F}_\theta$ if there exists a superior system j deemed feasible with respect to $\boldsymbol{q}^{(\theta)}$ (i.e., $j \in \mathcal{F}_\theta \cap SS_i$). When M_θ and $\mathcal{F}_\theta$ both contain only one system, then M_θ is set to an empty set. This is because the only system that is potentially the best system with respect to $\boldsymbol{q}^{(\theta)}$ is also the only feasible system with respect to $\boldsymbol{q}^{(\theta)}$. Therefore we can safely identify this system as the best feasible system with respect to $\boldsymbol{q}^{(\theta)}$ and remove it from set M_θ. Comparison is performed to all pairs of systems (i, j) such that $i, j \in M = \cup_{\theta=1}^{d+1} M_\theta$ and their superiority is unknown yet (i.e., $i \notin SS_j$ and $j \notin SS_i$). Within the comparison phase, system j is eliminated from M_θ and $\mathcal{F}_\theta$ whenever it is found inferior to another system i already deemed feasible to $\boldsymbol{q}^{(\theta)}$ for $\theta = 1, 2, \ldots, d+1$. If the superior system's feasibility with respect to $\boldsymbol{q}^{(\theta)}$ is unknown, then system i is added to the superior set SS_j. The procedure stops when $|M| = 0$. A detailed description of $\mathcal{FAP}+$ is provided in Algorithm 11.

4.4.2 Statistical Validity for the Simultaneously-Running Procedure

In this section, we prove the statistical validity for the proposed simultaneously-running procedure $\mathcal{FAP}+$.

Let β_f and β_c denote the nominal error of feasibility check for one constraint of single system with fixed threshold and the nominal error of comparison between two systems, respectively. We first present following lemmas.

Lemma 11. *For a particular θ and any system $i \in S_u$, the* **[Feasibility Check]** *steps in $\mathcal{FAP}+$ run to completion ensure*

$$\Pr\left(\mathcal{A}^*(i)\right) \geq 1 - \min\{s, d\}\beta_f.$$

Lemma 12. *For any system $i \in B$, the* **[Feasibility Check]** *steps in $\mathcal{FAP}+$ run to completion ensure*

$$\Pr\left(\mathcal{B}_1(i)\right) \geq 1 - 2s\beta_f.$$

Lemma 13. *Given an existing best desirable system $[b_\theta]$ and i such that $x_i \leq x_{[b_\theta]} - \delta$, the* **[Comparison]** *steps for system i and $[b_\theta]$ in $\mathcal{FAP}+$ run to completion ensure*

$$\Pr\left(\mathrm{CS}_i^{(\theta)}\right) \geq 1 - \beta_c.$$

Algorithm 11 $\mathcal{FAP}+$

[**Setup**:] Choose confidence level $1-\alpha$, tolerance levels $\epsilon_1,\ldots,\epsilon_s$, indifference-zone parameter δ, threshold vectors $\{\boldsymbol{q}^{(1)},\boldsymbol{q}^{(2)},\ldots,\boldsymbol{q}^{(d)}\}$, and associated index vectors $\{\boldsymbol{I}^{(1)},\boldsymbol{I}^{(2)},\ldots,\boldsymbol{I}^{(d)}\}$. Set $M=\Gamma, SS_i=\emptyset$ and $Z_{i\ell}^m=2$ for all $i\in M,\ell=1,\ldots,s$, and $m=1,\ldots,d_\ell$. Set $M_\theta=\Gamma$ for $\theta=1,\ldots,d+1$. Set $\mathcal{F}_\theta=\emptyset$ for $\theta=1,\ldots,d$ and $\mathcal{F}_{d+1}=\Gamma$. Choose $\beta_f,\beta_c>0$ such that β_f and β_c satisfy

$$(1-2s\beta_f)^{\min\{d+1,k\}}+(1-\min(s,d)\beta_f)^k+(d+1)\left[(1-\beta_c)^{k-1}-1\right]-1=1-\alpha,$$

if systems are simulated independently; and

$$1-\min\{d+1,k\}(2s\beta_f)-k\min(s,d)\beta_f-(d+1)(k-1)\beta_c=1-\alpha,$$

if systems are simulated under CRN.

Set η_f and η_c such that $g(\eta_f)=\beta_f$ and $g(\eta_c)=\beta_c$.
[**Initialization**:]
for each system $i\in M$ **do**

- Obtain n_0 observations from system i
- Compute $\bar{X}_i(n_0),\bar{Y}_{i\ell}(n_0),S^2_{X_{ij}}(n_0)$, and $S^2_{Y_{i\ell}}(n_0)$ for all $i,j\in M$, where $i\neq j$, and $\ell=1,\ldots,s$.
- Set $r=n_0,\mathrm{ON}_i=\{1,\ldots,s\}$, and $\mathrm{ON}_{i\ell}=\{1,\ldots,d_\ell\}$ for all $\ell=1,\ldots,s$.
- Set $v_{i\ell}^{\mathrm{UB}}=\infty$ and $v_{i\ell}^{\mathrm{LB}}=-\infty$ for $i\in M$ and $\ell=1,\ldots,s$.

end for
[**Feasibility Check**:]
for $i\in M$ **do**
  for $\ell\in\mathrm{ON}_i$ **do**
    $v_{i\ell}^{\mathrm{UB}}=\min(v_{i\ell}^{\mathrm{UB}},\bar{Y}_{i\ell}(r)+R(r;\epsilon_\ell,\eta_f,S^2_{i\ell}(n_0))/r_i)$.
    $v_{i\ell}^{\mathrm{LB}}=\max(v_{i\ell}^{\mathrm{LB}},\bar{Y}_{i\ell}(r)-R(r;\epsilon_\ell,\eta_f,S^2_{i\ell}(n_0))/r_i)$.
    for $m\in\mathrm{ON}_{i\ell}$ **do**
      If $v_{i\ell}^{\mathrm{UB}}\le q_\ell^m$, set $Z_{i\ell}^m=1$ and $\mathrm{ON}_{i\ell}=\mathrm{ON}_{i\ell}\backslash\{m\}$;
      If $v_{i\ell}^{\mathrm{LB}}\ge q_\ell^m$, set $Z_{i\ell}^m=0$ and $\mathrm{ON}_{i\ell}=\mathrm{ON}_{i\ell}\backslash\{m\}$.
    end for
    If $\mathrm{ON}_{i\ell}=\emptyset$, set $\mathrm{ON}_i=\mathrm{ON}_i\backslash\{\ell\}$.
  end for
  for $\theta=1,\ldots,d$ **do**
    If $\prod_{\ell=1}^s Z_{i\ell}^{I_\ell^{(\theta)}}=0$, then remove system i from M_θ.
    If $\prod_{\ell=1}^s Z_{i\ell}^{I_\ell^{(\theta)}}=1$, then

- Add system i to $\mathcal{F}_\theta$.
- If there exists system j such that $j\in\mathcal{F}_\theta\cup M_\theta$ with $i\in SS_j$, then eliminate j from $\mathcal{F}_\theta$ and M_θ, and delete SS_j.

  end for
end for
For $\theta=1,\ldots,d+1$, if $|M_\theta|=1$ and $M_\theta=\mathcal{F}_\theta$, then set $M_\theta=\emptyset$.
[**Comparison**:] Set $M=\cup_{\theta=1}^{d+1}M_\theta$.
for $i,j\in M$ s.t. $i\neq j,i\notin SS_j,j\notin SS_i$, and

$$r\bar{X}_i(r)>r\bar{X}_j(r)+R(r;\delta,\eta_c,S^2_{X_{ij}}(n_0)),$$

do
  Add system i into SS_j, eliminate j from $\mathcal{F}_{d+1}$ and M_{d+1} (if $j\in\mathcal{F}_{d+1}\cup M_{d+1}$), and
  for $\theta=1,\ldots,d$ **do**
    If $i\in\mathcal{F}_\theta$ and $j\in M_\theta$, then eliminate j from $\mathcal{F}_\theta$ and M_θ, and delete SS_j.
  end for
end for
For $\theta=1,\ldots,d+1$, if $|M_\theta|=1$ and $M_\theta=\mathcal{F}_\theta$, then set $M_\theta=\emptyset$.
[**Stopping Condition**:]
Set $M=\cup_{\theta=1}^{d+1}M_\theta$.
If $|M|=0$, then stop and return the set of systems in $\mathcal{F}_\theta$ as the portfolio of best systems. Otherwise, for all $i\in M$, take one additional observation $X_{i,r+1}$, and compute $\bar{X}_i(r+1)$ and $\bar{Y}_{i\ell}(r+1)$ for all $\ell\in\mathrm{ON}_i$. Then set $r=r+1$, and go to [**Feasibility Check**].

The proofs of Lemmas 11, 12, and 13 are essentially same as those of Lemmas 8, 9, and 10 because both α'_f of $\mathcal{FAP}^{\mathcal{R}}$ and β_f of $\mathcal{FAP}+$ are the nominal error of feasibility check for one constraint of single system with a fixed threshold, and both α'_c of $\mathcal{FAP}^{\mathcal{R}}$ and β_c of $\mathcal{FAP}+$ are the nominal error of comparison between an system in $S_a^{(\theta)} \cup S_{de}^{(\theta)} \setminus \{[b_\theta]\}$ and a best system $[b_\theta]$ for one particular θ. We now present the following theorem to prove the statistical validity of $\mathcal{FAP}+$.

Theorem 7. *Under Assumptions 4 and 5, $\mathcal{FAP}+$ guarantees*

$$\Pr\{\text{CS}\} \geq 1 - \alpha.$$

Proof. To ensure the correct selection, we need correct feasibility decisions for systems in B, correct feasibility decisions for systems in S_u, and correct comparison between $[b_\theta]$ and $i \in (S_a^{(\theta)} \cup S_{de}^{(\theta)})$ and $i \neq [b_\theta]$ for each $\theta \in I_b$. We consider the same definition of $\mathcal{B}_1^*$ from Section 4.3.2. We then have

$$\begin{aligned}\Pr\{\text{CS}\} &\geq \Pr\left(\mathcal{B}_1^* \cap (\cap_{i \in S_u} \mathcal{A}^*(i)) \cap \left(\cap_{\theta \in I_b} \cap_{i \in (S_a^{(\theta)} \cup S_{de}^{(\theta)}), i \neq [b_\theta]} \text{CS}_i^{(\theta)}\right)\right) \\ &\geq \Pr(\mathcal{B}_1^*) + \Pr(\cap_{i \in S_u} \mathcal{A}^*(i)) + \Pr\left(\cap_{\theta \in I_b} \cap_{i \in (S_a^{(\theta)} \cup S_{de}^{(\theta)}), i \neq [b_\theta]} \text{CS}_i^{(\theta)}\right) - 2.\end{aligned}$$

We discuss the cases depending on whether systems are simulated independently or under CRN. When systems are simulated independently, by Lemma 12, we know that

$$\Pr(\mathcal{B}_1^*) \geq (1 - 2s\beta_f)^{j_b}.$$

By Lemma 11, we also have

$$\Pr(\cap_{i \in S_u} \mathcal{A}^*(i)) \geq (1 - \min(s, d)\beta_f)^{j_u}.$$

By similar arguments as stated in the proof of Theorem 1, we get

$$\Pr\left(\cap_{\theta \in I_b} \cap_{i \in (S_a^{(\theta)} \cup S_{de}^{(\theta)}), i \neq [b_\theta]} \text{CS}_i^{(\theta)}\right) \geq \sum_{\theta \in I_b} (1 - \beta_c)^{|S_a^{(\theta)}| + |S_{de}^{(\theta)}| - 1} - (|I_b| - 1),$$

As $|S_a^{(\theta)}| + |S_{de}^{(\theta)}| \leq k$, we know that

$$\begin{aligned}\Pr\left(\cap_{\theta \in I_b} \cap_{i \in (S_a^{(\theta)} \cup S_{de}^{(\theta)}), i \neq [b_\theta]} \text{CS}_i^{(\theta)}\right) &\geq \sum_{\theta \in I_b} (1 - \beta_c)^{k-1} - (|I_b| - 1) = |I_b| (1 - \beta_c)^{k-1} - (|I_b| - 1) \\ &= |I_b| \left[(1 - \beta_c)^{k-1} - 1\right] + 1 \geq (d+1) \left[(1 - \beta_c)^{k-1} - 1\right] + 1,\end{aligned}$$

where the last inequality holds as $1 \le |I_b| \le d+1$ and $(1-\beta_c)^{k-1} - 1 \le 0$.

Thus, we know that

$$\begin{aligned}\Pr(\text{CS}) &\ge (1-2s\beta_f)^{j_b} + (1-\min(s,d)\beta_f)^{j_u} + (d+1)\left[(1-\beta_c)^{k-1} - 1\right] + 1 - 2 \\ &\ge (1-2s\beta_f)^{\min\{d+1,k\}} + (1-\min(s,d)\beta_f)^{k} + (d+1)\left[(1-\beta_c)^{k-1} - 1\right] - 1 \\ &= 1-\alpha,\end{aligned}$$

where the second inequality holds as the minimum of $(1-2s\beta_f)^{j_b}$ is achieved when $j_b = \min\{d+1,k\}$ and the minimum of $(1-\min(s,d)\beta_f)^{j_u}$ is achieved when $j_u = k$.

When systems are simulated under CRN, by Lemma 11, 12, and 13, we know that

$$\begin{aligned}\Pr(\mathcal{B}_1^*) &\ge 1 - j_b(2s\beta_f); \\ \Pr\left(\cap_{i\in S_u}\mathcal{A}^*(i)\right) &\ge 1 - j_u(\min(s,d)\beta_f); \\ \Pr\left(\cap_{\theta\in I_b}\cap_{i\in(S_a^{(\theta)}\cup S_{de}^{(\theta)}),i\neq[b_\theta]} \text{CS}_i^{(\theta)}\right) &\ge 1 - \sum_{\theta\in I_b}\sum_{i\in(S_a^{(\theta)}\cup S_{de}^{(\theta)}),i\neq[b_\theta]} \Pr\left(\text{ICS}_i^{(\theta)}\right) \\ &\ge 1 - \sum_{\theta\in I_b}\sum_{i\in(S_a^{(\theta)}\cup S_{de}^{(\theta)}),i\neq[b_\theta]} \Pr\left(\text{ICS}_i^{(\theta)}\right) \\ &\ge 1 - \sum_{\theta\in I_b}\left(|S_a^{(\theta)}| + |S_{de}^{(\theta)}| - 1\right)\beta_c \ge 1-(d+1)(k-1)\beta_c,\end{aligned}$$

where we use $\text{ICS}_i^{(\theta)}$ to denote the complement event of $\text{CS}_i^{(\theta)}$ and the last inequality holds as $|S_a^{(\theta)}| + |S_{de}^{(\theta)}| \le k$. Thus, we know that

$$\begin{aligned}\Pr(\text{CS}) &\ge 1 - j_b(2s\beta_f) - j_u\min(s,d)\beta_f - (d+1)(k-1)\beta_c \\ &\ge 1 - \min\{d+1,k\}(2s\beta_f) - k\min(s,d)\beta_f - (d+1)(k-1)\beta_c \\ &= 1-\alpha,\end{aligned}$$

where the second inequality holds as the upper bound of $j_b(2s\beta_f)$ is achieved when $j_b = \min\{d+1,k\}$ and $0 \le j_u \le k$. □

4.5 Experimental Results

In this section, we present experimental results to demonstrate the performance of procedures $\mathcal{FAP}^{\mathcal{R}}$, $\mathcal{FAP}$, and $\mathcal{FAP}+$. We compare the performance of the proposed procedures with two alternative procedures that iteratively apply $\mathcal{AK}$ and $\mathcal{AK}+$ due to [1] if a single constraint is considered, or $\mathcal{HAK}$ and $\mathcal{HAK}+$ due to [5] if multiple constraints are considered for $\boldsymbol{q}^{(\theta)}$, $\theta = 1, 2, \ldots, d$. For $\boldsymbol{q}^{(d+1)}$, $\mathcal{KN}$ due to [10] is applied because $\boldsymbol{P}(d+1)$ is an unconstrained R&S problem. Procedures $\mathcal{AK}$ and $\mathcal{HAK}$ are sequentially-running procedures for constrained R&S with a fixed threshold vector, while $\mathcal{AK}+$ and $\mathcal{HAK}+$ are simultaneously-running procedures for constrained R&S with a fixed threshold vector. We refer to the procedure that iteratively implements $\mathcal{AK}$ or $\mathcal{AK}+$ as Restart$^{\mathcal{AK}}$ or Restart$^{\mathcal{AK}+}$ and to the procedure that iteratively implements $\mathcal{HAK}$ or $\mathcal{HAK}+$ as Restart$^{\mathcal{HAK}}$ or Restart$^{\mathcal{HAK}+}$. Notice that Restart$^{\mathcal{AK}}$ (Restart$^{\mathcal{AK}+}$) is a special case of Restart$^{\mathcal{HAK}}$ (Restart$^{\mathcal{HAK}+}$) when the number of constraints is one. We provide the algorithm statement and discussion of the statistical validity of procedures Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ in Appendices C.1 and C.2, respectively.

All the experimental results are based on 10,000 macro replications with $\alpha = 0.05$ and $n_0 = 20$. We set $\delta = \epsilon_\ell = 1/\sqrt{n_0}$, where $\ell = 1, \ldots, s$. Section 4.5.1 defines non-overlapping regions that are needed to describe our experimental configuration. The experimental configurations we consider are given in Section 4.5.2. Sections 4.5.3 and 4.5.4 provide our results regarding the validity and the efficiency of the proposed procedures, respectively.

4.5.1 Non-overlapping Regions

In this section, we introduce definitions of non-overlapping regions that we need to explain the mean setting for our experiments.

We let $\mathcal{R}_\theta$ be defined as follows:

$$\mathcal{R}_\theta = \begin{cases} \left\{(z_1, z_2, \ldots, z_s) : z_\ell \le q_\ell^{(\theta)}, \ell = 1, 2, \ldots, s\right\}, & \text{if } \theta = 1; \\ \left\{(z_1, z_2, \ldots, z_s) : z_\ell \le q_\ell^{(\theta)}, \ell = 1, 2, \ldots, s\right\} \setminus \cup_{\kappa=1}^{\theta-1} \mathcal{R}_\kappa, & \text{if } \theta = 2, \ldots, d; \\ \mathbb{R}^s \setminus \cup_{\kappa=1}^{d} \mathcal{R}_\kappa, & \text{if } \theta = d+1. \end{cases}$$

Notice that $\mathcal{R}_\theta$ denotes the part of a feasible region with respect to threshold vector $\boldsymbol{q}^{(\theta)}$ that does not overlap with those of threshold vectors $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta-1)}$. For example, Figure 18 shows non-overlapping regions when a single constraint is in presence. The axis corresponds to the secondary performance mean and for $\theta = 2, \ldots, d$, $\mathcal{R}_\theta$ is the set of real values that are less than or equal to q_1^θ and greater than $q_1^{\theta-1}$ (i.e, $(q_1^{\theta-1}, q_1^\theta]$), while $\mathcal{R}_1$ is the set of values that are less than or equal to q_1^1 and $\mathcal{R}_{d+1}$ is the set of values that are greater than q_1^d. Any system whose secondary performance measure mean falls in $\mathcal{R}_\theta$, where $\theta \geq 2$, is feasible with respect to q_1^θ but infeasible to all $q_1^{\theta'}$ where $\theta' < \theta$. A system whose secondary performance mean falls in $\mathcal{R}_1$ is feasible with respect to all the thresholds.

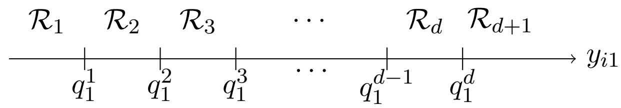
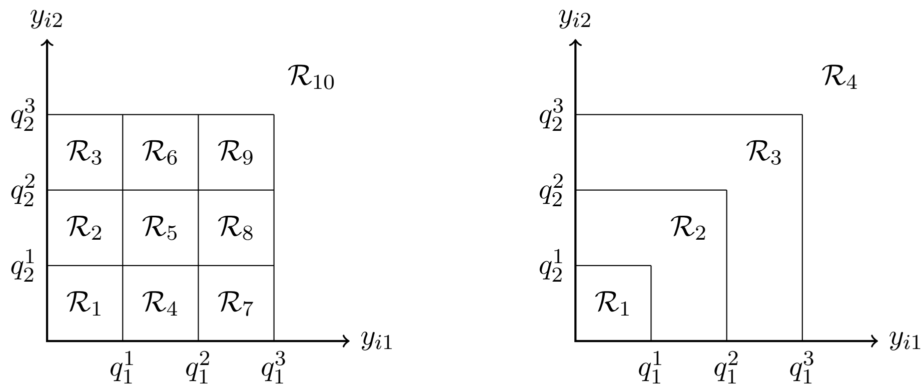


Figure 18: Non-overlapping regions for a single constraint with d thresholds

When the number of constraints is greater than one, one needs to first formulate the input threshold vectors based on how the decision maker wants to prioritize each constraint. In the experimental section with multiple constraints, we consider two of the formulations discussed in Section 3.5 of Chapter 3, namely ranked constraints and equally important constraints.

The ranked constraint formulation ranks the constraints based on their importance and relaxes the least important constraint first while keeping the other constraints fixed at the current threshold values, and then moves to the second least important constraint, etc. For example, Figure 19a shows the non-overlapping regions when there are two constraints where the first constraint is more important than the second constraint, and both constraints have three threshold values.

The equally important constraints formulation assumes that all the constraints are equally important. The decision maker wants to relax all constraints by one threshold at the same time. If the constraints do not all have the same number of thresholds, then constraints that have gone through all their thresholds keep the "loosest" threshold (i.e., $q_\ell^{d_\ell}$ for constraint ℓ) while the other constraints are relaxed. Figure 19b shows the non-overlapping region when there are two constraints and each constraint has three threshold values.



(a) Ranked constraints (b) Equally important constraints

Figure 19: Feasible region for two constraints

4.5.2 Experimental Configurations

For the primary performance measure, we consider an increasing means configuration where we set $x_i = \mathrm{E}[X_{in}] = (i-1)\delta$ for $i = 1, \ldots, k$. For the secondary performance measures, the first threshold value for each constraint is set to 0 and the remaining threshold values for constraint ℓ increase by $2\epsilon_\ell$ (i.e., $q_\ell^1 = 0$ and $q_\ell^m = q_\ell^{m-1} + 2\epsilon_\ell$ for any $m = 2, 3, \ldots, d_\ell$). We consider the three configurations discussed below.

- $\mathcal{P}(\theta)$: All the systems fall in the non-overlapping region $\mathcal{R}_\theta$. In this configuration, we test two extreme cases when all systems fall in $\mathcal{R}_1$ (i.e., all systems are feasible with respect to all the threshold vectors) and when all systems fall in $\mathcal{R}_{d+1}$ (i.e., all systems are infeasible to $\boldsymbol{q}(\theta)$, $\theta = 1, 2, \ldots, d$). As the proof of Theorem 1 shows the lower bound of Pr{CS} is achieved at $j_d = 0$, which means not all systems are in $\mathcal{R}_1$, we also test a case when all systems fall in $\mathcal{R}_2$.

- $\mathcal{Q}(\theta)$: $\mathcal{R}_{\theta'}$ has no system for $\theta' = 1, 2, \ldots, \theta - 1$, $\mathcal{R}_\theta$ contains $\max\{0, k-(d+1)+\theta\}$ systems, and $\mathcal{R}_{\theta'}$ has one system for $\theta' = \theta+1, \ldots, d+1$. In addition, the system in $\mathcal{R}_{\theta'}$ is the best feasible system with respect to $\boldsymbol{q}^{(\theta')}$ for each $\theta' = \theta+1, \ldots, d+1$. For example, in $\mathcal{Q}(1)$, $\mathcal{R}_1$ has $k-d$ systems and each $\mathcal{R}_{\theta'}$ for $\theta' = 2, \ldots, d+1$ has one system. In $\mathcal{Q}(2)$, $\mathcal{R}_1$ has no system, $\mathcal{R}_2$ has $k-d+1$ systems, and each $\mathcal{R}_{\theta'}$ for $\theta' = 3, \ldots, d+1$ has one system. Note that $\mathcal{Q}(2)$ corresponds to the case when $j_d = 0$.

- $\mathcal{E}$: Systems are evenly distributed among non-overlapping regions $\mathcal{R}_1, \ldots, \mathcal{R}_{d+1}$. More specifically, R_θ contains $\lceil \frac{k}{d+1} \rceil$ systems, where $\theta = 1, \ldots, d+1$.

For system i that falls in the non-overlapping region $\mathcal{R}_\theta$, the secondary performance mean is set to $y_{i\ell} = q_\ell^{(\theta)} - \epsilon_\ell$ where $\ell = 1, \ldots, s$. Finally, the variances for both primary and secondary performance measures are set to one.

4.5.3 Statistical Validity of the Proposed Procedures

In this section, we test the performance of procedures $\mathcal{FAP}^{\mathcal{R}}, \mathcal{FAP}$, and $\mathcal{FAP}+$ and check their statistical validity. We first consider the single constraint case. We consider four threshold values (i.e., $d_1 = 4$) and choose $q_1 \in \{0, 2\epsilon_\ell, 4\epsilon_\ell, 6\epsilon_\ell\}$. Table **??** shows the estimated probability of correct selection (PCS) and the required number of observations for procedures $\mathcal{FAP}^{\mathcal{R}}, \mathcal{FAP}$, and $\mathcal{FAP}+$ based on the configurations discussed in Section 4.5.2.

Table 23: Average number of observations and observed PCS (reported in parentheses) of $\mathcal{FAP}^{\mathcal{R}}, \mathcal{FAP}$, and $\mathcal{FAP}+$ for $k = 10$ systems and $s = 1$ constraint under different configurations

	$\mathcal{P}(1)$	$\mathcal{P}(2)$	$\mathcal{P}(5)$	$\mathcal{Q}(1)$	$\mathcal{Q}(2)$	$\mathcal{E}$
$\mathcal{FAP}^{\mathcal{R}}$	2261.03	2519.32	2265.26	3797.92	3797.92	4475.31
	(0.997)	(0.992)	(0.994)	(0.989)	(0.990)	(0.988)
$\mathcal{FAP}$	1558.47	1744.50	1559.39	2687.70	2470.18	3143.88
	(0.999)	(0.993)	(0.992)	(0.990)	(0.988)	(0.988)
$\mathcal{FAP}+$	1329.59	1648.92	1582.51	2260.28	2193.59	2649.59
	(0.998)	(0.995)	(0.994)	(0.992)	(0.990)	(0.989)

Table 23 shows that all three procedures guarantee statistical validity under the configurations we tested. The three procedures spend fewer observations than the other configurations Configurations $\mathcal{P}(\theta)$ for $\theta = 1, 2, 5$ because configurations $\mathcal{P}(1), \mathcal{P}(2)$, and $\mathcal{P}(5)$ need fewer pairwise comparions than the other configurations and thus are easier. The estimated PCS of the three procedures are similar for $\mathcal{P}(1), \mathcal{P}(2)$, and $\mathcal{P}(5)$. Configurations $\mathcal{Q}(1), \mathcal{Q}(2)$, and $\mathcal{E}$ also show similar estimated PCS among three procedures but are smaller compared to configurations $\mathcal{P}(\theta)$, while they are above the nominal level $1 - \alpha = 0.95$.

We then consider the two constraints case. We choose the two formulations of the input threshold vectors as provided in Figure 9 (i.e., $d_1 = d_2 = 3$) and set $q_\ell \in \{0, 2\epsilon_\ell, 4\epsilon_\ell\}$ where $\ell = 1, 2$. We also consider the same experimental configurations as discussed in Section 4.5.2. One may notice that as we consider nine threshold vectors in total under the ranked constraints formulation, configuration $\mathcal{E}$ is same as $\mathcal{Q}(1)$. Therefore, we only consider five configurations under the ranked constraints formulation. Tables 24 and 25 provide the results for the ranked constraints formulation and the equally important constraints formulation, respectively.

Table 24: Average number of observations and observed PCS (reported in parentheses) of $\mathcal{FAP}^{\mathcal{R}}, \mathcal{FAP}$, and $\mathcal{FAP}+$ for $k = 10$ systems and $s = 2$ constraints based on the ranked constraints formulation

Ranked constraints	$\mathcal{P}(1)$	$\mathcal{P}(2)$	$\mathcal{P}(5)$	$\mathcal{Q}(1)$	$\mathcal{Q}(2)$
$\mathcal{FAP}^{\mathcal{R}}$	3116.24	3277.09	3113.16	6581.47	6432.05
	(0.999)	(0.998)	(1.000)	(0.988)	(0.990)
$\mathcal{FAP}$	2182.63	2310.26	2187.99	4498.24	4336.65
	(0.999)	(0.998)	(1.000)	(0.991)	(0.993)
$\mathcal{FAP}+$	1832.26	2120.08	1671.03	3351.58	3342.27
	(0.999)	(0.998)	(1.000)	(0.992)	(0.993)

Table 25: Average number of observations and observed PCS (reported in parentheses) of $\mathcal{FAP}^{\mathcal{R}}, \mathcal{FAP}$, and $\mathcal{FAP}+$ for $k = 10$ systems and $s = 2$ constraints based on the equally important constraints formulation

	$\mathcal{P}(1)$	$\mathcal{P}(2)$	$\mathcal{P}(5)$	$\mathcal{Q}(1)$	$\mathcal{Q}(2)$	$\mathcal{E}$
$\mathcal{FAP}^{\mathcal{R}}$	2669.90	2927.21	2675.94	3755.71	3571.29	4405.75
	(0.998)	(0.998)	(0.999)	(0.994)	(0.994)	(0.994)
$\mathcal{FAP}$	2182.63	2416.48	2187.99	2909.64	2811.58	3267.44
	(0.999)	(0.999)	(1.000)	(0.998)	(0.998)	(0.998)
$\mathcal{FAP}+$	1567.34	1654.82	1430.25	2146.16	2006.07	2420.13
	(0.998)	(0.998)	(0.999)	(0.994)	(0.997)	(0.994)

From Tables 24 and 25, we see that all three procedures provide statistical guarantees under the experimental configurations we tested. The results of the two constraints case show a similar pattern compared with those of the single constraint case. Configurations $\mathcal{P}(\theta)$ for $\theta = 1, 2, 5$ are easier than the other configurations, spending fewer observations, and $\mathcal{Q}(1)$ and $\mathcal{Q}(2)$ show lower PCS than $\mathcal{P}(\theta)$. The three procedures have similar value of the estimated PCS under each configuration.

Note that although we do not provide a proof for the statistical validity of procedure $\mathcal{FAP}$, our experimental results do not yield any violations to its validity. As we also see that $\mathcal{FAP}$ is more efficient than $\mathcal{FAP}^{\mathcal{R}}$, we omit experimental results for $\mathcal{FAP}^{\mathcal{R}}$ in the following section for the sake of space.

4.5.4 Efficiency of the Proposed Procedures

In this section, we discuss the efficiency of the proposed procedures by showing the experimental results of the alternative procedures Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ when a single constraint is considered and Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ when two constraints are considered. As $\mathcal{Q}(1), \mathcal{Q}(2)$, and $\mathcal{E}$ appear to be more difficult than $\mathcal{P}(1), \mathcal{P}(2)$, and $\mathcal{P}(5)$ based on estimated PCS, we test these configurations only.

Table 26 shows the experimental result for Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ when a single constraint is considered. We see that the proposed procedures $\mathcal{FAP}$ and $\mathcal{FAP}+$ show a significant reduction in terms of the required number observations as the competing procedure Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ take about two times as many observations. This is expected as Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{AK}+}$ need to be applied repeatedly from scratch with respect to four threshold values. We expect the savings in the number of observations with our procedure procedures to increase further as the number of thresholds increases. Note that $\mathcal{FAP}+$ also requires fewer observations compared with $\mathcal{FAP}$ under all configurations considered.

Table 26: Average number of observations and observed PCS (reported in parentheses) of $\mathcal{FAP}, \mathcal{FAP}+$, Restart$^{\mathcal{AK}}$, and Restart$^{\mathcal{AK}+}$ for $k = 10$ systems and $s = 1$ constraint

	$\mathcal{Q}(1)$	$\mathcal{Q}(2)$	$\mathcal{E}$
$\mathcal{FAP}$	2687.70	2470.18	3143.88
	(0.990)	(0.988)	(0.988)
Restart$^{\mathcal{AK}}$	6051.93	6115.48	5628.75
	(0.993)	(0.993)	(0.988)
$\mathcal{FAP}+$	2260.28	2193.59	2649.59
	(0.992)	(0.990)	(0.989)
Restart$^{\mathcal{AK}+}$	6896.11	6477.99	6622.89
	(0.986)	(0.981)	(0.983)

We then show the experimental results for Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ when two constraints are considered. We see that $\mathcal{FAP}$ and $\mathcal{FAP}+$ still show large savings in terms of the required number of observations. Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ spend about three to four times more observations than $\mathcal{FAP}$ and $\mathcal{FAP}+$ under the ranked constraints formulation and about two times as many observations under the equally important constraints formulation. As the ranked constraints formulation has more threshold vectors compared with that of the equally important constraints formulation, it is expected that the savings is larger under the ranked constraints formulation. Similar as in the single constraint case, $\mathcal{FAP}+$ is more efficient than $\mathcal{FAP}$ under both the ranked constraints and equally important constraints formulations.

Table 27: Average number of observations and observed PCS (reported in parentheses) of Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ for $k = 10$ systems and $s = 2$ constraints

	Ranked constraints		Equally important constraints		
	$\mathcal{Q}(1)$	$\mathcal{Q}(2)$	$\mathcal{Q}(1)$	$\mathcal{Q}(2)$	$\mathcal{E}$
$\mathcal{FAP}$	4498.24	4336.65	2909.64	2811.58	3267.44
	(0.991)	(0.993)	(0.998)	(0.998)	(0.998)
Restart$^{\mathcal{HAK}}$	15015.95	15180.44	5132.71	5990.34	4668.57
	(0.993)	(0.993)	(0.996)	(0.991)	(0.993)
$\mathcal{FAP}+$	3351.58	3342.27	2146.16	2006.07	2420.13
	(0.992)	(0.993)	(0.994)	(0.997)	(0.994)
Restart$^{\mathcal{HAK}+}$	15242.12	15012.79	5171.17	5327.06	5714.66
	(0.989)	(0.989)	(0.997)	(0.994)	(0.990)

4.6 Conclusion

We consider constrained R&S with subjective constraints and identify a portfolio of best systems with respect to each threshold vector. We propose two procedures: sequentially- and simultaneously-running procedures. We discuss the statistical validity of the proposed procedures and provide experimental results that show that the proposed procedures dominate alternative procedures that repeatedly solve the problem for each thresholds vector in terms of the required number of observations.

CHAPTER V

FUTURE RESEARCH DIRECTIONS

In this chapter, we discuss some possible directions for future research.

Subjective constraints can appear in many applications when the decision maker has flexibility with constraint thresholds. One possible direction is to use the procedures proposed and analyzed in Chapters 2, 3, and 4 in applications that identify feasible systems or select the best system when subjective stochastic constraints are considered. For example, for a resource allocation problem within health care, the decision maker may consider three performance measures regarding a working schedule of health care staff: (1) the overutilization rate of staff; (2) the average waiting time of patients; and (3) the expected cost, and she may also be willing to consider multiple threshold values for each performance measure. If the decision maker treats all three performance measures equally, she may simply want to determine the feasibility of each working schedule with respect to various combinations of threshold values across all the performance measures, where a schedule that is declared infeasible to a threshold value can be considered inferior to those that are declared feasible to the threshold value of a specific performance measure. In this case, the procedures discussed in Chapter 2 can be used. On the other hand, if the decision maker values the performance measure (1) more than (2) and (3), she may want to formulate a constrained optimization problem and identify the schedule with the lowest overutilization rate that also satisfies constraints (2) and (3) with the "tightest" threshold values possible. The procedures discussed in Chapter 3 can be a good fit to solve this problem. Furthermore, if the decision maker not only wants to identify the best schedule that is feasible to the most preferred threshold values possible but also wants to consider the trade-off between the primary and secondary performance measures, she may utilize the procedures proposed in Chapter 4 to find a portfolio of best schedules and assess how the identity of the best schedule changes depending on the set of threshold values.

Another direction for future research would be to incorporate subjective constraints into two other major approaches for solving constrained R&S problems, namely the OCBA and the Bayesian approaches. The current research discussed in Chapters 2, 3, and 4 is based on the IZ approach. As the IZ approach guarantees the probability of a correct decision, it is useful when the decision maker wishes to have a guarantee of correctness. However, if the number of observations or simulation time is limited, procedures based on the OCBA or the Bayesian approach are useful. These procedures attempt to identify an optimal budget allocation that maximizes the probability of correct selection or correct decision. One interesting future research direction is to determine how the OCBA and the Bayesian approaches can solve constrained R&S problems in the presence of subjective stochastic constraints.

APPENDIX A

In this section, we address a special case of adding thresholds when implementing the procedure $\mathcal{RF}$ in Appendix A.1 and discuss the existence and uniqueness of the algorithm parameter η in Appendix A.2. Appendix A.3 proves Lemma 1 for $c = \infty$. Appendices A.4, A.5, A.6, and A.7 provide the algorithm statements and statistical validity proofs for the four alternative procedures, namely Recycle$^{\mathcal{B}}$, Restart$^{\text{prod}}$, Restart$^{\text{sum}}$, and Restart$^{\text{max}}$, that we compared with $\mathcal{RF}$.

A.1 Adding Threshold Values for Constraint ℓ When $d_\ell = 1$

Let ℓ represent a specific constraint with only one threshold value (i.e., $d_\ell = 1$). We consider a case where a new threshold value is added for the constraint ℓ. If $\beta_1, \ldots, \beta_s$ are determined following (ii) in Algorithm 1, namely $\beta_{\ell'} = \beta/D$ for $\ell' = 1, 2, \ldots, s$, where $D = \sum_{\ell'=1}^{s} \min\{d_{\ell'}, 2\}$, then adding a new threshold value for ℓ requires changing the value of D which, in turn, changes all of $\beta_1, \ldots, \beta_s$. Therefore it is necessary to execute Algorithm 1 with $d_\ell = d_\ell + 1$ and ON $= \{1, \ldots, s\}$ for every system $i \in \Theta$. On the other hand, if $\beta_1, \ldots, \beta_s$ are determined by (i) in Algorithm 1, then adding a new threshold value to constraint ℓ affects β_ℓ only and thus the feasibility check needs to be re-done for constraint ℓ only with $d_\ell = d_\ell + 1$ and ON $= \{\ell\}$.

We recommend setting $\beta_1, \ldots, \beta_s$ according to (i) in Algorithm 1 if the decision maker knows that there is a possibility of adding more thresholds to any constraints ℓ with $d_\ell = 1$. Then the decision maker needs to implement Algorithm 12 if she adds a second threshold value. It is recommended that one save observations $Y_{i\ell r_i}$ corresponding to any constraint ℓ with $d_\ell = 1$ just in case that the decision maker wants to add a second threshold value for constraint ℓ later and thus has to implement Algorithm 12. For those constraints with two or more threshold values, there is no need to store observations if $\beta_1, \ldots, \beta_s$ are specified as in (i) in Algorithm 1. Another possibility is to directly set β_ℓ as if $d_\ell > 1$ for all $\ell = 1, \ldots, s$ if one wants to add thresholds later. In this case, one does not need to save data.

Algorithm 12 When $q_\ell^{d_\ell+1}$ is added for constraint ℓ such that $d_\ell = 1$ and $\beta_1, \ldots, \beta_s$ are chosen as in (i)

Set η_ℓ such that $g(\eta_\ell) = \beta_\ell$, where $\beta_\ell = \beta/(2s)$ and β satisfies (4).
for each system $i \in \Theta$ **do**
  Set ON $= \{\ell\}$ and ON$_\ell = \{1, \ldots, d_\ell + 1\}$. Set $v_{i\ell}^{\text{UB}} = \infty$ and $v_{i\ell}^{\text{LB}} = -\infty$. Set $r_i = n_0$.
  Go to [**Feasibility Check**] of Procedure $\mathcal{RF}$.
end for

A.2 The Existence of η_ℓ

In this section, we show the existence of a unique solution of η_ℓ to $g(\eta_\ell) = \beta_\ell$. As our experimental results are based on the cases when $c = 1$ and $c = \infty$, we only discuss these two cases in this section.

When $c = 1$, we have

$$g(\eta_\ell) = \frac{1}{2}\left(1 + 2\eta_\ell\right)^{-(n_0-1)/2}.$$

Then we have

$$\frac{\partial g(\eta_\ell)}{\partial \eta_\ell} = -\frac{(n_0-1)}{2}(1 + 2\eta_\ell)^{-(n_0+1)/2} < 0,$$

for all $\eta_\ell > 0$. We also know that $\lim_{\eta_\ell \to \infty} g(\eta_\ell) = 0$ and $g(0) = \frac{1}{2}$. Therefore, $g(\eta_\ell) = \beta_\ell$ has a unique solution $\eta_\ell = \frac{1}{2}\left[(2\beta_\ell)^{-2/(n_0-1)} - 1\right]$ for all $\beta_\ell \in (0, \frac{1}{2}]$.

When $c = \infty$, we use $f(x)$ to denote the probability density function of a chi-squared distribution with $n_0 - 1$ degrees of freedom. Then, by taking derivative with respect to η_ℓ, we have

$$\begin{aligned}
\frac{\partial}{\partial \eta_\ell} g(\eta_\ell) &= \frac{\partial}{\partial \eta_\ell}\int_0^\infty \frac{f(x)}{1+\exp(2\eta_\ell x)}dx \\
&= \int_0^\infty \frac{\partial}{\partial \eta_\ell}\frac{f(x)}{1+\exp(2\eta_\ell x)}dx \\
&= -\int_0^\infty \frac{2x\exp(2\eta_\ell x) f(x)}{\left[1+\exp(2\eta_\ell x)\right]^2}dx \\
&\le -\mathrm{E}\left[\frac{2\chi^2_{n_0-1}\exp(2\eta_\ell \chi^2_{n_0-1})}{\left[1+\exp(2\eta_\ell \chi^2_{n_0-1})\right]^2}\right] \\
&< 0.
\end{aligned}$$

The second equality holds due to the fact that

$$\left|\frac{\partial}{\partial \eta_\ell}\frac{f(x)}{1+\exp(2\eta_\ell x)}\right| = \left|\frac{2x\exp(2\eta_\ell x)f(x)}{\left[1+\exp(2\eta_\ell x)\right]^2}\right| = 2xf(x)\frac{\exp(2\eta_\ell x)}{\left[1+\exp(2\eta_\ell x)\right]^2} \le 2xf(x), \text{ for } x, \eta_\ell \ge 0,$$

and $\int_0^\infty 2xf(x)dx = 2\mathrm{E}[\chi^2_{n_0-1}] = 2(n_0-1) < \infty$ (Billingsley, 1986). This means that $g(\eta_\ell)$ is decreasing when $c = \infty$. However, $g(0) = 1$ and

$$\lim_{\eta_\ell \to \infty} g(\eta_\ell) = \lim_{\eta_\ell \to \infty} \int_0^\infty \frac{f(x)}{1+\exp(2\eta_\ell x)} dx = \int_0^\infty \lim_{\eta_\ell \to \infty} \frac{f(x)}{1+\exp(2\eta_\ell x)} dx = 0,$$

where the second equality holds due to bounded convergence theorem. Therefore, $g(\theta_\ell) = \beta_\ell$ has a unique solution for all $\beta_\ell \in (0, 1]$. Thus, a simple search method such as the bi-section search will find the unique η_ℓ value when $c = \infty$.

A.3 Proof of Lemma 1 for $c = \infty$

We prove Lemma 1 when $c = \infty$ in this section. We first present the following lemma that is useful for our proof.

Lemma A1. *(Karlin and Taylor, 1975, Theorem 7.5.2) Let* $\{\mathcal{B}(t, \Delta, \sigma^2, x) : t \geq 0\}$ *be a Brownian motion process with drift* $\Delta \neq 0$*, variance* σ^2*, and the starting point* x *when* $t = 0$*. The probability that the process reaches the level* $a > x$ *before hitting* $-a < x$ *is given by*

$$\Pr\{\mathcal{B}(T, \Delta, \sigma^2, x) = a\} = \frac{\exp(-2\Delta x/\sigma^2) - \exp(2\Delta a/\sigma^2)}{\exp(-2\Delta a/\sigma^2) - \exp(2\Delta a/\sigma^2)},$$

where $T = \min\{t : \mathcal{B}(t, \Delta, \sigma^2, x) \notin (-a, a)\}$*, i.e. the first time when the drifted Brownian motion hits* $-a$ *or* a*.*

We now prove Lemma 1 when $c = \infty$.

Proof. Consider system i and constraint ℓ with mean $y_{i\ell}$ and threshold value q_ℓ^m where $m = 1, \ldots, d_\ell$. Let ϵ_ℓ be the fixed tolerance level and $(-R, R)$ be the straight-line continuation region, where $R = (n_0 - 1)\eta_\ell S^2_{i\ell}(n_0)/\epsilon_\ell$.

Assume system i is unacceptable with respect to constraint ℓ for threshold q_ℓ^m, i.e., $y_{i\ell} \geq q_\ell^m + \epsilon_\ell$. Define T_d and T_c as follows:

$$T_d = \min\left\{t \in \mathbb{Z}^+, t \geq n_0 : \mathcal{B}\left(t, y_{i\ell} - q_\ell^m, \sigma^2_{i\ell}, 0\right) \notin (-R, R)\right\},$$

$$T_c = \min\left\{t \in \mathbb{R}^+, t \geq n_0 : \mathcal{B}\left(t, y_{i\ell} - q_\ell^m, \sigma^2_{i\ell}, 0\right) \notin (-R, R)\right\},$$

where $\mathbb{Z}^+$ and $\mathbb{R}^+$ denote the set of positive integers and the set of positive real numbers, respectively. That is we define T_d/T_c as the first integer/continuous passage time of the drifted Brownian motion $\mathcal{B}\left(t, y_{i\ell} - q_\ell^m, \sigma_{i\ell}^2, 0\right)$, respectively. Then we have

$$\begin{aligned}
\Pr\left(\mathrm{CD}_{i\ell}(q_\ell^m)\right) &= \Pr\left(\sum_{n=1}^{T_d}(Y_{i\ell n} - q_\ell^m) \geq R\right) = \mathrm{E}\left[\Pr\left(\sum_{n=1}^{T_d}\left(\frac{Y_{i\ell n} - q_\ell^m}{\sigma_{i\ell}}\right) \geq \frac{R}{\sigma_{i\ell}}\middle| S_{i\ell}^2(n_0)\right)\right] \\
&\geq \mathrm{E}\left[\Pr\left(\mathcal{B}\left(T_c, \frac{y_{i\ell} - q_\ell^m}{\sigma_{i\ell}}, 1, 0\right) \geq \frac{R}{\sigma_{i\ell}}\middle| S_{i\ell}^2(n_0)\right)\right] \\
&\geq \mathrm{E}\left[\Pr\left(\mathcal{B}\left(T_c, \frac{\epsilon_\ell}{\sigma_{i\ell}}, 1, 0\right) \geq \frac{R}{\sigma_{i\ell}}\middle| S_{i\ell}^2(n_0)\right)\right].
\end{aligned}$$

The first inequality holds because of the fact that the sample mean $\bar{Y}_{i\ell}(n_0)$ and sample variance $S_{i\ell}^2(n_0)$ of normal random variables are independent, and because observing at discrete time reduces the chance of error (Jennison, Johnstone and Turnbull, 1980). The second inequality holds due to the assumption that $y_{i\ell} - q_\ell^m \geq \epsilon_\ell$.

We then have the following derivation,

$$\begin{aligned}
&\mathrm{E}\left[\Pr\left(\mathcal{B}\left(T_c, \frac{\epsilon_\ell}{\sigma_{i\ell}}, 1, 0\right) \geq \frac{(n_0-1)\eta_\ell S_{i\ell}^2(n_0)}{\epsilon_\ell \sigma_{i\ell}}\middle| S_{i\ell}^2(n_0)\right)\right] \\
&= \mathrm{E}\left[\frac{1 - \exp\left(2\eta_\ell \frac{(n_0-1)S_{i\ell}^2(n_0)}{\sigma_{i\ell}^2}\right)}{\exp\left(-2\eta_\ell \frac{(n_0-1)S_{i\ell}^2(n_0)}{\sigma_{i\ell}^2}\right) - \exp\left(2\eta_\ell \frac{(n_0-1)S_{i\ell}^2(n_0)}{\sigma_{i\ell}^2}\right)}\right],
\end{aligned}$$

due to Lemma A1. By the fact that $\frac{(n_0-1)S_{i\ell}^2(n_0)}{\sigma_{i\ell}^2}$ follows a χ^2 distribution with $n_0 - 1$ degrees of freedom, we have

$$\begin{aligned}
&\mathrm{E}\left[\frac{1 - \exp\left(2\eta_\ell \frac{(n_0-1)S_{i\ell}^2(n_0)}{\sigma_{i\ell}^2}\right)}{\exp\left(-2\eta_\ell \frac{(n_0-1)S_{i\ell}^2(n_0)}{\sigma_{i\ell}^2}\right) - \exp\left(2\eta_\ell \frac{(n_0-1)S_{i\ell}^2(n_0)}{\sigma_{i\ell}^2}\right)}\right] \\
&= \mathrm{E}\left[\frac{1 - \exp(2\eta_\ell \chi_{n_0-1}^2)}{\exp(-2\eta_\ell \chi_{n_0-1}^2) - \exp(2\eta_\ell \chi_{n_0-1}^2)}\right] \\
&= \mathrm{E}\left[\frac{\left(1 - \exp(2\eta_\ell \chi_{n_0-1}^2)\right)\exp(2\eta_\ell \chi_{n_0-1}^2)}{\left(\exp(-2\eta_\ell \chi_{n_0-1}^2) - \exp(2\eta_\ell \chi_{n_0-1}^2)\right)\exp(2\eta_\ell \chi_{n_0-1}^2)}\right] \\
&= \mathrm{E}\left[\frac{\left(1 - \exp(2\eta_\ell \chi_{n_0-1}^2)\right)\exp(2\eta_\ell \chi_{n_0-1}^2)}{1 - \left[\exp(2\eta_\ell \chi_{n_0-1}^2)\right]^2}\right] \\
&= \mathrm{E}\left[\frac{\exp(2\eta_\ell \chi_{n_0-1}^2)}{1 + \exp(2\eta_\ell \chi_{n_0-1}^2)}\right] = 1 - \mathrm{E}\left[\frac{1}{1 + \exp(2\eta_\ell \chi_{n_0-1}^2)}\right] \\
&= 1 - \int_0^\infty \frac{1}{1 + \exp(2\eta_\ell x)} \times \frac{1}{2^{(n_0-1)/2}\Gamma((n_0-1)/2)} x^{(n_0-1)/2-1} e^{-x/2} dx = 1 - \beta_\ell,
\end{aligned}$$

where the last equality holds because of the definition of $g(\cdot)$ in (3) and the fact that η_ℓ is the solution to $g(\eta_\ell) = \beta_\ell$.

The above results also hold for $y_{i\ell} \leq q_\ell^m - \epsilon_\ell$. Finally, $\Pr\left(\mathrm{CD}_{i\ell}(q_\ell^m)\right) = 1 \geq 1 - \beta_\ell$ when $q_\ell^m - \epsilon_\ell < y_{i\ell} < q_\ell^m + \epsilon_\ell$. Hence, when $c = \infty$, Lemma 1 follows. □

A.4 Algorithm Statement and Proof of Statistical Validity for the Recycle$^{\mathcal{B}}$ Procedure

We discuss the statistical validity of the Recycle$^{\mathcal{B}}$ procedure in this section. The full description of Recycle$^{\mathcal{B}}$ is provided in Algorithm 13. The Recycle$^{\mathcal{B}}$ procedure is essentially same as the $\mathcal{RF}$ procedure except that β_ℓ is defined differently and we do not need to keep track of $v_{i\ell}^{\mathrm{LB}}, v_{i\ell}^{\mathrm{UB}}$, and $\mathrm{LAST}_{i\ell}$ (because Recycle$^{\mathcal{B}}$ is not designed for adding threshold values later).

Algorithm 13 Procedure Recycle$^{\mathcal{B}}$

[**Setup**:] Choose confidence level $1-\alpha$, tolerance level ϵ_ℓ, and thresholds $\left\{q_\ell^1, q_\ell^2, \ldots, q_\ell^{d_\ell}\right\}$ for constraint $\ell = 1, 2, \ldots, s$. Also, choose the value of c and set $\Theta = \{1, 2, \ldots, k\}$. For $\ell = 1, \ldots, s$, set η_ℓ such that $g(\eta_\ell) = \beta_\ell$, where β satisfies (4), and either

(i) $\beta_\ell = \beta/(s \cdot d_\ell)$ for $\ell = 1, \ldots, s$, or

(ii) $\beta_\ell = \beta/D$ and $D = \sum_{\ell=1}^{s} d_\ell$ for $\ell = 1, \ldots, s$.

for each system $i \in \Theta$ **do**

[**Initialization**:]

- Obtain n_0 observations $Y_{i\ell 1}, Y_{i\ell 2}, \ldots, Y_{i\ell n_0}$ for $\ell = 1, 2, \ldots, s$.
- Compute $S_{i\ell}^2(n_0)$ for $\ell = 1, 2, \ldots, s$.
- Set $r_i = n_0$, $\mathrm{ON} = \{1, 2, \ldots, s\}$, and $\mathrm{ON}_\ell = \{1, 2, \ldots, d_\ell\}$ for $\ell = 1, 2, \ldots, s$.

[**Feasibility Check**:]

for $\ell \in \mathrm{ON}$ **do**

for $m \in \mathrm{ON}_\ell$ **do**

If $\sum_{n=1}^{r_i}(Y_{i\ell n} - q_\ell^m) \geq R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))$, set $Z_{i\ell}^m = 0$ and $\mathrm{ON}_\ell = \mathrm{ON} \backslash \{m\}$.

If $\sum_{n=1}^{r_i}(Y_{i\ell n} - q_\ell^m) \leq -R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))$, set $Z_{i\ell}^m = 1$ and $\mathrm{ON}_\ell = \mathrm{ON} \backslash \{m\}$.

end for

If $\mathrm{ON}_\ell = \emptyset$, set $\mathrm{ON} = \mathrm{ON} \setminus \{\ell\}$.

end for

[**Stopping Condition**:]

- If $\mathrm{ON} = \emptyset$, return $Z_{i\ell}^m$ for $\ell = 1, 2, \ldots, s$ and $m = 1, 2, \ldots, d_\ell$.
- Otherwise, set $r_i = r_i + 1$, take one additional observation $Y_{i\ell r_i}$ for $\ell \in \mathrm{ON}$, and go to [**Feasibility Check**].

end for

We now show the statistical validity of the Recycle$^{\mathcal{B}}$ procedure in the following theorem.

Theorem A1. *Procedure Recycle*$^{\mathcal{B}}$ *guarantees* PCD $\geq 1-\alpha$.

Proof. We prove the theorem based on whether systems are simulated independently or with correlation.

If the systems are simulated independently (i.e., no CRN), then the Bonferroni inequality yields

$$\text{PCD} = \Pr\left(\cap_{i=1}^{k}\cap_{\ell=1}^{s}\cap_{m=1}^{d_\ell}\text{CD}_{i\ell}(q_\ell^m)\right) \geq \prod_{i=1}^{k}\left[1-\sum_{\ell=1}^{s}\sum_{m=1}^{d_\ell}\Pr(\text{ICD}_{i\ell}(q_\ell^m))\right].$$

With β_ℓ from the first choice (i) of Algorithm 13, Lemma 1 and equation (4) yield

$$\begin{aligned}
\text{PCD} &\geq \prod_{i=1}^{k}\left(1-\sum_{\ell=1}^{s}\sum_{m=1}^{d_\ell}\beta_\ell\right)\\
&= \prod_{i=1}^{k}\left(1-\sum_{\ell=1}^{s}\sum_{m=1}^{d_\ell}\frac{\beta}{sd_\ell}\right)\\
&= \prod_{i=1}^{k}\left(1-\sum_{\ell=1}^{s}\frac{\beta}{s}\right) = \prod_{i=1}^{k}(1-\beta)\\
&= \prod_{i=1}^{k}\left(1-\left(1-(1-\alpha)^{1/k}\right)\right) = 1-\alpha.
\end{aligned}$$

With β_ℓ from the second choice (ii) of Algorithm 13, Lemma 1 and equation (4) yield

$$\begin{aligned}
\text{PCD} &\geq \prod_{i=1}^{k}\left(1-\sum_{\ell=1}^{s}\sum_{m=1}^{d_\ell}\frac{\beta}{D}\right)\\
&= \prod_{i=1}^{k}\left(1-\sum_{\ell=1}^{s}\frac{\beta}{D}d_\ell\right) = \prod_{i=1}^{k}(1-\frac{\beta}{D}\sum_{\ell=1}^{s}d_\ell)\\
&= \prod_{i=1}^{k}(1-\beta) = \prod_{i=1}^{k}\left(1-\left(1-(1-\alpha)^{1/k}\right)\right) = 1-\alpha.
\end{aligned}$$

If the systems are simulated with correlation, then the Bonferroni inequality yields

$$\text{PCD} = \Pr\left(\cap_{i=1}^{k}\cap_{\ell=1}^{s}\cap_{m=1}^{d_\ell}\text{CD}_{i\ell}(q_\ell^m)\right) \geq 1-\sum_{i=1}^{k}\sum_{\ell=1}^{s}\sum_{m=1}^{d_\ell}\Pr(\text{ICD}_{i\ell}(q_\ell^m)).$$

If β_ℓ is set based on the first choice (i), then Lemma 1 and equation (4) yield

$$\begin{aligned}
\text{PCD} &\geq 1 - \sum_{i=1}^{k}\sum_{\ell=1}^{s}\sum_{m=1}^{d_\ell}\beta_\ell \\
&= 1 - \sum_{i=1}^{k}\sum_{\ell=1}^{s}\sum_{m=1}^{d_\ell}\frac{\beta}{sd_\ell} \\
&= 1 - \sum_{i=1}^{k}\sum_{\ell=1}^{s}\frac{\beta}{s} \\
&= 1 - \sum_{i=1}^{k}\beta \\
&= 1 - k\beta = 1 - k(\alpha/k) = 1 - \alpha.
\end{aligned}$$

If β_ℓ is set based on the second choice (ii), then Lemma 1 and equation (4) yield

$$\begin{aligned}
\text{PCD} &\geq 1 - \sum_{i=1}^{k}\sum_{\ell=1}^{s}\sum_{m=1}^{d_\ell}\beta_\ell \\
&= 1 - \sum_{i=1}^{k}\sum_{\ell=1}^{s}\sum_{m=1}^{d_\ell}\frac{\beta}{D} \\
&= 1 - \sum_{i=1}^{k}\frac{\beta}{D}\left(\sum_{\ell=1}^{s}d_\ell\right) \\
&= 1 - \sum_{i=1}^{k}\beta \\
&= 1 - k\beta = 1 - k(\alpha/k) = 1 - \alpha. \qquad \square
\end{aligned}$$

A.5 Algorithm Statement and Proof of Statistical Validity for the Restart$^{\text{prod}}$ *Procedure*

In this section, we provide the full description of the Restart$^{\text{prod}}$ procedure and discuss its statistical validity. The full description of Restart$^{\text{prod}}$ is presented in Algorithm 14.

Notice that Restart$^{\text{prod}}$ performs feasibility check for each combination of thresholds and each combination has one threshold on each constraint. Thus, unlike $\mathcal{RF}$, Recycle$^{\mathcal{B}}$, and Restart$^{\text{sum}}$, Restart$^{\text{prod}}$ has only one way of setting the implementation parameter β_ℓ for all $\ell = 1, \ldots, s$.

We now show the statistical validity of the Restart$^{\text{prod}}$ by proving the following theorem.

Theorem A2. *Procedure Restart*$^{\text{prod}}$ *guarantees* PCD $\geq 1 - \alpha$.

Algorithm 14 Procedure Restart$^{\text{prod}}$

[**Setup**:] Choose confidence level $1-\alpha$, tolerance level ϵ_ℓ, and set $D = \prod_{\ell=1}^{s} d_\ell$. Set threshold vectors $\boldsymbol{q}^1, \boldsymbol{q}^2, \ldots, \boldsymbol{q}^D$ as all the possible combinations of thresholds on all constraints, where $\boldsymbol{q}^d$ denotes the dth combination of the thresholds and $d = 1, \ldots, D$. Also, choose the value of c and set $\Theta = \{1, 2, \ldots, k\}$. For $\ell = 1, \ldots, s$, set η_ℓ such that $g(\eta_\ell) = \beta_\ell$, where

$$\beta_\ell = \begin{cases} [1-(1-\alpha)^{1/(kD)}]/s & \text{when systems are independent,} \\ [1-(1-\alpha)^{1/D}]/(ks) & \text{when systems are dependent.} \end{cases}$$

.

for $d = 1, \ldots, D$ **do**

  for each system $i \in \Theta$ **do**

    [**Initialization**:]

    - Obtain n_0 observations $Y_{i\ell 1}^d, Y_{i\ell 2}^d, \ldots, Y_{i\ell n_0}^d$ for $\ell = 1, \ldots, s$.
    - Compute the sample variance of $Y_{i\ell 1}^d, Y_{i\ell 2}^d, \ldots, Y_{i\ell n_0}^d$ as $S_{i\ell d}^2(n_0)$ for $\ell = 1, \ldots, s$.
    - Set $r_i = n_0$ and $\text{ON} = \{1, \ldots, s\}$.

    [**Feasibility Check**:]

    for $\ell \in \text{ON}$ **do**

      If $\sum_{n=1}^{r_i}(Y_{i\ell n}^d - q_\ell^{m_\ell}) \geq R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell d}^2(n_0))$, set $Z_{i\ell}^d = 0$ and $\text{ON} = \text{ON} \backslash \{\ell\}$;

      Else if $\sum_{n=1}^{r_i}(Y_{i\ell n}^d - q_\ell^{m_\ell}) \leq -R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell d}^2(n_0))$, set $Z_{i\ell}^d = 1$ and $\text{ON} = \text{ON} \backslash \{\ell\}$.

    end for

    [**Stopping Condition:**]

    If $\text{ON} = \emptyset$, return $Z_{i\ell}^d$ for $\ell = 1, \ldots, s$. Otherwise, set $r_i = r_i + 1$, take one additional observation $Y_{i\ell r_i}$ for $\ell \in \text{ON}$ and go to [**Feasibility Check**].

  end for

end for

Proof. We let $\boldsymbol{q}^d = (q_1^{m_1^d}, \ldots, q_s^{m_s^d})$, where $d = 1, \ldots, D$, denote the dth combination of the thresholds, where m_ℓ^d is the index of the threshold on constraint ℓ in this combination of thresholds for $\ell = 1, \ldots, s$. Then for one particular threshold vector $\boldsymbol{q}^d$, we have the probability of correct decision for system i as

$$\Pr\left(\cap_{\ell=1}^{s} \text{CD}_{i\ell}(q_\ell^{m_\ell^d})\right) \geq 1 - \sum_{\ell=1}^{s} \beta_\ell, \tag{12}$$

where the inequality holds due to the Bonferroni inequality and Lemma 1. Restart$^{\text{prod}}$ performs feasibility check for each combination of thresholds independently. We then consider the cases when systems are simulated independently or with correlation.

If the systems are simulated independently, then equation (12) yields

$$\begin{aligned}
\mathrm{PCD} &= \Pr\left(\cap_{d=1}^{D} \cap_{i=1}^{k} \cap_{\ell=1}^{s} \mathrm{CD}_{i\ell}(q_\ell^{m_\ell^d})\right) \geq \prod_{d=1}^{D}\prod_{i=1}^{k}\left(1-\sum_{\ell=1}^{s}\beta_\ell\right) \\
&= \prod_{d=1}^{D}\prod_{i=1}^{k}\left(1-\sum_{\ell=1}^{s}\left[1-(1-\alpha)^{1/(kD)}\right]/s\right) = \prod_{d=1}^{D}\prod_{i=1}^{k}(1-\alpha)^{1/(kD)} \\
&= 1-\alpha.
\end{aligned}$$

If the systems are simulated with CRN, then the Bonferroni inequality and Lemma 1 yields

$$\begin{aligned}
\mathrm{PCD} &= \Pr\left(\cap_{d=1}^{D} \cap_{i=1}^{k} \cap_{\ell=1}^{s} \mathrm{CD}_{i\ell}(q_\ell^{m_\ell^d})\right) = \prod_{d=1}^{D} \Pr\left(\cap_{i=1}^{k} \cap_{\ell=1}^{s} \mathrm{CD}_{i\ell}(q_\ell^{m_\ell^d})\right) \\
&\geq \prod_{d=1}^{D}\left[1-\sum_{i=1}^{k}\sum_{\ell=1}^{s}\Pr(\mathrm{ICD}_{i\ell}(q_\ell^{m_\ell^d}))\right] = \prod_{d=1}^{D}\left[1-\sum_{i=1}^{k}\sum_{\ell=1}^{s}\beta_\ell\right] \\
&= \prod_{d=1}^{D}\left(1-\sum_{i=1}^{k}\sum_{\ell=1}^{s}\left[1-(1-\alpha)^{1/D}\right]/(ks)\right) = 1-\alpha. \qquad \square
\end{aligned}$$

A.6 Algorithm Statement and Proof of Statistical Validity for the $\mathrm{Restart}^{\mathrm{sum}}$ *Procedure*

In this section, we provide the full description of the $\mathrm{Restart}^{\mathrm{sum}}$ procedure and discuss its statistical validity. $\mathrm{Restart}^{\mathrm{sum}}$ is motivated by the following considerations (relative to $\mathrm{Restart}^{\mathrm{prod}}$). $\mathrm{Restart}^{\mathrm{sum}}$ makes feasibility decisions iteratively for each threshold value of each constraint, while $\mathrm{Restart}^{\mathrm{prod}}$ may make multiple decisions for each such threshold value (because $\mathrm{Restart}^{\mathrm{prod}}$ determines feasibility for each combination of threshold values of all constraints). Therefore, $\mathrm{Restart}^{\mathrm{sum}}$ usually performs fewer restarts than $\mathrm{Restart}^{\mathrm{prod}}$, and thus usually needs fewer observations compared with $\mathrm{Restart}^{\mathrm{prod}}$. The full description of $\mathrm{Restart}^{\mathrm{sum}}$ is provided in Algorithm 15.

We show the statistical validity of the $\mathrm{Restart}^{\mathrm{sum}}$ procedure based on the following theorem.

Theorem A3. *Procedure* $\mathrm{Restart}^{\mathrm{sum}}$ *guarantees* $\mathrm{PCD} \geq 1-\alpha$.

Proof. For the $\mathrm{Restart}^{\mathrm{sum}}$ procedure, as we restart a feasibility check for each system, each constraint, and each threshold value, the $\mathrm{CD}_{i\ell}(q_\ell^m)$ events are independent for all

Algorithm 15 Procedure Restart$^{\text{sum}}$

[**Setup**:] Choose confidence level $1-\alpha$, tolerance level ϵ_ℓ, and thresholds $\left\{q_\ell^1, q_\ell^2, \ldots, q_\ell^{d_\ell}\right\}$ for constraint $\ell = 1, 2, \ldots, s$. Also, choose the value of c and set $\Theta = \{1, 2, \ldots, k\}$. Set $D = \sum_{\ell=1}^{s} d_\ell$ and for $\ell = 1, \ldots, s$, set η_ℓ such that $g(\eta_\ell) = \beta_\ell$, where either

(i)

$$\beta_\ell = \begin{cases} 1 - \sqrt[k \cdot s d_\ell]{1-\alpha}, & \text{when systems are independent,} \\ (1 - \sqrt[s d_\ell]{1-\alpha})/k, & \text{when systems are dependent,} \end{cases}$$

or

(ii)

$$\beta_\ell = \begin{cases} 1 - \sqrt[k \cdot D]{1-\alpha}, & \text{when systems are independent,} \\ (1 - \sqrt[D]{1-\alpha})/k, & \text{when systems are dependent.} \end{cases}$$

for each system $i \in \Theta$ **do**
  for $\ell = 1, \ldots, s$ **do**
    for each threshold $m = 1, \ldots, d_\ell$ **do**
      [**Initialization**:]

- Obtain n_0 observations $Y_{i\ell 1}^m, Y_{i\ell 2}^m, \ldots, Y_{i\ell n_0}^m$.
- Compute the sample variance of $Y_{i\ell 1}^m, Y_{i\ell 2}^m, \ldots, Y_{i\ell n_0}^m$ as $S_{i\ell m}^2(n_0)$.
- Set $r_i = n_0$.

      [**Feasibility Check**:]
      If $\sum_{n=1}^{r_i}(Y_{i\ell n}^m - q_\ell^{m_\ell}) \geq R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell m}^2(n_0))$, return $Z_{i\ell}^m = 0$;
      Else if $\sum_{n=1}^{r_i}(Y_{i\ell n}^m - q_\ell^{m_\ell}) \leq -R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell m}^2(n_0))$, return $Z_{i\ell}^m = 1$;
      Else, set $r_i = r_i + 1$, take one additional observation $Y_{i\ell r_i}^m$ and go to [**Feasibility Check**].
    end for
  end for
end for

$\ell = 1, \ldots, s$ and $m = 1, \ldots, d_\ell$. We prove the theorem based on whether systems are simulated independently or with correlation.

If the systems are simulated independently, then

$$\text{PCD} = \Pr\left(\cap_{i=1}^{k} \cap_{\ell=1}^{s} \cap_{m=1}^{d_\ell} \text{CD}_{i\ell}(q_\ell^m)\right) = \prod_{i=1}^{k} \prod_{\ell=1}^{s} \prod_{m=1}^{d_\ell} \Pr\left(\text{CD}_{i\ell}(q_\ell^m)\right).$$

With β_ℓ from the first choice (i) of Algorithm 15, Lemma 1 yields

$$\begin{aligned}
\text{PCD} &\geq \prod_{i=1}^{k} \prod_{\ell=1}^{s} \prod_{m=1}^{d_\ell} (1 - \beta_\ell) \\
&= \prod_{i=1}^{k} \prod_{\ell=1}^{s} (1 - \beta_\ell)^{d_\ell} = \prod_{i=1}^{k} \prod_{\ell=1}^{s} \left(1 - (1 - \sqrt[ksd_\ell]{1-\alpha})\right)^{d_\ell} \\
&= \prod_{i=1}^{k} \prod_{\ell=1}^{s} \sqrt[ks]{1-\alpha} = 1 - \alpha.
\end{aligned}$$

With β_ℓ from the second choice (ii) of Algorithm 15, Lemma 1 yields

$$\begin{aligned}
\text{PCD} &\geq \prod_{i=1}^{k} \prod_{\ell=1}^{s} \prod_{m=1}^{d_\ell} (1 - \beta_\ell) \\
&= \prod_{i=1}^{k} \prod_{d=1}^{D} \left(1 - (1 - \sqrt[kD]{1-\alpha})\right) \\
&= \left(\sqrt[kD]{1-\alpha}\right)^{kD} = 1 - \alpha.
\end{aligned}$$

If the systems are simulated with CRN, then the Bonferroni inequality yields

$$\text{PCD} = \Pr\left(\cap_{i=1}^{k} \cap_{\ell=1}^{s} \cap_{m=1}^{d_\ell} \text{CD}_{i\ell}(q_\ell^m)\right) \geq \prod_{\ell=1}^{s} \prod_{m=1}^{d_\ell} \left(1 - \sum_{i=1}^{k} \Pr(\text{ICD}_{i\ell}(q_\ell^m))\right).$$

With β_ℓ from the first choice (i) of Algorithm 15, Lemma 1 yields

$$\begin{aligned}
\text{PCD} &\geq \prod_{\ell=1}^{s} \prod_{m=1}^{d_\ell} \left(1 - \sum_{i=1}^{k} \beta_\ell\right) \\
&= \prod_{\ell=1}^{s} \prod_{m=1}^{d_\ell} \left(1 - (1 - \sqrt[sd_\ell]{1-\alpha})\right) = \prod_{\ell=1}^{s} \sqrt[s]{1-\alpha} = 1 - \alpha.
\end{aligned}$$

With β_ℓ from the second choice (ii) of Algorithm 15, Lemma 1 yields

$$\begin{aligned}
\text{PCD} &\geq \prod_{\ell=1}^{s} \prod_{m=1}^{d_\ell} \left(1 - \sum_{i=1}^{k} \beta_\ell\right) \\
&= \prod_{d=1}^{D} \left(1 - (1 - \sqrt[D]{1-\alpha})\right) \\
&= \left(\sqrt[D]{1-\alpha}\right)^{D} = 1 - \alpha.
\end{aligned}$$

□

A.7 *Algorithm Statement and Proof of Statistical Validity for the* $\text{Restart}^{\max}$ *Procedure*

In this section, we provide the full description of the $\text{Restart}^{\max}$ procedure and discuss its statistical validity. $\text{Restart}^{\max}$ performs feasibility checks for each threshold vector which is formed by choosing one threshold from each constraint in a pre-defined order. Once the feasibility of all thresholds on some constraints is determined, the threshold vector is formed based on the thresholds from the remaining constraints by omitting the constraints whose thresholds have all received feasibility decisions. The procedure terminates when all the thresholds on all the constraints have received their feasibility decisions. Therefore, $\text{Restart}^{\max}$ requires $\max_{\ell=1,\ldots,s} d_\ell$ restarts and performs better compared with $\text{Restart}^{\text{prod}}$ and $\text{Restart}^{\text{sum}}$. Similar to $\text{Restart}^{\text{prod}}$, each restart of $\text{Restart}^{\max}$ involves one threshold on each remaining constraint. Therefore, $\text{Restart}^{\max}$ only has one way of setting the implementation parameter β_m. The full description of $\text{Restart}^{\max}$ is provided in Algorithm 16.

We show the statistical validity of the $\text{Restart}^{\max}$ procedure in the following theorem.

Theorem A4. *Procedure* $\text{Restart}^{\max}$ *guarantees* $\text{PCD} \geq 1-\alpha$.

Proof. For the $\text{Restart}^{\max}$ procedure, as we restart a feasibility check for D threshold vectors, the $\text{CD}_{i\ell}(q_\ell^m)$ events are independent for all $m = 1,\ldots,d_\ell$ when i,ℓ are fixed. We prove the theorem based on whether systems are simulated independently or with correlation.

If the systems are simulated independently, then the Bonferroni inequality yields

$$\begin{aligned}
\text{PCD} &= \Pr\left(\cap_{m=1}^{D}\cap_{i=1}^{k}\cap_{\ell=1,d_\ell\geq m}^{s}\text{CD}_{i\ell}(q_\ell^m)\right) = \prod_{m=1}^{D}\prod_{i=1}^{k}\left(1-\sum_{\ell=1}^{s}\mathcal{I}(d_\ell \geq m)\cdot\Pr(\text{ICD}_{i\ell}(q_\ell^m))\right)\\
&\geq \prod_{m=1}^{D}\prod_{i=1}^{k}\left(1-\sum_{\ell=1}^{s}\mathcal{I}(d_\ell\geq m)\beta_m\right) = \prod_{m=1}^{D}\prod_{i=1}^{k}\left(1-\sum_{\ell=1}^{s}\mathcal{I}(d_\ell\geq m)\cdot\frac{1-\sqrt[k\cdot D]{1-\alpha}}{\sum_{\ell'=1}^{s}\mathcal{I}(d_{\ell'}\geq m)}\right)\\
&= \prod_{m=1}^{D}\prod_{i=1}^{k}\sqrt[k\cdot D]{1-\alpha} = 1-\alpha.
\end{aligned}$$

If the systems are simulated with CRN, then the Bonferroni inequality yields

$$\text{PCD} = \Pr\left(\cap_{m=1}^{D}\cap_{i=1}^{k}\cap_{\ell=1,d_\ell\geq m}^{s}\text{CD}_{i\ell}(q_\ell^m)\right) \geq \prod_{m=1}^{D}\left(1-\sum_{i=1}^{k}\sum_{\ell=1}^{s}\mathcal{I}(d_\ell\geq m)\Pr(\text{ICD}_{i\ell}(q_\ell^m))\right)$$

Algorithm 16 Procedure Restart$^{\max}$

[**Setup**:] Choose confidence level $1-\alpha$, tolerance level ϵ_ℓ, and thresholds $\left\{q_\ell^1, q_\ell^2, \ldots, q_\ell^{d_\ell}\right\}$ for constraint $\ell = 1, 2, \ldots, s$. Also, choose the value of c and set $\Theta = \{1, 2, \ldots, k\}$. Set $D = \max_{\ell=1,\ldots,s} d_\ell$. For $m = 1, \ldots, D$, set η_m such that $g(\eta_m) = \beta_m$, where

$$\beta_m = \begin{cases} (1 - \sqrt[k\cdot D]{1-\alpha})/\sum_{\ell=1}^s \mathcal{I}(d_\ell \geq m), & \text{when systems are independent,} \\ (1 - \sqrt[D]{1-\alpha})/(k\sum_{\ell=1}^s \mathcal{I}(d_\ell \geq m)), & \text{when systems are dependent.} \end{cases} \tag{13}$$

for $m = 1, \ldots, D$ **do**

[**Initialization**:]

- Obtain n_0 observations $Y_{i\ell 1}^m, Y_{i\ell 2}^m, \ldots, Y_{i\ell n_0}^m$ for $\ell = 1, \ldots, s$ with $d_\ell \geq m$.
- Compute the sample variance of $Y_{i\ell 1}^m, Y_{i\ell 2}^m, \ldots, Y_{i\ell n_0}^m$ as $S_{i\ell m}^2(n_0)$ for $\ell = 1, \ldots, s$ with $d_\ell \geq m$.
- Set $r_i = n_0$ and $\text{ON} = \{\ell = 1, \ldots, s | d_\ell \geq m\}$.

for each system $i \in \Theta$ **do**

[**Feasibility Check**:]

for $\ell \in \text{ON}$ **do**

If $\sum_{n=1}^{r_i}(Y_{i\ell n}^m - q_\ell^m) \geq R(r_i; \epsilon_\ell, \eta_m, S_{i\ell m}^2(n_0))$, set $Z_{i\ell}^m = 0$ and $\text{ON} = \text{ON}\backslash\{\ell\}$;

Else if $\sum_{n=1}^{r_i}(Y_{i\ell n}^m - q_\ell^m) \leq -R(r_i; \epsilon_\ell, \eta_m, S_{i\ell m}^2(n_0))$, set $Z_{i\ell}^m = 1$ and $\text{ON} = \text{ON}\backslash\{\ell\}$.

end for

[**Stopping Condition**:]

If $\text{ON} = \emptyset$, return $Z_{i\ell}^m$ for $\ell = 1, \ldots, s$ with $m \leq d_\ell$. Otherwise, set $r_i = r_i + 1$, take one additional observation $Y_{i\ell r_i}^m$ and go to [**Feasibility Check**].

end for

end for

$$\geq \prod_{m=1}^{D}\left(1-\sum_{i=1}^{k}\sum_{\ell=1}^{s}\mathcal{I}(d_\ell \geq m)\beta_m\right) = \prod_{m=1}^{D}\left(1-\sum_{i=1}^{k}\sum_{\ell=1}^{s}\mathcal{I}(d_\ell \geq m)\cdot\frac{1-\sqrt[D]{1-\alpha}}{k\sum_{\ell'=1}^{s}\mathcal{I}(d_{\ell'}\geq m)}\right)$$

$$= \prod_{m=1}^{D}\left(1-k\frac{1-\sqrt[D]{1-\alpha}}{k}\right) = 1-\alpha. \qquad \square$$

APPENDIX B

In this section, we provide discussions about the algorithm description of procedures Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ (discussed in Chapter 3) and their statistical validity in Appendices B.1 and B.2, respectively.

B.1 Procedures Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{HAK}}$

In this section, we discuss the algorithms of Restart$^{\mathcal{AK}}$ and Restart$^{\mathcal{HAK}}$ and their statistical validity.

As Restart$^{\mathcal{AK}}$ is essentially a special case of Restart$^{\mathcal{HAK}}$ when the number of constraints in consideration is one, therefore, we omit the discussion on the algorithm statement and the statistical validity of procedure Restart$^{\mathcal{AK}}$ for the sake of space.

Procedure Restart$^{\mathcal{HAK}}$ perform $\mathcal{HAK}$ for threshold vectors $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(\theta^*)}$ independently. A detailed description of algorithm is shown in Algorithm 17.

Algorithm 17 Procedure Restart$^{\mathcal{HAK}}$

[**Setup**:] Select the overall nominal confidence level $1-\alpha$. Choose tolerance levels $\epsilon_1, \ldots, \epsilon_s$, indifference-zone parameter δ, threshold vectors $\{\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(d)}\}$. Set $\alpha' = 1-(1-\alpha)^{1/d}$.

for $\theta = 1, \ldots, d$ **do**

  [**Setup for** $\mathcal{HAK}$:] Same as in $\mathcal{HAK}$ except that α is replaced by α'.

  [**Initialization**], [**Feasibility Check**], [**Feasibility Stopping Rule**], [**Setup for Comparison**], [**Comparison**], and [**Comparison Stopping Rule**] are the same as in $\mathcal{HAK}$.

  [**Stopping Condition**:] If one system is found in [**Comparison Stopping Rule**], terminate the algorithm and select the system as the best. If no system is found in [**Feasibility Stopping Rule**] and $\theta = d$, declare no feasible system exists with respect to the given threshold vectors.

end for

As $\mathcal{HAK}$ is heuristic and Restart$^{\mathcal{HAK}}$ essentially applies $\mathcal{HAK}$ for threshold vectors $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(\theta^*)}$, we cannot prove the statistical validity of Restart$^{\mathcal{HAK}}$. However, if we consider a variation of $\mathcal{HAK}$, namely $\mathcal{HAK}^{\mathcal{R}}$ ("restart"), with the following two changes in the [**Setup for Comparison**], we are able to prove the statistical validity of procedure Restart$^{\mathcal{HAK}^{\mathcal{R}}}$ that implements $\mathcal{HAK}^{\mathcal{R}}$ for threshold vectors $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(\theta^*)}$:

- Instead of using the observations of the primary performance measure $X_{i1}, \ldots, X_{ir_i}$ collected from the [**Feasibility Check**], we perform a completely new simulation and collect $X_{i1}, \ldots, X_{in_0}$ for system $i \in F$, and compute $\bar{X}_i(n_0)$ and $S^2_{X_{ij}}(n_0)$ for $i, j \in F$.
- Change $\beta_2 = \alpha_2/(|F|-1)$ to $\beta_2 = \begin{cases} 1-(1-\alpha_2)^{1/(k-1)}, & \text{if systems are simulated independently;} \\ \alpha_2/(k-1), & \text{if systems are simulated under CRN.} \end{cases}$

Note that [5] use F to denote the set of systems that are declared feasible with respect to $\boldsymbol{q}^{(\theta)}$ from Phase I.

Before we prove the statistical validity of Restart$^{\mathcal{HAK}^{\mathcal{R}}}$, we first present the following two lemmas.

Lemma B1. *For system i and constraint ℓ with threshold q_ℓ, the* [**Feasibility Check**] *steps in $\mathcal{HAK}^{\mathcal{R}}$ ensure* $\Pr(\mathrm{CD}_{i\ell}(q_\ell)) \geq 1-\beta_1$.

Lemma B2. *Given i such that $x_i \leq x_{[b]} - \delta$, the* [**Comparison**] *steps for system i and $[b]$ in $\mathcal{HAK}^{\mathcal{R}}$ ensure*

$$\Pr\left(\mathrm{CS}_i\right) \geq 1-\beta_2.$$

The proofs of Lemmas B1 and B2 are essentially same as those of Lemmas 7 and 10 becase α'_f (α'_c) from $\mathcal{ZAK}^{\mathcal{R}}$ and β_1 (β_2) from $\mathcal{HAK}^{\mathcal{R}}$ both denote the nominal error of feasibility check for one constraint of one single system with a fixed threshold (comparison between an inferior system and the best system $[b]$).

We then prove the statistical validity of Restart$^{\mathcal{HAK}^{\mathcal{R}}}$ in the following theorem.

Theorem B1. *Under Assumptions 2 and 3, the procedure* Restart$^{\mathcal{HAK}^{\mathcal{R}}}$ *guarantees*

$$\Pr\{\mathrm{CS}\} \geq 1-\alpha.$$

Proof. As Restart$^{\mathcal{HAK}^{\mathcal{R}}}$ implements $\mathcal{HAK}^{\mathcal{R}}$ for thresholds $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(d)}$, for each threshold vector $\boldsymbol{q}^{(\theta)}$, we let $S_d^{(\theta)}, S_a^{(\theta)}$, and $S_u^{(\theta)}$ denote the set of desirable, acceptable, and unacceptable systems with respect to $\boldsymbol{q}^{(\theta)}$, and consider the following events.

$$\begin{aligned} \mathcal{A}^{(\theta)} &= \text{ all systems in } S_u^{(\theta)} \text{ are declared infeasible with respect to } \boldsymbol{q}^{(\theta)}; \\ \mathcal{B}_1^{(\theta)} &= \text{ system } [b] \text{ is declared feasible to } \boldsymbol{q}^{(\theta)}; \\ \mathcal{B}_2^{(\theta)} &= \text{ system } [b] \text{ is selected as the best system among systems in } S_d^{(\theta)} \cup S_a^{(\theta)}. \end{aligned}$$

We also let $\mathrm{CS}^{(\theta)}$ denote the event of selecting a desirable or acceptable system with respect to threshold vector $\boldsymbol{q}^{(\theta)}$ whose primary performance measure is no worse than δ away from that of system $[b]$. To ensure $\mathrm{CS}^{(\theta)}$, we need to ensure all $\mathcal{A}^{(\theta)}, \mathcal{B}_1^{(\theta)}$, and $\mathcal{B}_2^{(\theta)}$ if system $[b]$ is desirable with respect to $\boldsymbol{q}^{(\theta)}$. If system $[b]$ is not a desirable system with respect to $\boldsymbol{q}^{(\theta)}$, $\mathcal{B}_1^{(\theta)}$ and $\mathcal{B}_2^{(\theta)}$ are not defined and we only need to ensure $\mathcal{A}^{(\theta)}$. That is

$$\Pr\left(\mathrm{CS}^{(\theta)}\right) \geq \begin{cases} \Pr\left(\mathcal{A}^{(\theta)} \cap \mathcal{B}_1^{(\theta)} \cap \mathcal{B}_2^{(\theta)}\right), & \text{if system } b \text{ is desirable with respect to } \boldsymbol{q}^{(\theta)} \\ \Pr\left(\mathcal{A}^{(\theta)}\right), & \text{if system } b \text{ is not desirable with respect to } \boldsymbol{q}^{(\theta)} \end{cases}$$

As $\mathrm{CS}^{(\theta)}$ achieves its lower bound when system $[b]$ is desirable with respect to $\boldsymbol{q}^{(\theta)}$, we only focus on this case. One may also notice that $\mathcal{A}^{(\theta)}$ and $\mathcal{B}_1^{(\theta)}$ are independent if systems are simulated independently and are dependent if systems are simulated under CRN. As we discard observations from Phase I and completely restart for Phase II in $\mathcal{HAK}^{\mathcal{R}}$, and $\mathcal{B}_2^{(\theta)}$ involves making the correct selection from all systems in $S_a^{(\theta)} \cup S_d^{(\theta)}$, $\mathcal{B}_2^{(\theta)}$ is independent from $\mathcal{A}^{(\theta)}$ and $\mathcal{B}_1^{(\theta)}$. Then, we have

$$\Pr\left(\mathrm{CS}^{(\theta)}\right) \geq \begin{cases} \Pr\left(\mathcal{A}^{(\theta)}\right) \times \Pr\left(\mathcal{B}_1^{(\theta)}\right) \times \Pr\left(\mathcal{B}_2^{(\theta)}\right), & \text{if systems are simulated indepedently} \\ \left[\Pr\left(\mathcal{A}^{(\theta)}\right) + \Pr\left(\mathcal{B}_1^{(\theta)}\right) - 1\right] \times \Pr\left(\mathcal{B}_2^{(\theta)}\right), & \text{if systems are simulated under CRN.} \end{cases}$$

We let $j_u^{(\theta)}$ denote the number of unacceptable systems with respect to $\boldsymbol{q}^{(\theta)}$, i.e., $j_u^{(\theta)} = |S_u^{(\theta)}|$.

We then discuss the cases depending on whether systems are simulated independently or under CRN. When systems are simulated indepedently, by Lemma B1, we have

$$\begin{aligned} \Pr\left(\mathcal{A}^{(\theta)}\right) &\geq \Pr\left(\cap_{i \in S_u^{(\theta)}} \cap_{\ell=1}^{s} \mathrm{CD}_{i\ell}(q_\ell^{(\theta)})\right) = \prod_{i \in S_u^{(\theta)}} \Pr\left(\cap_{\ell=1}^{s} \mathrm{CD}_{i\ell}(q_\ell^{(\theta)})\right) \\ &= \left[\Pr\left(\cap_{\ell=1}^{s} \mathrm{CD}_{i\ell}(q_\ell^{(\theta)})\right)\right]^{j_u^{(\theta)}} \geq \left[1 - \sum_{\ell=1}^{s} \Pr\left(\mathrm{ICD}_{i\ell}(q_\ell^{(\theta)})\right)\right]^{j_u^{(\theta)}} \geq (1 - s\beta_1)^{j_u^{(\theta)}}; \\ \Pr\left(\mathcal{B}_1^{(\theta)}\right) &= \Pr\left(\cap_{\ell=1}^{s} \mathrm{CD}_{i\ell}(q_\ell^{(\theta)})\right) \geq 1 - \sum_{\ell=1}^{s} \Pr\left(\mathrm{ICD}_{i\ell}(q_\ell^{(\theta)})\right) \geq 1 - s\beta_1, \end{aligned}$$

where we use $\mathrm{ICD}_{i\ell}(q_\ell^m)$ to denote the event of incorrect feasibility decision with respect to q_ℓ^m.

We use the same notation of N_{ij} as defined in the proof of Theorem 1.

$$\begin{aligned}
\Pr(\mathcal{B}_2^{(\theta)}) &\geq \Pr\left(\cap_{i\in\left(S_a^{(\theta)}\cup S_d^{(\theta)}\right),i\neq b}\mathrm{CS}_i\right)\\
&= \mathrm{E}\left[\Pr\left\{\cap_{i\in\left(S_a^{(\theta)}\cup S_d^{(\theta)}\right),i\neq b}\mathrm{CS}_i\middle| X_{[b]1},\ldots,X_{[b],N_{[b]}},S^2_{X_{i[b]}}(n_0)\right\}\right]\\
&= \mathrm{E}\left[\prod_{i\in\left(S_a^{(\theta)}\cup S_d^{(\theta)}\right),i\neq b}\Pr\left\{\mathrm{CS}_i\middle| X_{[b]1},\ldots,X_{[b],N_{[b]}},S^2_{X_{i[b]}}(n_0)\right\}\right]\\
&\geq \prod_{i\in\left(S_a^{(\theta)}\cup S_d^{(\theta)}\right),i\neq b}\mathrm{E}\left[\Pr\left\{\mathrm{CS}_i\middle| X_{[b]1},\ldots,X_{[b],N_{[b]}},S^2_{X_{i[b]}}(n_0)\right\}\right]\\
&= \prod_{i\in\left(S_a^{(\theta)}\cup S_d^{(\theta)}\right),i\neq b}\Pr\left\{\mathrm{CS}_i\right\} \geq \prod_{i\in\left(S_a^{(\theta)}\cup S_d^{(\theta)}\right),i\neq b}(1-\beta_2) \geq (1-\beta_2)^{k-j_u^{(\theta)}-1},
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\Pr\left(\mathrm{CS}^{(\theta)}\right) &\geq (1-s\beta_1)^{j_u^{(\theta)}+1}\times(1-\beta_2)^{k-j_u^{(\theta)}-1}\\
&\geq (1-s\beta_1)^k\times(1-\beta_2)^{k-1}\\
&= (1-\alpha_1)(1-\alpha_2) > 1-(\alpha_1+\alpha_2) = 1-\alpha'.
\end{aligned}$$

When systems are simulated under CRN, by Lemma B1, we have

$$\Pr\left(\mathcal{A}^{(\theta)}\right) \geq \Pr\left(\cap_{i\in S_u^{(\theta)}}\cap_{\ell=1}^{s}\mathrm{CD}_{i\ell}(q_\ell^{(\theta)})\right) \geq 1-\sum_{i\in S_u^{(\theta)}}\sum_{\ell=1}^{s}\mathrm{CD}_{i\ell}(q_\ell^{(\theta)}) \geq 1-j_u^{(\theta)}s\beta_1;$$

$$\Pr\left(\mathcal{B}_1^{(\theta)}\right) \geq 1-s\beta_1.$$

$$\Pr(\mathcal{B}_2^{(\theta)}) \geq \Pr\left(\cap_{i\in\left(S_a^{(\theta)}\cup S_d^{(\theta)}\right),i\neq b}\mathrm{CS}_i\right) \geq 1-\sum_{i\in\left(S_a^{(\theta)}\cup S_d^{(\theta)}\right),i\neq b}\Pr\left(\mathrm{ICS}_i\right) \geq 1-(k-j_u^{(\theta)}-1)\beta_2$$

Thus, we have

$$\begin{aligned}
\Pr\left(\mathrm{CS}^{(\theta)}\right) &\geq \left[1-(j_u^{(\theta)}+1)s\beta_1\right]\left[1-(k-j_u^{(\theta)}-1)\beta_2\right]\\
&\geq (1-ks\beta_1)\times(1-(k-1)\beta_2)\\
&= (1-\alpha_1)(1-\alpha_2) > 1-(\alpha_1+\alpha_2) = 1-\alpha'.
\end{aligned}$$

Thus, we see that $\Pr\left(\mathrm{CS}^{(\theta)}\right) \geq 1-\alpha'$ regardless whether systems are simulated independently or under CRN. Therefore, we have

$$\begin{aligned}\Pr\{\mathrm{CS}\} = \Pr\{\cap_{\theta=1}^{\theta^*}\mathrm{CS}^{(\theta)}\} \geq \Pr\{\cap_{\theta=1}^{d}\mathrm{CS}^{(\theta)}\} = \prod_{\theta=1}^{d}\Pr\left(\mathrm{CS}^{(\theta)}\right) \\ \geq (1-\alpha')^d = (1-(1-(1-\alpha)^{1/d}))^d = 1-\alpha. \end{aligned} \qquad \square$$

As Restart$^{\mathcal{HAK}}$ reuses the observations from Phase I and assigns the error in Phase II more efficiently, it is expected to perform better than Restart$^{\mathcal{HAK}^{\mathcal{C}}}$. Although we cannot prove the statistical validity of Restart$^{\mathcal{HAK}}$, we have not found any experiments that violate the statistical guarantee. We believe that Restart$^{\mathcal{HAK}^{\mathcal{R}}}$ and Restart$^{\mathcal{HAK}}$ are appropriate choices of sequentially-running approaches for comparison with $\mathcal{ZAK}^{\mathcal{R}}$ and $\mathcal{ZAK}$, respectively.

B.2 Procedures Restart$^{\mathcal{AK}+}$ and Restart$^{\mathcal{HAK}+}$

In this section, we discuss the algorithms of Restart$^{\mathcal{AK}+}$ and Restart$^{\mathcal{HAK}+}$ and their statistical validity.

Similar as discussed in Appendix 17, as Restart$^{\mathcal{AK}+}$ is a special case of Restart$^{\mathcal{HAK}+}$ when the number of constraints is one, we omit the discussion on Restart$^{\mathcal{AK}+}$.

Restart$^{\mathcal{HAK}+}$ performs procedure $\mathcal{HAK}+$ due to [5] independently for the threshold vectors $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(\theta^*)}$. A detailed algorithm description is shown in Algorithm 18.

Algorithm 18 Procedure Restart$^{\mathcal{HAK}+}$

[**Setup**:] Select the overall nominal confidence level $1-\alpha$. Choose tolerance levels $\epsilon_1, \ldots, \epsilon_s$, indifference-zone parameter δ, and threshold vectors $\{\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(d)}\}$. Set $\alpha' = 1-(1-\alpha)^{1/d}$.
for $\theta = 1, \ldots, d$ **do**
    [**Setup for** $\mathcal{HAK}$+:] Same as in $\mathcal{HAK}$+ except that α is replaced by α'.
    [**Initialization**], [**Feasibility Check**], [**Comparison**], and [**Stopping Rule**] are the same as in $\mathcal{HAK}$+.
    [**Stopping Condition**:] If one system is found in [**Stopping Rule**], terminate the algorithm and select the system as the best. If no system is found in [**Stopping Rule**] and $\theta = d$, declare no feasible system exists with respect to the given threshold vectors.
end for

We then prove the statistical validity of Restart$^{\mathcal{HAK}+}$ in the following theorem.

Theorem B2. *Under Assumptions 2 and 3, the procedure* $\text{Restart}^{\mathcal{HAK}+}$ *guarantees*

$$\Pr\{\text{CS}\} \geq 1 - \alpha.$$

Proof. As $\text{Restart}^{\mathcal{HAK}+}$ implements $\mathcal{HAK}+$ to $\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(\theta^*)}$ independently, we use the same notation of $\text{CS}^{(\theta)}$ as defined in Appendix B.1. By Theorems 4.4 and 4.8 of [5], we know that $\Pr\left(\text{CS}^{\theta}\right) \geq 1 - \alpha'$. Thus we have

$$\Pr\{\text{CS}\} = \Pr\left\{\cap_{\theta=1}^{\theta^*} \text{CS}^{(\theta)}\right\} \geq \Pr\left\{\cap_{\theta=1}^{d} \text{CS}^{(\theta)}\right\} = \prod_{\theta=1}^{d} \Pr\left(\text{CS}^{(\theta)}\right)$$
$$\geq (1 - \alpha')^d = (1 - (1 - (1 - \alpha)^{1/d}))^d = 1 - \alpha. \qquad \square$$

APPENDIX C

In this section, we describe procedures Restart$^{\mathcal{HAK}}$ and Restart$^{\mathcal{HAK}+}$ (discussed in Chapter 4) and address their statistical validity in Appendices C.1 and C.2, respectively. Note that the procedures included in Appendix C are different from those in Appendix B.

C.1 Algorithm Statements and Discussion of Statistical Validity for the Restart$^{\mathcal{HAK}}$ Procedure

In this section, we provide the algorithm statement of the Restart$^{\mathcal{HAK}}$ procedure and discuss its statistical validity.

To implement Restart$^{\mathcal{HAK}}$, we perform the $\mathcal{HAK}$ procedure due to [5] independently to threshold vectors $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(d)}$ and followed with a R&S problem solved by procedure $\mathcal{KN}$ due to [10]. A detailed description of algorithm is shown in Algorithm 19.

Algorithm 19 Procedure Restart$^{\mathcal{HAK}}$

[**Setup**:] Select the overall nomial confidence level $1-\alpha$. Choose tolerance level ϵ_ℓ, indifference-zon parameter δ, and threshold vectors $\{\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(d)}\}$. Set $\alpha' = 1 - (1-\alpha)^{1/(d+1)}$.
for $\theta = 1, \ldots, d$ **do**
  [**Setup for** $\mathcal{HAK}$:] Same as in $\mathcal{HAK}$ except that α is replaced by α'.
  [**Initialization**],[**Feasibility Check**],[**Feasibility Stoppoing Rule**], [**Setup for Comparison**], [**Comparison**], and [**Comparison Stopping Rule**] are same as in $\mathcal{HAK}$.
end for
Perform $\mathcal{KN}$ with nominal error α'.
[**Stopping Condition**:] Select the best feasible system found in [**Comparison Stopping Rule**] as the best system with respect to the given threshold vectors and select the best feasible system found by $\mathcal{KN}$ as the best system with respect to $\boldsymbol{q}^{(d+1)}$.

One may notice that Restart$^{\mathcal{HAK}}$ is a special case of the procedure discussed in Appendix B.1 when $\theta^* = d$. The proof of the statistical validity of Restart$^{\mathcal{HAK}}$ can be found in Appendix B.1.

C.2 Algorithm Statements and Proofs of Statistical Validity for the Restart$^{\mathcal{HAK}+}$ Procedure

In this section, we provide the algorithm statement of Restart$^{\mathcal{HAK}+}$ and discuss its statistical validity.

Similar to Restart$^{\mathcal{HAK}}$, we perform the $\mathcal{HAK}+$ procedure due to [5] independently to the threshold vectors $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(d)}$ and followed by applying procedure $\mathcal{KN}$ with respect to a R&S problem. A detailed algorithm description is shown in Algorithm 20.

Algorithm 20 Procedure Restart$^{\mathcal{HAK}+}$

[**Setup**:] Select the overall nomial confidence level $1-\alpha$. Choose tolerance level ϵ_ℓ, indifference-zon parameter δ, threshold vectors $\{\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(d)}\}$. Set $\alpha' = 1-(1-\alpha)^{1/(d+1)}$.

for $\theta = 1, \ldots, d$ **do**

  [**Setup for** $\mathcal{HAK}+$:] Same as in $\mathcal{HAK}+$ except that α is replaced by α'.

  [**Initialization**],[**Feasibility Check**], [**Comparison**], and [**Stopping Rule**] are same as in $\mathcal{HAK}+$.

end for

Perform $\mathcal{KN}$ with nominal error α'.

[**Stopping Condition**:] Select the best feasible system found in [**Comparison Stopping Rule**] as the best system with respect to the given threshold vectors and select the best feasible system found by $\mathcal{KN}$ as the best system with respect to $\boldsymbol{q}^{(d+1)}$.

Similar as discussed in Appendix C.2, Restart$^{\mathcal{HAK}+}$ is a special case of the procedure discussed in Appendix C.1 when $\theta^* = d$. The proof of the statistical validity of Restart$^{\mathcal{HAK}+}$ can be found in Appendix C.1.

REFERENCES

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

[2] Batur, D. and Kim, S.-H., "Finding feasible systems in the presence of constraints on multiple performance measures," *ACM Trans. Model. Comput. Simul.*, vol. 20, Oct. 2010.

[3] Feng, M. and Staum, J., "Green simulation: Reusing the output of repeated experiments," *ACM Trans. Model. Comput. Simul.*, vol. 27, Oct. 2017.

[4] Gao, S. and Chen, W., "A partition-based random search for stochastic constrained optimization via simulation," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 740–752, 2017.

[5] Healey, C., Andradóttir, S., and Kim, S.-H., "Selection procedures for simulations with multiple constraints under independent and correlated sampling," *ACM Trans. Model. Comput. Simul.*, vol. 24, May 2014.

[6] Healey, C. M., 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.

[7] Hong, L. J., Nelson, B. L., and Xu, J., "Discrete optimization via simulation," in *Handbook of Simulation Optimization. International Series in Operations Research & Management Science* (Fu, M., ed.), vol. 216, pp. 9–44, Springer, 2015.

[8] 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.

[9] Kim, S.-H. and Nelson, B., "Selecting the best system: Theory and methods," in *The Handbook of OR & MS: Simulation* (Henderson, S. G. and Nelson, B. L., eds.), p. 501–534, Elsevier, 2006.

[10] Kim, S.-H. and Nelson, B. L., "A fully sequential procedure for indifference-zone selection in simulation," *ACM Trans. Model. Comput. Simul.*, vol. 11, p. 251–273, July 2001.

[11] Koenig, L. W. and Law, A. M., "A procedure for selecting a subset of size m containing the l best of k independent normal populations, with applications to simulation," *Communications in Statistics - Simulation and Computation*, vol. 14, no. 3, pp. 719–734, 1985.

[12] Law, A. and Kelton, D., *Simulation Modeling and Analysis*. Academic Press, 2000.

[13] Lee, L. H., Pujowidianto, N. A., Li, L.-W., Chen, C.-H., and Yap, C. M., "Approximate simulation budget allocation for selecting the best design in the presence of stochastic constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 11, pp. 2940–2945, 2012.

[14] Lee, M. L., Park, C., and Park, D. U., "Self-adjusting the tolerance level in a fully sequential feasibility check procedure," *European Journal of Operational Research*, vol. 271, no. 2, pp. 733–745, 2018.

[15] Pasupathy, R., Hunter, S. R., Pujowidianto, N. A., Lee, L. H., and Chen, C.-H., "Stochastically constrained ranking and selection via score," *ACM Trans. Model. Comput. Simul.*, vol. 25, Aug. 2014.

[16] Pichitlamken, J., Nelson, B. L., and Hong, L. J., "A sequential procedure for neighborhood selection-of-the-best in optimization via simulation," *European Journal of Operational Research*, vol. 173, no. 1, pp. 283–298, 2006.

[17] Tamhane, A. C., "Multiple comparisons in model I one-way ANOVA with unequal variances," *Communications in Statistics - Theory and Methods*, vol. 6, no. 1, pp. 15–32, 1977.

[18] Xie, J. and Frazier, P. I., "Sequential Bayes-optimal policies for multiple comparisons with a known standard," *Operations Research*, vol. 61, no. 5, pp. 1174–1189, 2013.