Pruning Inferior Systems Using Subjective Constraints with Sequentially Added Thresholds

Yuwei Zhou[a], Sigrún Andradóttir[b], Seong-Hee Kim[b], and Chuljin Park[c]

[a] Booth School of Business, University of Chicago, Chicago, IL 60637, USA; [b]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA; [c] Department of Industrial Engineering, Hanyang University, Seoul, 04763, South Korea.

ARTICLE HISTORY
Compiled October 30, 2023

ABSTRACT
We consider the problem of pruning inferior systems among a finite number of simulated systems using constraints that are stochastic (in that their performance measures need to be estimated through observations) and subjective (in that their thresholds can be tightened or relaxed). With subjective constraints, the decision maker can test multiple threshold values to determine how a set of feasible systems changes as constraints become more strict and use this information to prune systems or identify the system with the best performance. When the number of possible thresholds is large, the decision maker may want to start by obtaining the feasibility decisions with respect to a smaller subset of thresholds. Depending on the results, she can then add tighter or relaxed thresholds if many or no feasible systems have been identified. In this paper, we present a Multi-Pass Pruning ($\mathcal{MPP}$) procedure that starts with a smaller set of thresholds in a first pass and adds more thresholds sequentially in later passes with the goal of pruning inferior systems efficiently. We prove the statistical validity of the proposed procedure and numerically demonstrate its efficiency in terms of the required number of observations for pruning inferior systems.

KEYWORDS
Ranking and Selection, Feasibility Check, Stochastic Constraints, Fully Sequential Procedure, Green Simulation

1. Introduction

We consider the problem of pruning inferior systems among a finite number of simulated systems by comparing their performance with different standards (i.e., checking the feasibility of the systems with respect to different thresholds). For example, a decision maker wishes to implement an (s, S) inventory policy (namely ordering products up to S when the inventory level at a review period is below s, with no order placed when the inventory level is above or equal to s) with two performance measures. She hopes to identify combinations of the values of s and S among finitely many choices such that (i) the probability that a shortage occurs during a review period and (ii) the expected cost per review period are both small. One can formulate the above bi-objective optimization problem as a feasibility determination problem by incorporating

CONTACT Yuwei Zhou. Email: yuwei.zhou@chicagobooth.edu

the constraints that the shortage probability is no larger than q_1 and the expected cost is less than or equal to q_2. By identifying feasible systems with respect to the thresholds q_1 and q_2, those systems that are deemed infeasible are considered as inferior and can be pruned. When constraints are subjective, instead of choosing two fixed values for the threshold constants q_1 and q_2, the decision maker can allow the constraints to have several values for each threshold (such as 20 possible values $0.01, 0.02, 0.03, \ldots, 0.19, 0.2$ for q_1 and 121 possible values $115, 115.5, 116, \ldots, 174.5, 175$ for q_2) and observe how the feasible set varies with respect to each combination of thresholds in order to further prune inferior systems. For example, she can start with larger thresholds for both constraints, say 0.2 for threshold q_1 and 175 for threshold q_2, and repeat the feasibility check for smaller values for q_1 or q_2, such as 0.05 for q_1 and 120 for q_2, while there are multiple feasible solutions left. For each combination of thresholds, whenever infeasible systems are identified, the systems are considered inferior and can be pruned.

Ranking and Selection (R&S) is a classical and actively studied problem in the simulation community. R&S procedures are useful in identifying feasible systems or a system with the best expected performance among a finite number of systems whose performance is estimated through stochastic simulation. We refer to Kim and Nelson (2006) and Hong, Nelson, and Xu (2015) for detailed literature reviews of R&S with a single performance measure. In reality, decision makers may want to handle multiple performance measures (as in the inventory example discussed above) and Hunter et al. (2019) provide a review on the multi-objective simulation optimization problem. Several studies apply the Optimal Computing Budget Allocation (OCBA) approach to handle multiple performance measures. Lee et al. (2012) consider the primary performance measure as the objective and the remaining secondary performance measures as constraints. Lee et al. (2010) propose a multi-objective OCBA method to allocate computing budget among systems in order to minimize types I and II errors in selecting non-dominated systems.

Among the R&S procedures that use the indifference-zone (IZ) approach and deal with multiple performance measures, Andradóttir and Kim (2010) consider two performance measures and propose procedures to identify the best system in terms of the primary performance measure and subject to a constraint on the secondary performance measure. Healey, Andradóttir, and Kim (2014) propose procedures to identify the best system in the presence of multiple secondary performance measures. Batur and Kim (2010) propose procedures to identify the feasibility of systems with respect to multiple constraints with fixed thresholds. Andradóttir and Lee (2021) present a procedure to estimate a Pareto set with statistical guarantee. When the stochastic constraints are subjective (i.e., have multiple thresholds), Zhou et al. (2022) adopt the concept of "green simulation" and propose statistically valid procedures that recycle simulation observations to perform feasibility checks with respect to all threshold values on each performance measure. Zhou, Andradóttir, and Kim (2023) propose procedures to identify the system with the best possible primary performance measure in the presence of subjective stochastic constraints on secondary performance measures.

In practice, the decision maker may not value one performance over the others. Or the decision maker may be interested in understanding how the set of feasible systems changes when the thresholds vary for each constraint. These problems cannot be fully addressed by the procedures due to Zhou, Andradóttir, and Kim (2023). In this paper, rather than formulating the multi-objective optimization problem as a selection-of-the-best problem with subjective constraints as Zhou, Andradóttir, and Kim (2023) do, we will model it as the feasibility check problem with subjective constraints considered by Zhou et al. (2022).

When the objective of a feasibility check problem with subjective constraints is to solve a multi-objective optimization problem, it is reasonable to consider large numbers of possible values for the constraint thresholds in order to facilitate the comparison of the different systems. However, to perform feasibility checks for subjective constraints, Zhou et al. (2022) suggest to apply their proposed $\mathcal{RF}$ procedure with respect to all possible thresholds, even if that number is large. For the inventory example discussed above, to find the system with the smallest possible combination of shortage probability and expected cost, this would involve checking feasibility with respect to all 20 thresholds of the shortage probability constraint and all 121 thresholds of the expected cost constraint, and prune systems based on the feasibility decisions to those thresholds. If the decision maker's objective is to use feasibility decisions to prune inferior systems, then checking feasibility with respect to all possible thresholds for all systems can be inefficient. For example, if any system is already deemed feasible with respect to the threshold combination 0.1 and 120, then there is no need to perform additional feasibility checks from systems deemed infeasible with respect to 0.1 and 120. However, the $\mathcal{RF}$ procedure will keep collecting more observations from such systems, solely aiming to perform feasibility checks for all less preferred thresholds (i.e., $q_1 > 0.1$ and $q_2 > 120$), which is a waste of time and resources. In this case, a multi-pass approach is preferable, where a "pass" represents the feasibility checks for systems with respect to a subset of the set of thresholds. More specifically, the decision maker can start with thresholds $q_1 \in \{0.01, 0.1, 0.2\}$ and $q_2 \in \{115, 145, 175\}$ in the first pass (a total of three thresholds for each constraint). If no systems are feasible with respect to the most preferred threshold combination 0.01 and 115 but several systems are feasible with respect to threshold combination 0.1 and 145, the decision maker can consider additional thresholds $q_1 = 0.05$ and $q_2 \in \{120, 125, \ldots, 140\}$ in the second pass (one threshold for the first constraint and five thresholds for the second constraint). If feasible systems are identified with respect to threshold combination 0.05 and 120, she can further include additional thresholds that are multiples of 0.01 from 0.02 to 0.04 for the shortage probability constraint and thresholds that are multiples of 0.5 from 115.5 to 119.5 for the expected cost constraint (a total of three thresholds for the first constraint and nine thresholds for the second constraint). Such a multi-pass approach reduces the total number of thresholds considered from 20 to 7 and 121 to 17 for the shortage probability and the expected cost constraints, respectively, and may reduce the number of needed observations as well.

In this paper, we propose a Multi-Pass Pruning ($\mathcal{MPP}$) procedure that performs feasibility check with respect to a subset of possible thresholds during the first pass and allows the decision maker to sequentially add thresholds in the following passes. The main contributions of this paper include (1) suggesting new statistics that enable us to implement feasibility checks with sequentially added thresholds for subjective constraints without significantly increasing the data storage requirement, (2) proposing a computationally efficient procedure for the purpose of pruning inferior systems, (3) proving the statistical guarantee of the proposed procedure, and (4) demonstrating the computational efficiency and statistical validity of our new procedure through experiments.

The rest of this paper is organized as follows: Section 2 provides the background of our problem, including the problem formulation and a discussion of existing work. Section 3 proposes our multi-pass procedure to identify feasible systems in the presence of subjective stochastic constraints with sequentially added thresholds. Section 4 proves the statistical validity of our proposed procedure. The experimental results are provided in Section 5 and the concluding remarks are in Section 6.

2. Background

In this section, we provide the background for our problem. Section 2.1 describes the problem and notation and Section 2.2 discusses the existing procedure $\mathcal{RF}$ that is relevant to our problem.

2.1. *Problem and Notation*

In this section, we present our problem description and the required 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. Note that the observations across different systems may or may not be correlated depending on whether systems are simulated independently or under common random numbers (CRN). The expected value and variance for system i regarding performance measure ℓ are denoted as $y_{i\ell} = \mathrm{E}[Y_{i\ell n}]$ and $\sigma^2_{i\ell} = \mathrm{Var}(Y_{i\ell n})$. 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 $s \times s$ positive definite 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 (Law and Kelton, 2000). The observations of different performance measures from a system can be correlated, such as the shortage probability and expected cost in the inventory example of Section 1.

When each constraint contains one fixed threshold value with a given threshold vector $\boldsymbol{q} = (q_1, \ldots, q_s)$, Batur and Kim (2010) introduce procedure $\mathcal{F}_B$ to determine a set of systems i with $y_{i\ell} \leq q_\ell$ for all $\ell = 1, 2, \ldots, s$. In this paper, we consider subjective constraints whose threshold values vary. We let d_ℓ denote the number of threshold values that the decision maker is interested in for performance measure ℓ and let $q_{\ell m}$ denote the threshold value for performance measure ℓ with index m, where $m = 1, \ldots, d_\ell$.

Consider the feasibility check of a system with respect to constraint ℓ and threshold $q_{\ell m}$. Andradóttir and Kim (2010) propose the concept of a tolerance level, which is denoted by ϵ_ℓ for constraint ℓ and set to a positive real number by the decision maker. Any system i with $y_{i\ell} \leq q_{\ell m} - \epsilon_\ell$ is considered desirable and feasible with respect to constraint ℓ and threshold $q_{\ell m}$. The set of all desirable systems with respect to constraint ℓ and threshold $q_{\ell m}$ 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 $q_{\ell m}$, placing

them in the set $U_\ell(q_{\ell m})$. Systems that fall within the tolerance level of $q_{\ell m}$, so that $q_{\ell m} - \epsilon_\ell < y_{i\ell} < q_{\ell m} + \epsilon_\ell$ are acceptable, and are placed in the set $A_\ell(q_{\ell m})$:

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

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 $q_{\ell m}$, 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 $q_{\ell m}$ 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:

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

where $1 - \alpha$ is the nominal confidence level for the feasibility check.

The decision maker starts by choosing all possible threshold values for constraint ℓ and defining a set of possible thresholds, $\{q_{\ell 1}, q_{\ell 2}, \ldots, q_{\ell d_\ell}\}$. Without loss of generality, we adopt the conversion that $q_{\ell 1} < q_{\ell 2} < \cdots < q_{\ell d_\ell}$ for each $\ell = 1, \ldots, s$. Suppose that the decision maker performs the initial feasibility check with respect to a subset of the possible thresholds and adds additional possible thresholds in subsequent passes (possibly adaptively). We introduce the following notation.

$T_\ell \equiv$ the index set of all possible thresholds considered for constraint $\ell, \{1, 2, \ldots, d_\ell\}$;
$T_\ell^{(w)} \equiv$ the index set of the thresholds tested in pass $w \ge 1$ for constraint ℓ;
$H^{(w)} \equiv$ the index set of the constraints with added thresholds in pass $w \ge 1$ (i.e., ℓ such that $T_\ell^{(w)} \ne \emptyset$).

The thresholds tested in each pass $w \ge 1$ should be possible thresholds that have not been tested in previous passes, and hence $T_\ell^{(w)} \subseteq T_\ell \setminus (\cup_{u=1}^{w-1} T_\ell^{(u)})$ for $\ell = 1, 2, \ldots, s$.

Consider the inventory example in Section 1. The pre-defined threshold set for the shortage probability constraint (i.e., $\ell = 1$) is $\{0.01 + 0.01\gamma \mid 0 \le \gamma \le 19, \gamma \in \mathbb{Z}\}$ and for the expected cost constraint (i.e., $\ell = 2$) is $\{115 + 0.5\gamma \mid 0 \le \gamma \le 120, \gamma \in \mathbb{Z}\}$. This means that $T_1 = \{1, 2, \ldots, 20\}$ and $T_2 = \{1, 2, \ldots, 121\}$. The decision maker wants to run the first pass with thresholds $\{0.01, 0.1, 0.2\}$ for the shortage probability constraint and with $\{115, 145, 175\}$ for the expected cost constraint. Then we have $T_1^{(1)} = \{1, 10, 20\}$ and $T_2^{(1)} = \{1, 61, 121\}$. Assume that she adds thresholds $\{120, 125, 130, 135, 140\}$ for the expected cost constraint in the second pass and no additional thresholds for the shortage probability constraint (unlike in Section 1). Then we have $H^{(2)} = \{2\}, T_1^{(2)} = \emptyset$, and $T_2^{(2)} = \{11, 21, 31, 41, 51\}$.

The thresholds tested up to the wth pass are $q_{\ell m}$ where $m \in \cup_{u=1}^{w} T_\ell^{(u)}$ for $\ell = 1, \ldots, s$. We seek a statistical guarantee that the feasibility decisions under the multi-

pass approach are identical to those of $\mathcal{RF}$ when the thresholds $q_{\ell m}$ where $m \in \cup_{u=1}^{w} T_\ell^{(u)}, \ell = 1, \ldots, s$, are considered in one pass in $\mathcal{RF}$.

Throughout the paper, we need the additional notation defined below:

$$
\begin{aligned}
n_0 &\equiv \text{the initial sample size for each system } (n_0 \geq 2);\\
r_i &\equiv \text{the number of observations obtained so far for system } i \ (r_i \geq n_0);\\
\bar{Y}_{i\ell}(r_i) &\equiv \text{average value of } Y_{i\ell 1}, \ldots, Y_{i\ell r_i} \text{ for system } i \text{ and constraint } \ell;\\
S^2_{i\ell}(n_0) &\equiv \text{the sample variance of } Y_{i\ell 1}, \ldots, Y_{i\ell n_0} \text{ for system } i = 1, 2, \ldots, k \text{ and}\\
&\quad \text{constraint } \ell = 1, \ldots, s;\\
R(r_i; v, w, z) &\equiv \max\left\{0, \frac{(n_0 - 1)wz}{v} - \frac{v}{2c} r_i\right\} \text{ for } v, w, z \in \mathbb{R}^+ \text{ and } c \in \mathbb{N}^+;\\
g(\eta) &= \sum_{j=1}^{c} (-1)^{j+1} \left(1 - \frac{1}{2}\mathcal{I}(j = c)\right) \times \left(1 + \frac{2\eta(2c - j)j}{c}\right)^{-(n_0 - 1)/2},\\
&\quad \text{where } c \in \mathbb{N}^+ \text{ and } \mathcal{I}(\cdot) \text{ is the indicator function.}
\end{aligned}
$$

The non-negative function $R(r_i; \cdot)$ is used to specify an interval $(-R(r_i; \cdot), R(r_i; \cdot))$ called the continuation region after r_i observations have been collected from system i. To determine the continuation region, we need to choose the value of c. The shape of the continuation region $(-R(r_i; \cdot), R(r_i; \cdot))$ becomes a longer and narrower triangle as c increases and turns into two straight lines when $c = \infty$. The choice $c = 1$ guarantees a unique and easy solution when computing the implementation parameter η from $g(\eta)$. Kim and Nelson (2001) also suggests that $c = 1$ is a good choice when the decision maker does not have information about the systems' mean configuration. Zhou et al. (2022) consider both $c \in \mathbb{N}^+$ and $c = \infty$ and present expressions for $g(\eta)$ for in both cases. However, as the focus of this paper is not on the continuation region and as $c = 1$ is a good choice, we only consider $g(\eta)$ when $c \in \mathbb{N}^+$ and we set $c = 1$ for the experimental results in this paper.

2.2. *Existing Procedure*

In this section, we provide a brief overview of the procedure $\mathcal{RF}$ due to Zhou et al. (2022) that performs feasibility check for subjective constraints. $\mathcal{RF}$ is given in Algorithm 1 with the following definition of β,

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

As shown in Algorithm 1, Procedure $\mathcal{RF}$ determines the feasibility of system i with respect to threshold $q_{\ell m}$ on constraint ℓ as

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

In other words, $\mathcal{RF}$ constructs an interval $(\bar{Y}_{i\ell}(r_i) - R(r_i; \epsilon_\ell, \eta, S^2_{i\ell}(n_0))/r_i, \bar{Y}_{i\ell}(r_i) + R(r_i; \epsilon_\ell, \eta, S^2_{i\ell}(n_0))/r_i)$ whenever an observation $Y_{i\ell r_i}$ is collected and makes the feasi-

Algorithm 1 Procedure $\mathcal{RF}$

[**Setup**:]
Choose confidence level $1-\alpha$, tolerance level ϵ_ℓ, and thresholds $\{q_{\ell 1}, q_{\ell 2}, \ldots, q_{\ell d_\ell}\}$ for constraint $\ell = 1, 2, \ldots, s$. Also, choose the value of $c \in \mathbb{N}^+$ 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 (1), 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 $S^2_{i\ell}(n_0)$ 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 $\bar{Y}_{i\ell}(r_i) + R(r_i; \epsilon_\ell, \eta_\ell, S^2_{i\ell}(n_0))/r_i \le q_{\ell m}$, set $Z_{i\ell m} = 1$ and $\mathrm{ON}_\ell = \mathrm{ON}_\ell \setminus \{m\}$.
If $\bar{Y}_{i\ell}(r_i) - R(r_i; \epsilon_\ell, \eta_\ell, S^2_{i\ell}(n_0))/r_i \ge q_{\ell m}$, set $Z_{i\ell m} = 0$ and $\mathrm{ON}_\ell = \mathrm{ON}_\ell \setminus \{m\}$.
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 $Y_{i\ell r_i}$ and update $\bar{Y}_{i\ell}(r_i)$ for $\ell \in \mathrm{ON}$, then go to [**Feasibility Check**].
end for

bility decision for system i with respect to threshold $q_{\ell m}$ when the threshold $q_{\ell m}$ falls outside of the interval.

The following theorem from Zhou et al. (2022) shows that there are at most two *effective* thresholds on each constraint.

Theorem 2.1. *For system i with s constraints and thresholds $T_\ell = \{q_{\ell 1}, q_{\ell 2}, \ldots, q_{\ell d_\ell}\}$ for $\ell = 1, 2, \ldots, s$, the joint probability of correct decision with respect to thresholds $y_{i\ell} - \epsilon_\ell$ and $y_{i\ell} + \epsilon_\ell$ is a lower bound on the joint probability of correct decision with respect to all thresholds of constraint ℓ, i.e.,*

$$\Pr\left(\cap_{m=1}^{d_\ell} \mathrm{CD}_{i\ell}(q_{\ell m})\right) \ge \Pr\left(\mathrm{CD}_{i\ell}(y_{i\ell} - \epsilon_\ell), \mathrm{CD}_{i\ell}(y_{i\ell} + \epsilon_\ell)\right).$$

Theorem 2.1 implies that a procedure designed to deliver a correct decision with respect to the two *effective* thresholds $y_{i\ell} - \epsilon_\ell, y_{i\ell} + \epsilon_\ell$ on constraint ℓ will deliver correct decisions with respect to the other thresholds. $\mathcal{RF}$ is designed to accomplish this. Thus $\mathcal{RF}$ avoids setting the implementation parameter η in a very conservative way and scales well with respect to the number of thresholds. However, it requires all thresholds to be tested simultaneously in one pass for statistical validity, and thus the computation time for [**Feasibility Check**] increases as the number of thresholds increases.

3. Multi-Pass Pruning ($\mathcal{MPP}$) Procedure

In this section, we present a new procedure to identify feasible systems for subjective constraints when thresholds are added sequentially in multiple passes. Section 3.1 presents the pruning procedure for the first pass and Section 3.2 proposes the pruning procedure for the subsequent passes.

3.1. *First-Pass Pruning Procedure*

In this section, we propose a procedure with new statistics that performs feasibility check for the first-pass when the decision maker considers a subset of thresholds chosen from the set of all possible thresholds.

As discussed in Section 2.2, Procedure $\mathcal{RF}$ checks the two inequalities in Equation (2) for each threshold $q_{\ell m}$ whenever an observation is collected from constraint ℓ. Recall that the statistical guarantee we want to provide is that the feasibility decisions under the multi-pass approach are identical to those of $\mathcal{RF}$. One straightforward way to provide such guarantee is to save the sample paths $\bar{Y}_{i\ell}(r_i)$ of system i, where $r_i \geq n_0$, during the first pass and track down from the very first stage (i.e., from $r_i = n_0$) what would have happened if $\mathcal{RF}$ had been performed with all thresholds considered in the multiple passes. However, this is not desirable due to a data storage problem. Instead, we keep the following two statistics while system i is simulated:

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

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

where $v_{i\ell}^{\mathrm{UB}}$ is the minimum of the upper bounds (UB) and $v_{i\ell}^{\mathrm{LB}}$ is the maximum of the lower bounds (LB) that the possible thresholds $q_{\ell m}$ have been compared with so far. The first pass for the Multi-Pass Pruning ($\mathcal{MPP}$) procedure, namely $\mathcal{P}^{(1)}$, then determines system i feasible with respect to threshold $q_{\ell m}$ on constraint ℓ if $v_{i\ell}^{\mathrm{UB}} \leq q_{\ell m}$ and infeasible if $v_{i\ell}^{\mathrm{LB}} \geq q_{\ell m}$ where $m \in T_\ell^{(1)}$. We update the interval $(v_{i\ell}^{\mathrm{LB}}, v_{i\ell}^{\mathrm{UB}})$ whenever an observation $Y_{i\ell r_i}$ is collected and the feasibility decision is made for a particular threshold $q_{\ell m}$, where $m \in T_\ell^{(1)}$, once it falls outside of the interval for the first time. Figure 1 shows the behavior of $v_{i\ell}^{\mathrm{UB}}$ and $v_{i\ell}^{\mathrm{LB}}$ when feasibility check is performed with respect to thresholds $\{q_{\ell m_1}, q_{\ell m_2}\}$ on constraint ℓ during the first pass (i.e., $m_1, m_2 \in T_\ell^{(1)}$). In Figure 1, $q_{\ell m_1}$ falls outside of the interval (i.e., below $v_{i\ell}^{\mathrm{LB}}$) at r_{i1} for the first time and system i is then determined infeasible with respect to $q_{\ell m_1}$. The feasibility decision for $q_{\ell m_2}$ is determined at r_{i2} and system i is determined feasible with respect to $q_{\ell m_2}$ as $q_{\ell m_2}$ falls above $v_{i\ell}^{\mathrm{UB}}$.

The full description of $\mathcal{P}^{(1)}$ is provided in Algorithm 2 for k systems, s constraints, and thresholds $q_{\ell m}$ where $m \in T_\ell^{(1)}$ for $\ell = 1, \ldots, s$. We use $Z_{i\ell m}$, where $m \in T_\ell^{(1)}$, to indicate the feasibility of system i with respect to threshold $q_{\ell m}$ on constraint ℓ (i.e., $Z_{i\ell m} = 1$ if system i is feasible with respect to threshold $q_{\ell m}$ on constraint ℓ, and $Z_{i\ell m} = 0$ otherwise).

Note that in addition to only considering a subset of all possible thresholds whose indices are in $T_\ell^{(1)} \subseteq T_\ell$ and maintaining the variables $v_{i\ell}^{\mathrm{UB}}$ and $v_{i\ell}^{\mathrm{LB}}$, $\mathcal{P}^{(1)}$ incorporates

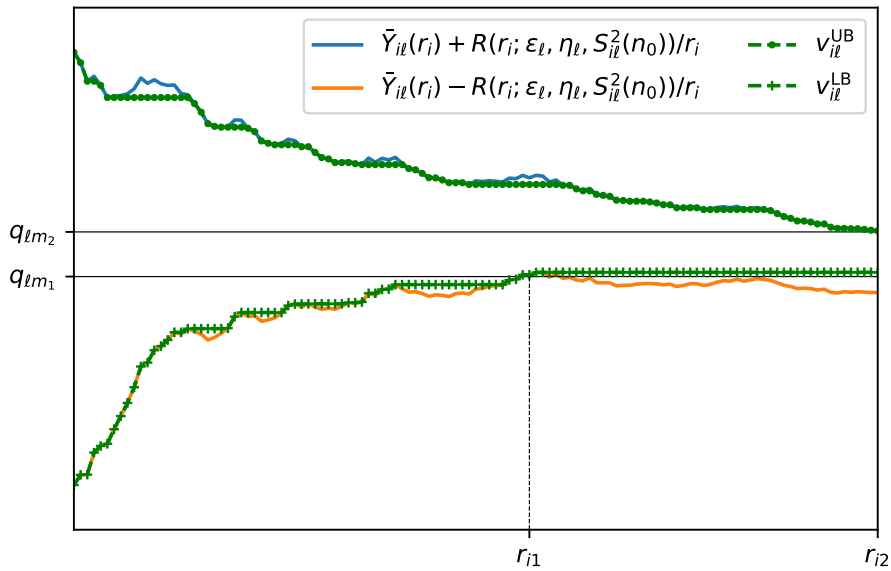

Figure 1. Behavior of $v^{\mathrm{UB}}_{i\ell}$ and $v^{\mathrm{LB}}_{i\ell}$ for the feasibility check with respect to thresholds $\{q_{\ell m_1}, q_{\ell m_2}\}$ on constraint ℓ, where $m_1, m_2 \in T^{(1)}_\ell$.

several other differences with $\mathcal{RF}$ as described below:

- We add variables $\mathrm{LAST}_{i\ell}$ in the description of $\mathcal{P}^{(1)}$. This is needed for subsequent passes of the feasibility check when $v^{\mathrm{UB}}_{i\ell} \leq v^{\mathrm{LB}}_{i\ell}$ so that we can directly conclude the correct feasibility decisions for the added thresholds. Note that $\mathcal{P}^{(1)}$ can overwrite the value of $\mathrm{LAST}_{i\ell}$ if both $v^{\mathrm{LB}}_{i\ell}$ and $v^{\mathrm{UB}}_{i\ell}$ are updated in one stage. A detailed discussion on the use of $\mathrm{LAST}_{i\ell}$, including an explanation that overwriting $\mathrm{LAST}_{i\ell}$ in the same stage does not result in unintended consequences, is provided in Section 3.2.
- We keep collecting observations $Y_{i\ell r_i}$ from constraint ℓ with $v^{\mathrm{UB}}_{i\ell} > v^{\mathrm{LB}}_{i\ell}$ and update $\bar{Y}_{i\ell r_i}(r_i)$ when $\mathrm{ON} \neq \emptyset$ even if $\ell \notin \mathrm{ON}$. Note that whenever we conduct one simulation replication, observations across all the constraints can be obtained. Therefore, obtaining observations from constraint ℓ such that $\ell \notin \mathrm{ON} \neq \emptyset$ does not increase the total number of required simulation replications. This additional data prepares for the case when the decision maker adds thresholds in later passes for such constraints (in order to guarantee statistical validity and increase efficiency). One may notice that when $v^{\mathrm{UB}}_{i\ell} \leq v^{\mathrm{LB}}_{i\ell}$, adding thresholds to constraint ℓ does not require additional observations to conclude their feasibility decisions because every possible threshold $q_{\ell m}$ satisfies $v^{\mathrm{UB}}_{i\ell} \leq q_{\ell m}$ or $v^{\mathrm{LB}}_{i\ell} \geq q_{\ell m}$ or both. In addition, when $v^{\mathrm{UB}}_{i\ell} \leq v^{\mathrm{LB}}_{i\ell}$, we utilize the final value of $\mathrm{LAST}_{i\ell}$ to conclude certain feasibility decisions in later passes. Collecting additional observations when $v^{\mathrm{UB}}_{i\ell} \leq v^{\mathrm{LB}}_{i\ell}$ might overwrite $\mathrm{LAST}_{i\ell}$ and lead to issues with statistical validity.
- We save the (final) random seed(s) for each system when the feasibility check is completed so that we can continue generating observations for the systems that match those of $\mathcal{RF}$ in future passes if needed.

Algorithm 2 Procedure $\mathcal{P}^{(w)}, w = 1$

[**Setup**:]

Choose confidence level $1-\alpha$, tolerance level ϵ_ℓ, and threshold index set $T_\ell^{(1)}$ for constraint $\ell = 1, 2, \ldots, s$. Also, choose the value of $c \in \mathbb{N}^+$ and set $\Theta = \{1, 2, \ldots, k\}$ and $H^{(1)} = \{1, 2, \ldots, s\}$. For $\ell = 1, \ldots, s$, set η_ℓ such that $g(\eta_\ell) = \beta_\ell$, where β satisfies (1), 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 $S^2_{i\ell}(n_0)$ for $\ell = 1, 2, \ldots, s$.
- Set $r_i = n_0$, $\mathrm{ON} = H^{(1)}$, and $\mathrm{ON}_\ell = T_\ell^{(1)}$ for $\ell = 1, 2, \ldots, s$.
- Set $v^{\mathrm{UB}}_{i\ell} = \infty$ and $v^{\mathrm{LB}}_{i\ell} = -\infty$ for $\ell = 1, 2, \ldots, s$.
- Set $\mathrm{LAST}_{i\ell}$ as an empty string for $\ell = 1, \ldots, s$.

[**Feasibility Check**:]

for $\ell = 1, 2 \ldots, s$ and $v^{\mathrm{UB}}_{i\ell} > v^{\mathrm{LB}}_{i\ell}$ **do**

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

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

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

If $v^{\mathrm{UB}}_{i\ell} \le q_{\ell m}$, set $Z_{i\ell m} = 1$ and $\mathrm{ON}_\ell = \mathrm{ON}_\ell \setminus \{m\}$.

If $v^{\mathrm{LB}}_{i\ell} \ge q_{\ell m}$, set $Z_{i\ell m} = 0$ and $\mathrm{ON}_\ell = \mathrm{ON}_\ell \setminus \{m\}$.

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 $\ell \in H^{(w)}$ and $m \in T_\ell^{(w)}$ and save the (final) random seed(s) for system i. Otherwise, set $r_i = r_i + 1$, and, for any ℓ such that $v^{\mathrm{UB}}_{i\ell} > v^{\mathrm{LB}}_{i\ell}$, take one additional observation $Y_{i\ell r_i}$ and update $\bar{Y}_{i\ell}(r_i)$. Then go to [**Feasibility Check**].

end for

3.2. *Pruning Procedure for Later Passes*

In this section, we propose a procedure to determine the feasibility for the added thresholds in the later passes after the first pass is complete. We let $w \ge 2$ be the index of the pass and consider a particular constraint $\ell \in H^{(w)}$. Recall that the thresholds for the wth pass need to be selected from the pre-defined threshold set $\{q_{\ell 1}, q_{\ell 2}, \ldots, q_{\ell d_\ell}\}$. That is, the threshold indices for the wth pass satisfy

$$T_\ell^{(w)} \subseteq T_\ell \setminus \left(\cup_{u=1}^{w-1} T_\ell^{(u)} \right).$$

If the decision maker decides to add a threshold $q_{\ell m}$ where $m \in T_\ell^{(w)}$, the feasibility decision that $\mathcal{RF}$ would have made for system i is retrieved during the wth pass. This is achieved by comparing the values of $v^{\mathrm{UB}}_{i\ell}$ and $v^{\mathrm{LB}}_{i\ell}$ with $q_{\ell m}$, and only collecting additional observations if needed (in which case the additional observations would be collected and employed as in $\mathcal{P}^{(1)}$). We now discuss how we determine feasibility for

$q_{\ell m}$ depending on the values of $v_{i\ell}^{\mathrm{UB}}$ and $v_{i\ell}^{\mathrm{LB}}$ at the end of pass $w-1$ based on three cases as follows.

Case 1: When $v_{i\ell}^{UB} \le q_{\ell m}$ and $v_{i\ell}^{\mathrm{LB}} < q_{\ell m}$ (or $v_{i\ell}^{LB} \ge q_{\ell m}$ and $v_{i\ell}^{\mathrm{UB}} > q_{\ell m}$):
If $v_{i\ell}^{\mathrm{UB}} \le q_{\ell m}$ and $v_{i\ell}^{\mathrm{LB}} < q_{\ell m}$, we immediately declare system i is feasible with respect to $q_{\ell m}$. Similarly, we immediately declare system i is infeasible with respect to $q_{\ell m}$ if $v_{i\ell}^{\mathrm{LB}} \ge q_{\ell m}$ and $v_{i\ell}^{\mathrm{UB}} > q_{\ell m}$.

Case 2: When $v_{i\ell}^{\mathrm{UB}} \le q_{\ell m} \le v_{i\ell}^{\mathrm{LB}}$:
Although $v_{i\ell}^{\mathrm{UB}} > v_{i\ell}^{\mathrm{LB}}$ in general, it is possible that $v_{i\ell}^{\mathrm{UB}} \le v_{i\ell}^{\mathrm{LB}}$ happens at the time when feasibility decisions for all thresholds on constraint ℓ of system i with indices in $T_\ell^{(w-1)}$ are concluded in the previous pass. If an added threshold value $q_{\ell m}$ satisfies $v_{i\ell}^{\mathrm{UB}} \le q_{\ell m} \le v_{i\ell}^{\mathrm{LB}}$, we need to know which value of $v_{i\ell}^{\mathrm{UB}}$ and $v_{i\ell}^{\mathrm{LB}}$ was updated last in the previous pass. Consider the example in Figure 2 where we use $q_{\ell m'}$ to denote the threshold whose feasibility decisions are made last in pass $w-1$. At the last stage of the feasibility check for constraint ℓ in the previous pass, we have $v_{i\ell}^{\mathrm{UB}} \le v_{i\ell}^{\mathrm{LB}}$. When the feasibility decision for $q_{\ell m}$ needs to be retrieved in the current pass, we have $v_{i\ell}^{\mathrm{UB}} \le q_{\ell m} \le v_{i\ell}^{\mathrm{LB}}$ and $v_{i\ell}^{\mathrm{UB}}$ is the last updated value in the previous pass. As shown in Figure 2, the lower bound of the interval

$$\left(\bar{Y}_{i\ell}(r_{i\ell}) - R(r_{i\ell};\epsilon_\ell,\eta_\ell,S_{i\ell}^2(n_0))/r_{i\ell}, \bar{Y}_{i\ell}(r_{i\ell}) + R(r_{i\ell};\epsilon_\ell,\eta_\ell,S_{i\ell}^2(n_0))/r_{i\ell}\right)$$

is greater than $v_{i\ell}^{\mathrm{LB}}$ before the last stage of the previous pass and $q_{\ell m}$ would have satisfied $v_{i\ell}^{\mathrm{LB}} \ge q_{\ell m}$ before it satisfied $v_{i\ell}^{\mathrm{UB}} \le q_{\ell m}$ if $q_{\ell m}$ had been included in the previous pass. Thus we should declare system i infeasible with respect to $q_{\ell m}$. In general, when $v_{i\ell}^{\mathrm{UB}} \le q_{\ell m} \le v_{i\ell}^{\mathrm{LB}}$, if the last updated value among $v_{i\ell}^{\mathrm{UB}}, v_{i\ell}^{\mathrm{LB}}$ in the previous pass is $v_{i\ell}^{\mathrm{UB}}$ (i.e., $\mathrm{LAST}_{i\ell} = \mathrm{UB}$), we declare the system infeasible with respect to $q_{\ell m}$ and we declare the system feasible with respect to $q_{\ell m}$ if the last updated value in the previous pass is $v_{i\ell}^{\mathrm{LB}}$ (i.e., $\mathrm{LAST}_{i\ell} = \mathrm{LB}$).

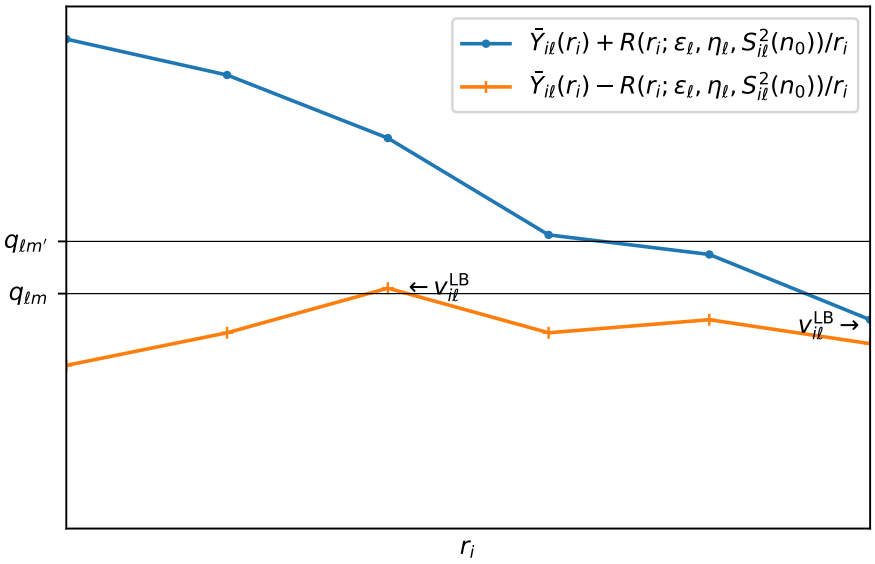

Figure 2. Crossing of $v_{i\ell}^{\mathrm{UB}}$ and $v_{i\ell}^{\mathrm{LB}}$ on constraint ℓ when retrieving feasibility decision for threshold $q_{\ell m}$, where $m \in T_\ell^{(w)}$ and $w \ge 2$.

Note that the value of $\text{LAST}_{i\ell}$ can be overwritten when both $v_{i\ell}^{\text{LB}}$ and $v_{i\ell}^{\text{UB}}$ are updated in one stage. For example, we see that the first few observations taken in Figure 1 result in $v_{i\ell}^{\text{UB}}$ decreasing and $v_{i\ell}^{\text{LB}}$ increasing. This means that both $v_{i\ell}^{\text{UB}}$ and $v_{i\ell}^{\text{LB}}$ are updated the same stages and thus $\text{LAST}_{i\ell}$ updates accordingly (and therefore is overwritten). However, overwriting $\text{LAST}_{i\ell}$ results in values $v_{i\ell}^{\text{UB}}$ of $\bar{Y}_{i\ell}(r_i) + R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/r_i$ and $v_{i\ell}^{\text{LB}}$ of $\bar{Y}_{i\ell}(r_i) - R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/r_i$. Since $R(r_{i\ell}; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))$ takes a non-negative value, it is guaranteed $v_{i\ell}^{\text{UB}} \geq v_{i\ell}^{\text{LB}}$ in such a case. As the variable $\text{LAST}_{i\ell}$ is only used when $v_{i\ell}^{\text{UB}} \leq v_{i\ell}^{\text{LB}}$, overwriting the value of $\text{LAST}_{i\ell}$ in the same stage does not have any unintended consequences for the feasibility check during the following pass when $R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0)) > 0$ (and hence $v_{i\ell}^{\text{UB}} > v_{i\ell}^{\text{LB}}$). Moreover, when $R(r_i; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0)) = 0$ in the final stage of pass $w-1$, overwriting $\text{LAST}_{i\ell}$ implies that $v_{i\ell}^{\text{LB}} = v_{i\ell}^{\text{UB}} = \bar{Y}_{i\ell}(r_i)$ and $\text{LAST}_{i\ell} = \text{UB}$; thus, the system is declared infeasible when implementing $\mathcal{P}^{(w)}$. In this case, $v_{i\ell}^{\text{UB}} \leq q_{\ell m} \leq v_{i\ell}^{\text{LB}}$ implies that $q_{\ell m} = v_{i\ell}^{\text{UB}} = v_{i\ell}^{\text{LB}} = \bar{Y}_{i\ell}(r_i)$ and it is easily seen that the value of $Z_{i\ell m}$ is also overwritten in Algorithm 1 and thus the system is concluded infeasible by $\mathcal{RF}$, which matches with the decision from $\mathcal{P}^{(w)}$. Furthermore, note that $\bar{Y}_{i\ell}(r_i) = v_{i\ell}^{\text{LB}} = v_{i\ell}^{\text{UB}} = q_{i\ell}$ occurs with zero probability under Assumption 1.

Case 3: When $v_{i\ell}^{\text{LB}} < q_{\ell m} < v_{i\ell}^{\text{UB}}$:
If $v_{i\ell}^{\text{LB}} < q_{\ell m} < v_{i\ell}^{\text{UB}}$, we cannot determine feasibility relative to threshold $q_{\ell m}$ based on the data collected in passes $1, \ldots, w-1$ and need to take additional observations. In this case, one needs to use the (final) random seed(s) saved for the system from the previous pass. This is essential for the proof of statistical validity discussed in Section 4.

We present the description of the retrieving process for pass $w \geq 2$, namely $\mathcal{P}^{(w)}$, in Algorithm 3. We determine the values of $Z_{i\ell m}$ for $\ell \in H^{(w)}$ and $m \in T_\ell^{(w)}$ in the wth pass for the feasibility of system i for the corresponding added thresholds $q_{\ell m}$.

4. Statistical Validity

We prove the statistical validity of our proposed procedure in this section. We first address the statistical validity for a single system in Section 4.1 and then discuss the overall probability of correct decision for multiple systems in Section 4.2.

4.1. *Statistical Validity of $\mathcal{MPP}$ for a Single System*

In this section, we prove the statistical validity of our proposed procedure for a single system. Recall that Procedure $\mathcal{RF}$ makes a decision for each threshold $q_{\ell m}$ for $m \in T_\ell$ when the interval

$$\left(\bar{Y}_{i\ell}(r) - \frac{R(r; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))}{r}, \bar{Y}_{i\ell}(r) + \frac{R(r; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))}{r}\right) \quad \text{for} \quad r = n_0, n_0+1, \ldots,$$

does not include the threshold $q_{\ell m}$ for the first time. We refer to such a stage as the first exit stage. The following Lemma 4.1 from Zhou et al. (2022) provides the statistical validity of Procedure $\mathcal{RF}$ (shown in Algorithm 1) for a single system.

Lemma 4.1. *For system i with s constraints and threshold constants $q_{\ell m}$ where $m \in T_\ell$ for $\ell = 1, \ldots, s$, Procedure $\mathcal{RF}$ makes a decision for each threshold based on its*

Algorithm 3 Procedure $\mathcal{P}^{(w)}, w \geq 2$

[**Setup**:]

Decide $H^{(w)}$, the set of constraints that need additional thresholds, for the wth pass. Choose the indices of the thresholds added, $T_\ell^{(w)}$, for $\ell \in H^{(w)}$ and set $\Theta = \{1, 2, \ldots, k\}$.

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

[**Initialization**:]

- Set $\text{ON} = H^{(w)}$ and $\text{ON}_\ell = T_\ell^{(w)}$ for $\ell \in H^{(w)}$.
- Obtain $r_i, \bar{Y}_{i\ell}(r_i), \text{LAST}_{i\ell}, v_{i\ell}^{\text{LB}}$, and $v_{i\ell}^{\text{UB}}$ for $\ell = 1, \ldots, s$ from Procedure $\mathcal{P}^{(w-1)}$ and $S_{i\ell}^2(n_0)$ from $\mathcal{P}^{(1)}$.
- Obtain the saved seed(s) for system i from Procedure $\mathcal{P}^{(w-1)}$ and use it (them) for generating observations from system i (if needed).

[**Initial Feasibility Check**:]

for $\ell \in H^{(w)}$ **do**

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

If $v_{i\ell}^{\text{UB}} \leq q_{\ell m}$ and $v_{i\ell}^{\text{LB}} < q_{\ell m}$, set $Z_{i\ell m} = 1$ and $\text{ON}_\ell = \text{ON}_\ell \setminus \{m\}$;

Else if $v_{i\ell}^{\text{LB}} \geq q_{\ell m}$ and $v_{i\ell}^{\text{UB}} > q_{\ell m}$, set $Z_{i\ell m} = 0$ and $\text{ON}_\ell = \text{ON}_\ell \setminus \{m\}$;

Else if $v_{i\ell}^{\text{UB}} \leq q_{\ell m} \leq v_{i\ell}^{\text{LB}}$,

if $\text{LAST}_{i\ell} = \text{UB}$, set $Z_{i\ell m} = 0$ and $\text{ON}_\ell = \text{ON}_\ell \setminus \{m\}$;

if $\text{LAST}_{i\ell} = \text{LB}$, set $Z_{i\ell m} = 1$ 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 all $\ell \in H^{(w)}$ and $m \in T_\ell^{(w)}$. Otherwise, set $r_i = r_i + 1$, and, for any ℓ such that $v_{i\ell}^{\text{UB}} > v_{i\ell}^{\text{LB}}$, take one additional observation $Y_{i\ell r_i}$ and update $\bar{Y}_{i\ell}(r_i)$. Then go to [**Feasibility Check**] of Procedure $\mathcal{P}^{(1)}$.

end for

first exit stage and guarantees $\Pr(\cap_{\ell=1}^s \text{CD}_{i\ell}) \geq 1 - \beta$.

The following lemma shows that if $\mathcal{RF}$ is implemented for thresholds $q_{\ell m}$, where $m \in \cup_{u=1}^w T_\ell^{(u)}$ (i.e., the thresholds considered through the execution of $\mathcal{P}^{(w)}$), then it still provides statistical validity.

Lemma 4.2. *Procedure $\mathcal{RF}$, executed with respect to $q_{\ell m}$, where $m \in \cup_{u=1}^w T_\ell^{(u)}$, guarantees*

$$\Pr\left(\cap_{\ell=1}^s \cap_{m \in \cup_{u=1}^w T_\ell^{(u)}} \text{CD}_{i\ell}(q_{\ell m})\right) \geq 1 - \beta.$$

Proof. Because $\cup_{u=1}^w T_\ell^{(u)} \subseteq T_\ell$ and $\cap_{m=1}^{d_\ell} \text{CD}_{i\ell}(q_{\ell m}) \subseteq \cap_{m \in \cup_{u=1}^w T_\ell^{(u)}} \text{CD}_{i\ell}(q_{\ell m})$ for $\ell = 1, 2, \ldots, s$, we have

$$\Pr\left(\cap_{\ell=1}^s \cap_{m \in \cup_{u=1}^w T_\ell^{(u)}} \text{CD}_{i\ell}(q_{\ell m})\right) \geq \Pr\left(\cap_{\ell=1}^s \cap_{m=1}^{d_\ell} \text{CD}_{i\ell}(q_{\ell m})\right) \geq 1 - \beta,$$

where the second inequality follows from Lemma 4.1. □

We use $\mathcal{MPP}^{(w)}$ to denote the $\mathcal{MPP}$ procedure with $w \geq 1$ passes (so that $\mathcal{P}^{(u)}$

is applied to thresholds $q_{\ell m}$, where $m \in T_\ell^{(u)}$, for $u = 1, \ldots, w$). Now we present the main theorem that proves that the $\mathcal{MPP}^{(w)}$ procedure guarantees statistical validity by showing that the feasibility decisions of $\mathcal{MPP}^{(w)}$ match those of $\mathcal{RF}$ with respect to thresholds $q_{\ell m}$, where $m \in \cup_{u=1}^{w} T_\ell^{(u)}$ and $\ell = 1, \ldots, s$.

Theorem 4.3. *Given system i with s constraints and index set $T_\ell = \{1, 2, \ldots, d_\ell\}$ for $\ell = 1, 2, \ldots, s$, the Multi-Pass Pruning procedure $\mathcal{MPP}^{(w)}$ guarantees*

$$\Pr\left(\cap_{\ell=1}^{s} \cap_{m \in \cup_{u=1}^{w} T_\ell^{(u)}} \mathrm{CD}_{i\ell}(q_{\ell m})\right) \geq 1 - \beta.$$

Proof. As $\mathcal{RF}$ makes the feasibility decisions at the first exit stage, we prove the theorem by showing that the feasibility decisions made by the $\mathcal{MPP}^{(w)}$ procedure with respect to threshold $q_{\ell m}$ where $m \in \cup_{u=1}^{w} T_\ell^{(u)}$ for $\ell = 1, \ldots, s$ are identical to those at the first exit stage, which in turn match the decisions made by $\mathcal{RF}$.

Procedure $\mathcal{P}^{(1)}$ sets the implementation parameters β_ℓ identical to those of $\mathcal{RF}$ for $\ell = 1, 2, \ldots, s$. In addition, the two inequalities that determine the values of $Z_{i\ell m}$ in the **[Feasibility Check]** step are essentially identical in the two procedures as well. The main difference between $\mathcal{RF}$ and $\mathcal{P}^{(1)}$ is that $\mathcal{P}^{(1)}$ keeps updating the values of $v_{i\ell}^{\mathrm{UB}}$ and $v_{i\ell}^{\mathrm{LB}}$ for every constraint ℓ such that $v_{i\ell}^{\mathrm{LB}} < v_{i\ell}^{\mathrm{UB}}$ whenever system i is simulated, even if $\ell \notin \mathrm{ON}$. However, this difference does not affect the values of $Z_{i\ell m}$ as $\mathrm{ON}_\ell = \emptyset$ for $\ell \notin \mathrm{ON}$; thus $Z_{i\ell m}$ are not updated and $\mathcal{RF}$ and $\mathcal{P}^{(1)}$ yield the same decisions for $m \in T_\ell^{(1)}$ and $\ell = 1, 2, \ldots, s$.

Let $w \geq 2$ and ℓ be an arbitrary constraint. To avoid the trivial case, we assume that the decision maker adds thresholds for constraint ℓ in pass w (i.e., $\ell \in H^{(w)}$). We need to consider five cases for each added threshold $q_{\ell m}$, where $m \in T_\ell^{(u)}$, after the completion of $\mathcal{P}^{(w-1)}$:

(1) *If $v_{i\ell}^{\mathrm{UB}} \leq q_{\ell m}$ and $v_{i\ell}^{\mathrm{LB}} < q_{\ell m}$, both $\mathcal{MPP}^{(w)}$ and $\mathcal{RF}$ declare system i feasible with respect to $q_{\ell m}$ by the first exist stage.*

Recall that r_i is the number of observations collected from system i to conclude feasibility decisions from all constraints after the completion of $\mathcal{P}^{(w-1)}$. It is clear from Algorithm 3 that $\mathcal{MPP}^{(w)}$ declares system i feasible with respect to $q_{\ell m}$. Moreover, if $v_{i\ell}^{\mathrm{UB}} \leq q_{\ell m}$ and $v_{i\ell}^{\mathrm{LB}} < q_{\ell m}$, then there exists $n_0 \leq n^{\mathrm{UB}} \leq r_i$ such that $\bar{Y}_{i\ell}(n^{\mathrm{UB}}) + R(n^{\mathrm{UB}}; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/n^{\mathrm{UB}} \leq q_{\ell m}$ but there does not exist $n_0 \leq n^{\mathrm{LB}} \leq r_i$ such that $\bar{Y}_{i\ell}(n^{\mathrm{LB}}) - R(n^{\mathrm{LB}}; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/n^{\mathrm{LB}} \geq q_{\ell m}$. Therefore, $\mathcal{RF}$ also declares system i feasible with respect to $q_{\ell m}$.

(2) *If $v_{i\ell}^{LB} \geq q_{\ell m}$ and $v_{i\ell}^{\mathrm{UB}} > q_{\ell m}$, both $\mathcal{MPP}^{(w)}$ and $\mathcal{RF}$ declare system i infeasible with respect to $q_{\ell m}$ by the first exist stage.*

By similar arguments as in Case 1, the above claim holds.

(3) *If $v_{i\ell}^{\mathrm{UB}} \leq q_{\ell m} \leq v_{i\ell}^{\mathrm{LB}}$ and $\mathrm{LAST}_{i\ell} = \mathrm{UB}$, both $\mathcal{MPP}^{(w)}$ and $\mathcal{RF}$ declare system i infeasible with respect to $q_{\ell m}$ by the first exist stage.*

As stated in Algorithm 3, it is clear that $\mathcal{MPP}^{(w)}$ declares system i infeasible with respect to threshold $q_{\ell m}$. Recall that r_i represents the number of observations from system i from all constraints after the completion of $\mathcal{P}^{(w-1)}$. Since $v_{i\ell}^{\mathrm{UB}} \leq q_{\ell m}$, there exists $n_0 \leq n^{\mathrm{UB}} \leq r_i$ such that $\bar{Y}_{i\ell}(n^{\mathrm{UB}}) + R(n^{\mathrm{UB}}; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/n^{\mathrm{UB}} \leq q_{\ell m}$. Similarly, due to $v_{i\ell}^{\mathrm{LB}} \geq q_{\ell m}$, there exists $n_0 \leq n^{\mathrm{LB}} \leq r_i$ such that $\bar{Y}_{i\ell}(n^{\mathrm{LB}}) - R(n^{\mathrm{LB}}; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/n^{\mathrm{LB}} \geq q_{\ell m}$. Moreover, given that $\mathrm{LAST}_{i\ell} = \mathrm{UB}$ and $\mathrm{LAST}_{i\ell}$ will not be updated when $v_{i\ell}^{\mathrm{UB}} \leq v_{i\ell}^{\mathrm{LB}}$ happens, we know that $n^{\mathrm{LB}} < n^{\mathrm{UB}}$ since $v_{i\ell}^{\mathrm{UB}}$ was updated later than $v_{i\ell}^{\mathrm{LB}}$. There-

fore, as $n_0 \leq n^{\mathrm{LB}} < n^{\mathrm{UB}} \leq r_i$, we see that $\mathcal{RF}$ declares system i infeasible with respect to threshold $q_{\ell m}$.

(4) *If* $v_{i\ell}^{\mathrm{UB}} \leq q_{\ell m} \leq v_{i\ell}^{\mathrm{LB}}$ *and* $\mathrm{LAST}_{i\ell} = \mathrm{LB}$*, both* $\mathcal{MPP}^{(w)}$ *and* $\mathcal{RF}$ *declare system* i *feasible with respect to* $q_{\ell m}$*.*

By similar arguments as in the previous case, Case 4 holds.

(5) *Finally, if* $v_{i\ell}^{\mathrm{LB}} < q_{\ell m} < v_{i\ell}^{\mathrm{UB}}$*,* $\mathcal{MPP}^{(w)}$ *takes more observations and reaches the same decision made by* $\mathcal{RF}$*.*

Based on Algorithm 3, it is clear that $\mathcal{MPP}^{(w)}$ needs more than the r_i observations obtained all previous passes (i.e., passes 1 through $w-1$) and additional observations are generated using the saved random seeds from pass $w-1$. Given that $v_{i\ell}^{\mathrm{LB}} < q_{\ell m} < v_{i\ell}^{\mathrm{UB}}$, there does not exist $n_0 \leq n^{\mathrm{UB}} \leq r_i$ such that $\bar{Y}_{i\ell}(n^{\mathrm{UB}}) + R(n^{\mathrm{UB}}; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/n^{\mathrm{UB}} \leq q_{\ell m}$ nor $n_0 \leq n^{\mathrm{LB}} \leq r_i$ such that

$$\bar{Y}_{i\ell}(n^{\mathrm{LB}}) - R(n^{\mathrm{LB}}; \epsilon_\ell, \eta_\ell, S_{i\ell}^2(n_0))/n^{\mathrm{LB}} \geq q_{\ell m},$$

which implies that $\mathcal{RF}$ also has not made a feasibility decision for $q_{\ell m}$ after taking r_i observations. Thus, both $\mathcal{MPP}^{(w)}$ and $\mathcal{RF}$ proceed to obtain the same observations until they reach the first exit stage for $q_{\ell m}$ and make the same feasibility decision.

Putting all five cases together proves that $\mathcal{MPP}^{(w)}$ makes the same decisions as $\mathcal{RF}$ when $\mathcal{RF}$ is implemented for the threshold $q_{\ell m}$ where $m \in \cup_{u=1}^{w} T_\ell^{(u)}$ for $\ell = 1, \ldots, s$. By Lemma 4.2, we know that $\mathcal{RF}$ guarantees

$$\Pr\left(\cap_{\ell=1}^{s} \cap_{m \in \cup_{u=1}^{w} T_\ell^{(u)}} \mathrm{CD}_{i\ell}(q_{\ell m})\right) \geq 1 - \beta$$

and so does $\mathcal{MPP}^{(w)}$. □

4.2. *Statistical Validity of $\mathcal{MPP}$ for Multiple Systems*

In this section, we extend Theorem 4.3 to the general case with multiple systems and multiple constraints.

Theorem 4.4. *Given pre-defined threshold sets* $\{q_{\ell 1}, q_{\ell 2}, \ldots, q_{\ell d_\ell}\}$ *for* $\ell = 1, 2, \ldots, s$*, the Multi-Pass Pruning procedure* $\mathcal{MPP}^{(w)}$ *guarantees*

$$\Pr\left(\cap_{i=1}^{k} \cap_{\ell=1}^{s} \cap_{m \in \cup_{u=1}^{w} T_\ell^{(u)}} \mathrm{CD}_{i\ell}(q_{\ell m})\right) \geq 1 - \alpha.$$

Proof. When systems are simulated with CRN, we have

$$\begin{aligned}
\mathrm{PCD} &= \Pr\left(\cap_{i=1}^{k} \cap_{\ell=1}^{s} \cap_{m \in \cup_{u=1}^{w} T_\ell^{(u)}} \mathrm{CD}_{i\ell}(q_{\ell m})\right) \\
&\geq 1 - \sum_{i=1}^{k} \left(1 - \Pr\left(\cap_{\ell=1}^{s} \cap_{m \in \cup_{u=1}^{w} T_\ell^{(u)}} \mathrm{CD}_{i\ell}(q_{\ell m})\right)\right) \\
&\geq 1 - k\beta = 1 - k\frac{\alpha}{k} = 1 - \alpha,
\end{aligned}$$

where the first inequality follows from the Bonferroni inequality, the second inequality holds due to Theorem 4.3, and the second equality holds due to equation (1).

Similarly, when systems are simulated independently, we have

$$\begin{aligned}\text{PCD} &= \Pr\left(\cap_{i=1}^{k}\cap_{\ell=1}^{s}\cap_{m\in\cup_{u=1}^{w}T_\ell^{(u)}}\text{CD}_{i\ell}(q_{\ell m})\right)\\ &= \prod_{i=1}^{k}\Pr\left(\cap_{\ell=1}^{s}\cap_{m\in\cup_{u=1}^{w}T_\ell^{(u)}}\text{CD}_{i\ell}(q_{\ell m})\right)\\ &\geq (1-\beta)^k = (1-(1-(1-\alpha)^{1/k}))^k = 1-\alpha,\end{aligned}$$

where the inequality holds due to Theorem 4.3 and the third equality follows by equation (1). □

5. Experiments

In this section, we provide numerical results to demonstrate the performance of our proposed procedure compared with that of $\mathcal{RF}$. We first test whether the feasibility decisions of chosen thresholds concluded by $\mathcal{MPP}$ are identical to those by $\mathcal{RF}$ in Section 5.1, where all possible thresholds are included throughout two passes for $\mathcal{MPP}$ and thus no pruning occurs. We then compare the performance of the two procedures in terms of the number of replications in Section 5.2, where the thresholds chosen for different passes of $\mathcal{MPP}$ are chosen adapatively and designed to prune inferior systems.

Each experiment is repeated 10,000 times with $\alpha = 0.05$. The initial sample size is $n_0 = 20$ except we also consider other values of n_0 in Section 5.2.2. The tolerance level is set $\epsilon_\ell = 1/\sqrt{n_0}$ for all $\ell = 1, \ldots, s$ for Sections 5.1 and 5.2.1 and as specified in Sections 5.2.2 and 5.2.3. We report how many times $\mathcal{RF}$ and $\mathcal{MPP}$ have the same feasibility decisions for thresholds considered by $\mathcal{MPP}$ in the experiments. As stated in Algorithms 2 and 3, whenever we perform feasibility checks for the thresholds on one constraint, we collect observations across all constraints from that system and one *replication* refers to one set of observations collected across all constraints. To measure efficiency, we use $\text{REP}^{(u)}$ to denote the average number of *replications* obtained during the execution of Procedure $\mathcal{P}^{(u)}$, where $u = 1, \ldots, w$. Note that $\text{REP}^{(u)}$ is only applicable for our multi-pass procedure. We also let REP denote the overall average number of replications throughout the experiments (this applies to both $\mathcal{MPP}$ and $\mathcal{RF}$). Since Zhou et al. (2022) report that the correlation between the primary and secondary performance measures does not have a significant impact on the experimental results, we assume the observations for all performance measures from each system are independent when $s > 1$. Furthermore, given that Zhou et al. (2022) report that applying CRN does not benefit feasibility checks for subjective constraints, we consider independent systems. Finally, $d_\ell > 1$ for $\ell = 1, \ldots, s$, throughout all of our experiments. This implies that approaches (i) and (ii) of selecting β_ℓ in Algorithms 1 and 2 are the same for all $\ell = 1, \ldots, s$.

5.1. *Statistical Validity*

To show that the decisions of $\mathcal{RF}$ and $\mathcal{MPP}^{(w)}$ are identical, we consider one system with two constraints and $w = 2$. The mean performance of the system with respect to the two constraints is set to $y_{i,1} = y_{i,2} = 0$, and the variances are set as $\sigma_{i,1}^2 = \sigma_{i,2}^2 = 1$.

We let both constraints have four thresholds as $q_{\ell 1} = -3\epsilon_\ell, q_{\ell 2} = -\epsilon_\ell, q_{\ell 3} = \epsilon_\ell$, and $q_{\ell 4} = 3\epsilon_\ell$ where $\ell = 1, 2$.

We test all thresholds for $\mathcal{RF}$. For $\mathcal{MPP}^{(2)}$, we consider three different scenarios as below:

- Scenario 1: $T_1^{(1)} = \{1, 4\}, T_2^{(1)} = \{2, 3\}, T_1^{(2)} = \{2, 3\}$, and $T_2^{(2)} = \{1, 4\}$;
- Scenario 2: $T_1^{(1)} = T_2^{(1)} = \{1, 4\}$, and $T_1^{(2)} = T_2^{(2)} = \{2, 3\}$;
- Scenario 3: $T_1^{(1)} = T_2^{(1)} = \{2, 3\}$, and $T_1^{(2)} = T_2^{(2)} = \{1, 4\}$.

Note that Scenario 1 concerns the difference in the difficulty of the feasibility checks between the two constraints. More specifically, the feasibility check for the first (second) constraint is easy (difficult) during $\mathcal{P}^{(1)}$, while it is the opposite for $\mathcal{P}^{(2)}$. The overall difficulty for $\mathcal{P}^{(1)}$ and $\mathcal{P}^{(2)}$ is the same. On the other hand, Scenarios 2 and 3 address the difference between the two passes, where Scenario 2 (3) has a relatively easier first (second) pass. During each pass, the overall difficulty for both constraints is the same.

Among the total 10,000 repeated runs, we count the number of runs when the feasibility decisions with respect to all thresholds tested match exactly for $\mathcal{RF}$ and $\mathcal{MPP}^{(2)}$. Table 1 shows the ratio of the runs that have all decisions matched along with the estimated PCD for the two procedures. We also report $\text{REP}^{(1)}, \text{REP}^{(2)}$, and REP. As all possible thresholds are tested (rather than excluding unnecessary thresholds) throughout the execution of $\mathcal{MPP}^{(2)}$, we expect that $\text{REP}^{(1)} + \text{REP}^{(2)} = \text{REP}$ for $\mathcal{MPP}^{(2)}$ and report the ratio of the runs with the same total number of replications.

Table 1. Average number of replications and estimated PCD for $k = 1$ system and $s = 2$ constraints for $\mathcal{RF}$ and $\mathcal{MPP}^{(2)}$, where $\mathcal{MPP}^{(2)}$ is tested under three scenarios.

	$\mathcal{MPP}^{(2)}$ Scenario 1	$\mathcal{MPP}^{(2)}$ Scenario 2	$\mathcal{MPP}^{(2)}$ Scenario 3	$\mathcal{RF}$
PCD	0.9583	0.9583	0.9583	0.9583
Ratio of matched decisions	100%			
$\text{REP}^{(1)}$	79.44	37.82	95.17	—
$\text{REP}^{(2)}$	15.73	57.36	0.00	—
REP	95.17	95.17	95.17	95.17
Ratio of matched REP	100%			

Table 1 shows that the two procedures have exactly the same feasibility decisions and use the same total number of replications as expected. Comparing the results of $\mathcal{MPP}^{(2)}$ under the three scenarios, we see that Scenario 2 has the lowest $\text{REP}^{(1)}$ and Scenario 3 achieves the highest $\text{REP}^{(1)}$. This is expected since Scenario 1 has the easiest first pass among the three scenarios (two easy constraints) while Scenario 3 has the most difficult first pass (two hard constraints). Similarly, as Scenario 1 has a more (less) difficult first pass compared with Scenario 2 (3) (because Scenario 1 has one difficult and one easy constraint), we also see that Scenario 1 has a larger (smaller) $\text{REP}^{(1)}$ compared with Scenario 2 (3). Furthermore, as the number of *effective* thresholds per constraint is at most two (see Theorem 2.1), it is clear that Scenario 1 requires additional replications for the second constraint but not for the first constraint during the second pass (as the second pass adds more difficult (easier) thresholds for the first (second) constraint). For similar reasons, Scenario 2 is expected to require more repli-

cations for both constraints, whereas Scenario 3 is not expected to require additional replications for either constraint. This matches with the results that Scenario 2 has a higher $\text{REP}^{(2)}$ than Scenario 1 and Scenario 3 incurs a zero $\text{REP}^{(2)}$.

5.2. *Efficiency*

We show the efficiency of $\mathcal{MPP}$ compared to $\mathcal{RF}$ in this section when the main goal is to prune inferior systems by finding feasible systems with respect to the most preferred thresholds possible. Section 5.2.1 considers multiple systems with a single constraint, and Section 5.2.2 provides results in cases where $\mathcal{MPP}$ yields large savings. Finally, Section 5.2.3 addresses the efficiency of $\mathcal{MPP}$ compared with $\mathcal{RF}$ when multiple systems and two constraints are considered in an inventory example (as described in Section 1).

Since the thresholds tested by $\mathcal{MPP}^{(w)}$ may be a subset of all possible thresholds, we let $\widetilde{\text{PCD}}$ be the probability of correct decision with respect to the thresholds tested for $\mathcal{RF}$ or $\mathcal{MPP}^{(w)}$, i.e.,

$$\widetilde{\text{PCD}} = \begin{cases} \text{PCD}, & \text{for } \mathcal{RF}, \\ \Pr\left(\cap_{i=1}^{k} \cap_{\ell=1}^{s} \cap_{m \in \cup_{u=1}^{w} T_\ell^{(u)}} \text{CD}_{i\ell}(q_{\ell m})\right), & \text{for } \mathcal{MPP}^{(w)}. \end{cases}$$

We report the estimated $\widetilde{\text{PCD}}$ in our experimental results.

5.2.1. Systems with One Constraint

We consider $k = 100$ systems with a single constraint and 100 thresholds (i.e., $d_1 = 100$). We set the difference between two consecutive thresholds to $2\epsilon_1$, i.e., $q_{1,m} = (2m-1)\epsilon_1$ where $m = 1, \ldots, 100$. We assume that the decision maker prefers systems with smaller means.

As $\mathcal{MPP}^{(2)}$ only tests a subset of thresholds through each pass and adds thresholds in the second pass depending on the feasibility decisions obtained in the first pass, it will perform feasibility checks with respect to a restricted set of thresholds that are close to the means of the potential best systems throughout the two passes. On the other hand, $\mathcal{RF}$ collects observations for all systems with respect to all possible thresholds considered, regardless of whether it is unnecessary to conclude feasibility decisions for some of the thresholds. We use the Concentrated Mean (CM) configuration to demonstrate the case when the mean of all systems are the same except for the best system and the common mean of all inferior systems is quite far from the mean of the best system. This configuration benefits $\mathcal{MPP}^{(2)}$ as $\mathcal{MPP}^{(2)}$ is likely to identify the smallest threshold that the best system is declared feasible to and use it to prune all inferior systems (as the mean difference between inferior systems and the best system is large, the pruning becomes easy), whereas $\mathcal{RF}$ can spend more observations to conclude feasibility decisions for all the inferior systems with respect to their closest thresholds. To be more practical, we also consider a Monotonically Increasing Means (MIM) configuration where there is one system in each intersection of the unacceptable and desirable regions of two consecutive thresholds. Specifically, we set the mean configurations as follows:

- CM: $y_{1,1} = 0$ and $y_{i,1} = 198\epsilon_1$ for $i = 2, \ldots, k$.
- MIM: $y_{i1} = 2(i-1)\epsilon_1$ for all $i = 1, \ldots, k$.

For $\mathcal{RF}$, we test all 100 thresholds together in one run. For $\mathcal{P}^{(1)}$ of $\mathcal{MPP}^{(2)}$, we consider thresholds $\{q_{1,10}, q_{1,20}, \ldots, q_{1,80}, q_{1,90}\}$, i.e., $T_1^{(1)} = \{10, 20, \ldots, 90\}$. Based on the feasibility decisions with respect to the thresholds from $\mathcal{P}^{(1)}$, if there is only one system declared feasible with respect to the tightest threshold, we terminate and select the single system as the best system. On the other hand, if there are multiple systems declared feasible with respect to the tightest threshold, we consider nine even tighter thresholds compared to this threshold. For example, if there are multiple systems declared feasible with respect to threshold $q_{1,10}$, we add $\{q_{1,1}, q_{1,2}, \ldots, q_{1,9}\}$ (with $T_1^{(2)} = \{1, 2, \ldots, 9\}$) for $\mathcal{P}^{(2)}$.

Table 2 presents the results of the estimated $\widetilde{\text{PCD}}$, the ratio of matched feasibility decisions for thresholds in $T_1^{(1)} \cup T_1^{(2)}$ over 10,000 repeated runs, and the number of replications required for $\mathcal{MPP}^{(2)}$ and $\mathcal{RF}$ under the CM and the MIM configurations. Table 3 shows the average number of feasible systems declared by $\mathcal{MPP}^{(2)}$ with respect to the tightest threshold considered in $\mathcal{P}^{(1)}$ and $\mathcal{P}^{(2)}$ under both the CM and MIM configurations. As none of the 10,000 replications execute $\mathcal{P}^{(2)}$ under the CM configuration, we do not report this value in Table 3.

Table 2. Average number of replications and estimated $\widetilde{\text{PCD}}$ for $k = 100$ systems and $s = 1$ constraint under the CM configuration.

	CM		MIM	
	$\mathcal{MPP}$	$\mathcal{RF}$	$\mathcal{MPP}$	$\mathcal{RF}$
$\widetilde{\text{PCD}}$	1.000	0.9570	0.9922	0.9638
Ratio of matched decisions	100%		100%	
$\text{REP}^{(1)}$	2,009.86	—	6,065.56	—
$\text{REP}^{(2)}$	0.00	—	1,391.22	—
REP	2,009.86	18,494.22	7,456.78	18,494.24

Table 3. Average number of surviving systems throughout the execution of $\mathcal{MPP}$.

	CM	MIM
$\mathcal{P}^{(1)}$	1.00	10.00
$\mathcal{P}^{(2)}$	—	1.00

We first discuss our results under the CM configuration. We see that both $\mathcal{MPP}^{(2)}$ and $\mathcal{RF}$ guarantee statistical validity. $\mathcal{MPP}^{(2)}$ achieves a higher $\widetilde{\text{PCD}}$ than $\mathcal{RF}$ because $\mathcal{MPP}^{(2)}$ only needs to guarantee the correct decisions for a subset of the thresholds while $\mathcal{RF}$'s PCD is with respect to all 100 thresholds. In terms of the subset of thresholds tested for both $\mathcal{MPP}^{(2)}$ and $\mathcal{RF}$, the feasibility decisions are matched perfectly at 100% throughout 10,000 repeated runs. In terms of the required replications to conclude the feasibility decisions, we observe that $\mathcal{MPP}^{(2)}$ requires only 10.87% of the replications $\mathcal{RF}$ used. This significant reduction is expected as $\mathcal{RF}$ performs feasibility checks with respect to all thresholds considered, whereas $\mathcal{MPP}^{(2)}$ only tests preferred thresholds and saves unnecessary replications needed to declare feasibility

with respect to clearly less preferred thresholds. More specifically, we see from Table 3 that there is only one system declared feasible to the tightest threshold, and thus, $\mathcal{MPP}^{(2)}$ does not require the second pass.

Overall, we observe a similar tendency for the MIM configuration as in the CM configuration. The main difference is that $\mathcal{MPP}^{(2)}$ requires 40.32% of the replications $\mathcal{RF}$ used, and this savings is relatively smaller than that from the CM configuration as expected. Note that the MIM configuration has systems spread out evenly over the 100 thresholds, while the majority of systems in the CM configuration are only feasible to the least preferred threshold. As shown in Table 3, $\mathcal{P}^{(1)}$ identifies an average of 10 feasible systems to the tightest threshold under the MIM configuration while identifies only one feasible system under the CM configuration. This means that with the same subset of thresholds chosen in $\mathcal{P}^{(1)}$, the MIM configuration is more likely to require the second pass in order to further prune inferior systems, and thus, requires more replications than the CM configuration.

One may also notice that $\mathcal{RF}$ achieves similar REP under both the CM and the MIM configurations. Although the systems' means are set differently, $\mathcal{RF}$ essentially declares feasibility with respect to *one effective* threshold for the best system (i.e., ϵ_1) and *two effective* thresholds for all inferior systems (i.e., $y_{i,1} - \epsilon_1$ and $y_{i,1} + \epsilon_1$) as discussed in Zhou et al. (2022). Thus, it is expected that $\mathcal{RF}$ results in a similar REP under both configurations.

5.2.2. Systems with Large Savings

Section 5.2.1 shows that when inferior systems are pruned, $\mathcal{MPP}$ achieves significant savings compared with $\mathcal{RF}$. In this section, we demonstrate that the savings of $\mathcal{MPP}$ can be huge in certain settings.

We consider the CM mean configuration as described in Section 5.2.1 for a single constraint case. We see that to identify the best system, a feasibility decision is basically needed for one critical threshold, i.e., any threshold that is in between the best and second best system means. In this case, $\mathcal{MPP}$ only requires one pass to achieve the objective (i.e., $w = 1$) and the decision maker chooses $T_1^{(1)} = \{50\}$ (i.e., the threshold in the middle of the best system mean and the mean of all inferior systems). She considers all thresholds for $\mathcal{RF}$. The difficulty of the feasibility checks of a specific constraint depends highly on the number of systems k considered and on the minimum distance between the thresholds and the systems' means (ϵ_1 for $\mathcal{RF}$). In the following experiments, we adjust the values of k and ϵ_1 to further test how these factors affect the relative performance of $\mathcal{MPP}$ and $\mathcal{RF}$. More specifically, we use CM(k, ϵ_1) to denote the CM configuration where k systems and tolerance level ϵ_1 are considered. Table 4 presents the results when $k \in \{100, 1000, 10000\}$ and $\epsilon_1 = 1/\sqrt{n_0} \approx 0.22$ and Table 5 shows the results when $\epsilon_1 \in \{1/\sqrt{n_0}, 0.1, 0.05\}$ and $k = 100$.

Table 4. Average number of replications and estimated $\widetilde{\text{PCD}}$ for $\mathcal{MPP}^{(1)}$, where $\epsilon_1 = 1/\sqrt{n_0}$ and $k \in \{100, 1000, 10000\}$.

	CM$(100, 1/\sqrt{n_0})$		CM$(1000, 1/\sqrt{n_0})$		CM$(10000, 1/\sqrt{n_0})$	
	$\mathcal{MPP}^{(1)}$	$\mathcal{RF}$	$\mathcal{MPP}^{(1)}$	$\mathcal{RF}$	$\mathcal{MPP}^{(1)}$	$\mathcal{RF}$
$\widetilde{\text{PCD}}$	1.000	0.957	1.000	0.954	1.000	0.956
Ratio of matched decisions	100%		100%		100%	
REP	2,000.00	18,494.24	20,000.00	268,895.14	200,000.00	3,741,675.70

Table 5. Average number of replications and estimated $\widetilde{\text{PCD}}$ for $\mathcal{MPP}^{(1)}$, where $k = 100$ and $\epsilon_1 \in \{1/\sqrt{n_0}, 0.1, 0.05\}$.

	CM$(100, 1/\sqrt{n_0})$		CM$(100, 0.1)$		CM$(100, 0.05)$	
	$\mathcal{MPP}^{(1)}$	$\mathcal{RF}$	$\mathcal{MPP}^{(1)}$	$\mathcal{RF}$	$\mathcal{MPP}^{(1)}$	$\mathcal{RF}$
$\widetilde{\text{PCD}}$	1.000	0.957	1.000	0.976	1.000	0.975
Ratio of matched decisions	100%		100%		100%	
REP	2,000.00	18,494.24	2,007.47	91,852.00	4,713.99	366,307.30

From Table 4, we see that $\mathcal{MPP}^{(1)}$ requires only $10.81\%, 7.44\%, 5.35\%$ REP compared of that of $\mathcal{RF}$ when $k = 100, 1000, 10000$ and $\epsilon_1 = 1/\sqrt{n_0}$, respectively. As $\mathcal{RF}$ spends unnecessary replications trying to conclude feasibility decisions with respect to all thresholds for all systems while $\mathcal{MPP}^{(1)}$ only performs feasibility check with respect to one critical threshold to identify the best system, the huge savings of $\mathcal{MPP}^{(1)}$ is expected. We further note that $\mathcal{MPP}^{(1)}$ stops in all cases after taking the minimum number $n_0 = 20$ replications per system. Therefore, we test $n_0 \in \{5, 10\}$ and $\epsilon_1 = 1/\sqrt{20}$ as well. To conserve space, we do not include the detailed results here, but for $n_0 = 10$, $\mathcal{MPP}^{(1)}$ only uses 3.32%, 2.00%, 1.55% REP compared to $\mathcal{RF}$, and for $n_0 = 5$, these percentages become 1.47%, 1.47%, 1.48% for $k = 100, 1000, 10000$, respectively.

We also see from Table 5 that $\mathcal{MPP}^{(1)}$ uses $10.81\%, 2.19\%, 1.29\%$ REP compared of that of $\mathcal{RF}$ when $n_0 = 20, k = 100$, and $\epsilon_1 = 1/\sqrt{n_0}, 0.1, 0.05$, respectively. We again consider $n_0 \in \{5, 10\}$ but omit the details for reasons of brevity. For $n_0 = 10$, these percentages become 3.32%, 1.37%, 1.33%, and for $n_0 = 5$, they are 1.47%, 1.42%, and 1.42% for $\epsilon_1 = 1/\sqrt{20}, 0.1, 0.05$, respectively. In conclusion, the results in this section show that $\mathcal{MPP}$ can achieve huge savings compared with $\mathcal{RF}$, especially when the number of systems and thresholds considered are large and take wide ranges of values.

5.2.3. Inventory Example

In this section, we demonstrate the performance of $\mathcal{MPP}^{(3)}$ based on an (s, S) inventory example as discussed in Section 1. We consider a similar problem setting as in Koenig and Law (1985), also tested in Zhou et al. (2022), where one review period is one month and the performance measures are estimated using the first 30 months. The two performance measures are the same as in Section 1, namely the probability that a shortage occurs during each review period ($\ell = 1$) and the expected cost per review period ($\ell = 2$). The expected cost includes the ordering cost, holding cost, and penalty cost when the demand is more than the inventory level. We set the ordering cost as 3 per item and a fixed cost of 32 per order. The holding cost is set as 1 per item between each pair of consecutive periods, and the penalty cost is 5 per item of each unsatisfied demand. The demand during each period is assumed to follow a Poisson distribution with mean 25. We assume demands over different review periods are independent. We consider 2,901 systems in total as $\Theta = \{(s, S) \mid 20 \le s \le 80, 40 \le S \le 100, s \in \mathbb{Z}^+, S \in \mathbb{Z}^+, \text{ and } s \le S\}$. To reduce the initialization bias, both performance measures are computed after the first 100 review periods and averaged over the subsequent 30 review periods. We obtain analytical results for both performance measures using an steady-state analysis of a Markov chain model. We estimate the correlation between the two performance measures among all 2,901 systems using simulation with 1,000,000 replications. The estimated correlations

range from -0.235 to 0.553.

We consider the same threshold setting as discussed in Section 1, where q_1 takes values in $\{0.01+0.01\gamma \mid 0 \leq \gamma \leq 19, \gamma \in \mathbb{Z}\}$ and q_2 takes values in $\{115+0.5\gamma \mid 0 \leq \gamma \leq 120, \gamma \in \mathbb{Z}\}$ thousands (20 values for q_1 and 121 values for q_2). We include all thresholds from both constraints for $\mathcal{RF}$. To choose the thresholds $\mathcal{MPP}^{(3)}$ uses to prune inferior systems in this multi-objective setting, we utilize the concept of preference order introduced in Zhou, Andradóttir, and Kim (2023). Preference order is used to describe how the decision maker prioritizes different constraints based on the given thresholds on each constraint. Zhou, Andradóttir, and Kim (2023) propose three preference orders, namely the ranked constraints, equally important constraints, and the total violation with ranked constraints formulations. In this section, we focus on the equally important constraints formulation, where the decision maker values both constraints equally and would like to tighten (relax) both constraints at the same time if multiple (no) feasible systems are identified with respect to the more (less) preferred threshold combination. More specifically, we first consider thresholds $q_1 \in \{0.01, 0.1, 0.2\}$ and $q_2 \in \{115, 145, 175\}$ for $\mathcal{P}^{(1)}$. If multiple systems are declared feasible to the most preferred possible threshold combination, we tighten both constraints by adding thresholds on a finer level such that q_1 is chosen with an increment of 0.05 and q_2 with an increment of 5 for $\mathcal{P}^{(2)}$. Similarly, we consider an even finer level of thresholds in $\mathcal{P}^{(3)}$ by choosing q_1 at an increment of 0.01 and q_2 at an increment of 0.5. For example, if multiple feasible systems are identified by $\mathcal{P}^{(1)}$ with respect to threshold combinations $(q_1, q_2) = (0.1, 145), (0.01, 175)$, and $(0.1, 115)$, but no systems are declared feasible with respect to combination $(q_1, q_2) = (0.01, 115)$, then consistent with the equally important constraints formulation, we choose $(0.1, 145)$ as the most preferred threshold combination and add thresholds $q_1 \in \{0.05\}$ and $q_2 \in \{120, 125, 130, 135, 140\}$ for $\mathcal{P}^{(2)}$. If multiple feasible systems are further identified with respect to threshold combination $(q_1, q_2) = (0.05, 120)$, we add $q_1 \in \{0.01 + 0.01\gamma \mid 1 \leq \gamma \leq 3, \gamma \in \mathbb{Z}\}$ and $q_2 \in \{115 + 0.5\gamma \mid 1 \leq \gamma \leq 9, \gamma \in \mathbb{Z}\}$ for $\mathcal{P}^{(3)}$. This is equivalent to setting $T_1^{(1)} = \{1, 10, 20\}, T_2^{(1)} = \{1, 61, 121\}, T_1^{(2)} = \{5\}, T_2^{(2)} = \{11, 21, 31, 41, 51\}, T_1^{(3)} = \{2, 3, 4\}$, and $T_2^{(3)} = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

To get a sense of the number of feasible systems with respect to different threshold combinations, we present Table 6 for the number of feasible systems with respect to $5 \times 13 = 65$ of the $20 \times 121 = 2420$ combinations of thresholds of both constraints based on their analytical values. As feasible systems are likely to be identified with respect to threshold combination $(q_1, q_2) = (0.1, 145)$ after the completion of $\mathcal{P}^{(1)}$ and are also likely to be identified with respect to $(q_1, q_2) = (0.05, 120)$ throughout $\mathcal{P}^{(2)}$, we further include Table 7 to show the number of feasible systems with respect to the threshold combinations, where $q_1 \in \{0.01, 0.02, \ldots, 0.05\}$ and $q_2 \in \{115, 115.5, 116, \ldots, 119.5, 120\}$. Note that we do not present the analytical values for all 2420 combinations (as discussed in Section 1) for simplicity.

Due to the nature of the two constraints (shortage probability is likely between 0.9 to 1 and the expected cost is likely between 110 to 180), we set the tolerance-level for the shortage probability constraint as $\epsilon_1 = 0.001$ and for the expected cost constraint as $\epsilon_2 = 0.1$. Table 8 presents the estimated $\widetilde{\text{PCD}}$ and the average number of replications needed for the $\mathcal{MPP}^{(3)}$ and $\mathcal{RF}$ procedures as well as the ratio of the matched feasibility decisions.

Similar to Sections 5.2.1 and 5.2.2, both $\mathcal{MPP}^{(3)}$ and $\mathcal{RF}$ guarantee statistical validity even though Assumption 1 is violated due to the observations on each constraint not being normally-distributed. We also see that $\mathcal{MPP}^{(3)}$ results in a slightly

Table 6. Number of feasible systems with respect to a grid of 65 combinations of constraint thresholds.

q_1 \ q_2	115	120	125	130	135	140	145	150	155	160	165	170	175
0.01	0	31	221	563	914	1210	1470	1705	1902	2052	2176	2265	2317
0.05	31	178	526	923	1274	1570	1830	2065	2262	2412	2536	2625	2677
0.1	74	309	675	1081	1432	1728	1988	2223	2420	2570	2694	2783	2835
0.15	94	345	711	1117	1468	1764	2024	2259	2456	2606	2730	2819	2871
0.2	108	364	730	1136	1487	1783	2043	2278	2475	2625	2749	2838	2890

Table 7. Number of feasible systems with respect to a finer grid of 55 threshold combinations.

q_1 \ q_2	115	115.5	116	116.5	117	117.5	118	118.5	119	119.5	120
0.01	0	0	0	0	0	1	5	12	17	27	31
0.02	0	0	2	8	15	23	31	40	47	61	67
0.03	6	10	18	27	36	45	56	67	77	92	104
0.04	18	24	34	44	56	65	79	91	102	125	142
0.05	31	38	49	60	73	84	99	117	130	158	178

Table 8. Average number of replications and estimated $\widetilde{\text{PCD}}$ for the inventory example.

	$\mathcal{MPP}^{(3)}$	$\mathcal{RF}$
$\widetilde{\text{PCD}}$	0.966	0.964
Ratio of matched decisions	100%	
$\text{REP}^{(1)}$	4,648,089	—
$\text{REP}^{(2)}$	9,123,765	—
$\text{REP}^{(3)}$	6,210,287	—
REP	19,982,140	65,293,513

higher $\widetilde{\text{PCD}}$ compared with $\mathcal{RF}$ since $\mathcal{MPP}^{(3)}$ only needs to guarantee correct decision for a subset of the thresholds while $\mathcal{RF}$ guarantees correct decisions for all possible thresholds. $\mathcal{MPP}^{(3)}$ achieves huge savings by only requiring about 30.60% of the replications compared with $\mathcal{RF}$. The values of $\text{REP}^{(1)}, \text{REP}^{(2)}$, and $\text{REP}^{(3)}$ likely depend on the difficulty of each pass and the number of replications collected already in previous passes. For each pass, the difficulty mainly depends on the number of surviving systems and whether the system means are close to the thresholds considered (see additional discussion in Section 5.2.2). Pass 1 is easy in general. Although we need to run replications from all 2,901 systems, the thresholds are easy to test (because many systems have means that are far from the thresholds considered, see Table 6). Pass 2 considers more preferred thresholds and tests only surviving systems from Pass 1, and thus, those thresholds can be closer to the surviving systems' means for the two constraints, which makes the feasibility checks more difficult. Also, since Pass 1 is relatively easy, the number of collected replications so far is not large, and hence

does not benefit Pass 2 significantly. This explains why $\text{REP}^{(2)}$ is higher than $\text{REP}^{(1)}$. Similar reasons apply to Pass 3, except with fewer surviving systems and many replications that Pass 3 can utilize from Passes 1 and 2, and hence Pass 3 does not require as many additional replications as Pass 2.

6. Conclusion

We consider the problem of pruning inferior systems among finitely many simulated systems using subjective stochastic constraints with sequentially added thresholds. When some systems are concluded feasible with respect to preferred thresholds, the decision maker can prune systems that are declared infeasible to those thresholds in order to avoid collecting unnecessary observations from the inferior systems. We propose an indifference-zone $\mathcal{MPP}$ procedure that initially tests a subset of thresholds and allows thresholds to be added sequentially if needed without requiring much data storage. We prove that $\mathcal{MPP}$ guarantees statistical validity and show by experiments that it can achieve large savings in terms of the required replications compared with $\mathcal{RF}$ if the decision maker aims to prune inferior systems using subjective constraints.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

The second author was supported by NSF under grant CMMI-2127778 and the last author was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (No. 2022R1F1A1063147).

References

Andradóttir, Sigrún, and Seong-Hee Kim. 2010. "Fully sequential procedures for comparing constrained systems via simulation." *Naval Research Logistics (NRL)* 57 (5): 403–421.

Andradóttir, Sigrún, and Judy S. Lee. 2021. "Pareto set estimation with guaranteed probability of correct selection." *European Journal of Operational Research* 292 (1): 286–298.

Batur, Demet, and Seong-Hee Kim. 2010. "Finding Feasible Systems in the Presence of Constraints on Multiple Performance Measures." *ACM Transactions on Modeling and Computer Simulation* 20 (3): Article 13.

Healey, Christopher, Sigrún Andradóttir, and Seong-Hee Kim. 2014. "Selection Procedures for Simulations with Multiple Constraints under Independent and Correlated Sampling." *ACM Transactions on Modeling and Computer Simulation* 24 (3): Article 14.

Hong, L. Jeff, Barry L. Nelson, and Jie Xu. 2015. "Discrete Optimization via Simulation." In *Handbook of Simulation Optimization*, edited by Michael C Fu, 9–44. New York, NY: Springer.

Hunter, Susan R., Eric A. Applegate, Viplove Arora, Bryan Chong, Kyle Cooper, Oscar Rincón-Guevara, and Carolina Vivas-Valencia. 2019. "An Introduction to Multiobjective Simulation Optimization." *ACM Transactions on Modeling and Computer Simulation* 29 (1): Article 7.

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

Kim, Seong-Hee, and Barry L. Nelson. 2006. "Selecting the Best System: Simulation." In *Handbooks in Operations Research and Management Science*, edited by Shane G. Henderson and Barry L. Nelson, 501–534. Elsevier.

Koenig, Lloyd W., and Averill M. Law. 1985. "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* 14 (3): 719–734.

Lee, Loo Hay, Ek Peng Chew, Suyan Teng, and David Goldsman. 2010. "Finding the non-dominated Pareto set for multi-objective simulation models." *IIE Transactions* 42 (9): 656–674.

Lee, Loo Hay, Nugroho Artadi Pujowidianto, Ling-Wei Li, Chun-Hung Chen, and Chee Meng Yap. 2012. "Approximate Simulation Budget Allocation for Selecting the Best Design in the Presence of Stochastic Constraints." *IEEE Transactions on Automatic Control* 57 (11): 2940–2945.

Zhou, Yuwei, Sigrún Andradóttir, and Seong-Hee Kim. 2023. "Selection of the Best in the Presense of Subjective Constraints." *Submitted for publication* .

Zhou, Yuwei, Sigrún Andradóttir, Seong-Hee Kim, and Chuljin Park. 2022. "Finding Feasible Systems for Subjective Constraints Using Recycled Observations." *INFORMS Journal on Computing* 34 (6): 3080–3095.