Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

FINDING FEASIBLE SYSTEMS FOR A STOCHASTIC CONSTRAINT WITH RELAXED TOLERANCE LEVELS

Chuljin Park[1], Sigrún Andradóttir[2], Seong-Hee Kim[2], and Yuwei Zhou[3]

[1]Dept. of Industrial Engineering, Hanyang University, Seoul, SOUTH KOREA
[2]H. Milton Stewart School of Industrial and Systems Eng., Georgia Tech, Atlanta, GA, USA
[3]Kelley School of Business, Indiana University, Bloomington, IN, USA

ABSTRACT

We consider the problem of finding feasible systems with respect to a stochastic constraint when the performance of each system needs to be evaluated via simulation. We develop a new procedure, referred to as the Indifference-Zone Relaxation procedure, to lessen inefficiencies of existing procedures derived under the assumption that all systems are exactly the tolerance level away from the threshold. Specifically, our procedure introduces a set of relaxed tolerance levels and simultaneously implements two subroutines for each relaxed tolerance level: one to identify clearly feasible systems and the other to exclude clearly infeasible systems. As a result, the proposed procedure allows early determination of feasibility for some systems, while maintaining the statistical guarantee. The efficiency of the procedure is investigated through experimental results.

1 INTRODUCTION

Ranking and selection (R&S) procedures have been developed and used for finding a system with the best performance among finitely many systems when the performance is evaluated by stochastic simulation. General approaches to R&S fall in several categories, including fully-sequential indifference-zone (IZ) procedures (Kim and Nelson 2006), optimal computing budget allocation (OCBA) procedures (Chen et al. 2000), and Bayesian procedures (Chick 2006).

Constrained R&S involves selecting the best system with respect to a primary performance measure while satisfying stochastic constraints on secondary performance measures. Andradóttir and Kim (2010) and Batur and Kim (2010) develop fully-sequential IZ feasibility check procedures (FCPs) to find a set of feasible systems with respect to a constraint and multiple constraints, respectively. These FCPs are employed as subroutines to select a system with the best performance among systems satisfying constraints on one or more secondary performance measures. Andradóttir and Kim (2010) and Healey et al. (2013) propose statistically valid procedures that select the best feasible system with at least a pre-specified probability under a stochastic constraint. Healey et al. (2014) extend the procedures to handle multiple stochastic constraints and to find the best system among the feasible systems. As OCBA procedures, Lee et al. (2012) provide a procedure that allocates a finite simulation budget to maximize the probability of correctly selecting the best system satisfying multiple stochastic constraints. Furthermore, Hunter and Pasupathy (2013), Pasupathy et al. (2015), and Gao and Chen (2017) enhance a feasibility determination for multiple stochastic constraints considering the OCBA and the large deviation theory. Solow et al. (2021) handle the constraints using Bayesian concepts. In this paper, we focus on the fully-sequential IZ procedures in the presence of a stochastic constraint.

The fully-sequential IZ FCPs introduce one tolerance level, that specifies how much the decision maker is willing to be off from the constant threshold for checking the feasibility of the systems. This is the least absolute difference in the performance measure and the constant threshold that the decision maker wants to detect. In practice, the true expected performance measure is unknown. When a system's performance measure is very close to the constant threshold, distinguishing this small difference requires a huge number of replications until the decision maker determines its feasibility. The tolerance level helps avoid such situations. Nevertheless, when the tolerance level is too small relative to the difference between the true expected performance measure and the constant threshold, computational costs for feasibility checks may be unnecessarily high. Therefore, a small value of the tolerance level may aggravate the computational burden when the number of systems becomes large and their performance measures quite vary. Lee et al. (2018) propose the adaptive feasibility check procedure that uses existing FCPs (Batur and Kim 2010) with different thresholds as its subroutines and introduce a decreasing sequence of tolerance levels within the subroutines. The procedure is designed to self-adjust tolerance levels regarding each system and each constraint by introducing two independent FCPs with different thresholds and the same tolerance level. Even though the procedure shows strong empirical performance in numerical and practical examples, it is found that the procedure may terminate without keeping the statistical guarantee when the expected performance measure is very close to the constant threshold.

In this paper, we propose a new FCP, referred to as the Indifference-Zone Relaxation ($\mathcal{IZR}$) procedure, that (i) introduces a set of relaxed tolerance levels and (ii) simultaneously runs two subroutines with different thresholds and the relaxed tolerance levels. In particular, a subroutine of $\mathcal{IZR}$ mainly distinguishes clearly infeasible systems, and the other subroutine distinguishes clearly feasible systems. Then, the algorithm makes the feasibility decision of the system whenever the two subroutines provide the same feasibility decision (i.e., either feasible or infeasible) for the same tolerance level. As a result, the feasibility decision of different systems may be made with different tolerance levels and this leads to saving computational costs while keeping the statistical guarantee with respect to the original tolerance level.

The rest of the paper is organized as follows: Section 2 provides the background for our problem. Section 3 proposes our $\mathcal{IZR}$ procedure and Section 4 presents numerical results for our procedure and compares its performance with that of an existing procedure. Concluding remarks are provided in Section 5. Note that Zhou et al. (2024) provide additional details for this paper, including the proof of the statistical guarantee of the procedure.

2 BACKGROUNDS

We consider k systems whose performance can be observed through stochastic simulation and one constraint on a performance measure. Let Θ denote the set of all systems (i.e., $\Theta = \{1, 2, \ldots, k\}$) and Y_{ij} for $j = 1, 2, \ldots,$ denote the jth simulation observation associated with the performance measure of system i. For any given system i, y_i denotes the expectation of Y_{ij} (i.e., $y_i = \mathbf{E}[Y_{ij}]$) and σ_i^2 denotes the variance of Y_{ij} (i.e., $\sigma_i^2 = \mathbf{Var}[Y_{ij}]$). Let q be a given constant threshold in the constraint. System i is defined as feasible if $y_i \leq q$, and infeasible otherwise. Our problem is to determine the set of systems with $y_i \leq q$. We assume that the observations satisfy the following assumption throughout the paper:

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

$$Y_{ij} \overset{iid}{\sim} N(y_i, \sigma_i^2),$$

where $\overset{iid}{\sim}$ denotes independent and identically distributed, and N denotes a normal distribution.

To define a correct decision event, Andradóttir and Kim (2010) introduce a tolerance level, denoted by ϵ, which is a user-specified positive real number. System i falls in one of the following three sets:

$$\begin{aligned} D &\equiv \{i \in \Theta \mid y_i \leq q - \epsilon\}\,; \\ A &\equiv \{i \in \Theta \mid q - \epsilon < y_i < q + \epsilon\}\,; \text{ and} \\ U &\equiv \{i \in \Theta \mid q + \epsilon \leq y_i\}\,. \end{aligned}$$

The systems in the sets D, A, and U are called desirable, acceptable, and unacceptable systems, respectively. A correct decision for system i is denoted by CD_i and it is defined as declaring system i feasible if $i \in D$

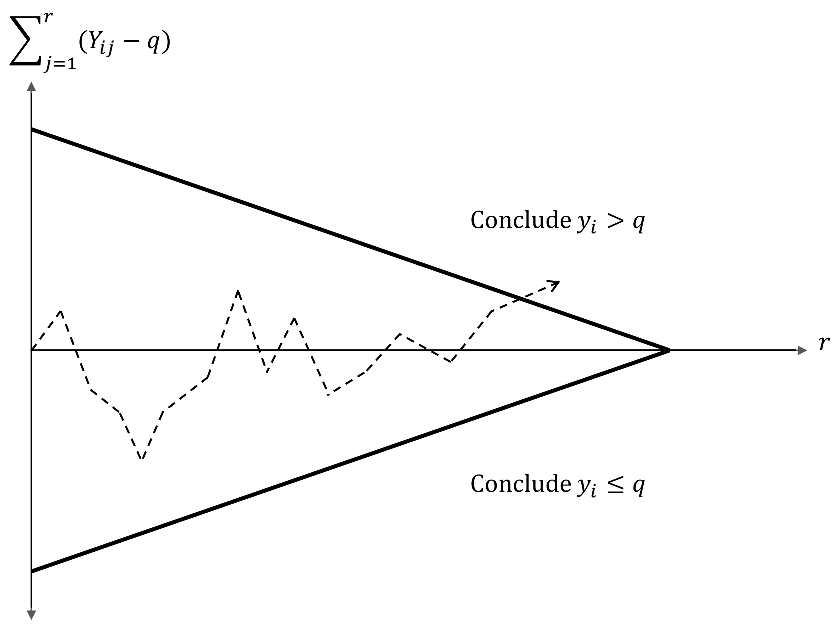


Figure 1: A triangular continuation region for system i.

and infeasible if $i \in U$. Any decision is considered as a correct decision for $i \in A$. It is noted that the acceptable region (which is equivalent to the IZ) is defined as the interval from $q - \epsilon$ to $q + \epsilon$. Then, a correct decision for the problem is defined as the event that the correct decisions for all systems are simultaneously made ($\text{CD} = \cap_{i=1}^{k}\text{CD}_i$), and a statistically valid procedure should satisfy $\mathbf{P}(\text{CD}) \geq 1 - \alpha$ where α is a confidence level.

Procedure $\mathcal{F}$ due to Andradóttir and Kim (2010) is one of the statistically valid procedures for feasibility determination. It is based on a fully-sequential R&S procedure in Fabian (1974) and Kim and Nelson (2001). Specifically, the procedure requires keeping track of a monitoring statistic that is a cumulative sum of the difference between Y_{ij} and q. An observation is sampled at each stage of the procedure while the statistic sojourns within a boundary called the continuation region. When this statistic first exits the continuation region, the feasibility decision is made. Figure 1 shows a sample path of the monitoring statistics for system i (dashed line) and the boundary of the continuation region (bold line). As shown in the figure, if the statistic first exits through the upper boundary, then we conclude that system i is infeasible regarding the constraint. On the other hand, if it first exits through the lower boundary, system i is declared as feasible regarding the constraint.

As our procedure is related to $\mathcal{F}$, we provide the detailed description of $\mathcal{F}$ which requires additional notation as follows:

$n_0 \equiv$ the initial sample size for each system ($n_0 \geq 2$);
$r \equiv$ the current stage number ($r \geq n_0$);
$S_i^2 \equiv$ the sample variance of $Y_{i1}, \ldots, Y_{in_0}$ for system i ($i = 1, 2, \ldots, k$);
$M \equiv$ the set of systems whose feasibility is not determined yet; and
$F \equiv$ the set of systems declared as feasible.

Moreover, we need the following functions to define the continuation region as in Kim and Nelson (2001):

$$R(r; v, w, z) = \max\left\{0, \frac{wz}{2cv} - \frac{v}{2c}r\right\}, \text{ for any } v, w, z \in \mathbb{R},\ v \neq 0,$$

$$g(\eta) = \sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1 - \frac{1}{2}\mathbb{I}(\ell = c)\right) \times \left(1 + \frac{2\eta(2c-\ell)\ell}{c}\right)^{-(n_0-1)/2},$$

where c is a user-specified positive integer and $\mathbb{I}(\cdot)$ is an indicator function. Then, a detailed description of the existing procedure $\mathcal{F}$ is shown in Algorithm 1.

Algorithm 1: Procedure $\mathcal{F}$

[**Setup**:] Choose initial sample size $n_0 \geq 2$, confidence level $0 < 1 - \alpha < 1$, $c \in \mathbb{N}^+$, and $\Theta = \{1, 2, \ldots, k\}$. Set tolerance level $\epsilon > 0$ and threshold q. Calculate $h_B^2 = 2c\eta_B(n_0 - 1)$, where $\eta_B > 0$ satisfies

$$g(\eta_B) = \beta_B = \begin{cases} 1 - (1-\alpha)^{1/k}, & \text{if systems are independent;} \\ \alpha/k, & \text{otherwise.} \end{cases} \tag{1}$$

[**Initialization**:] Obtain initial observations $Y_{i1}, Y_{i2}, \ldots, Y_{in_0}$ and compute S_i^2 for each $i \in \Theta$. Set $r = n_0$, $M = \{1, 2, \ldots, k\}$, and $F = \emptyset$.

[**Feasibility Check**:] Set $M^{old} = M$.

for $i \in M^{old}$ **do**

If $\sum_{j=1}^{r}(Y_{ij} - q) \leq -R(r; \epsilon, h_B^2, S_i^2)$, move i from M to F.

Else if $\sum_{j=1}^{r}(Y_{ij} - q) \geq R(r; \epsilon, h_B^2, S_i^2)$, eliminate i from M.

end for

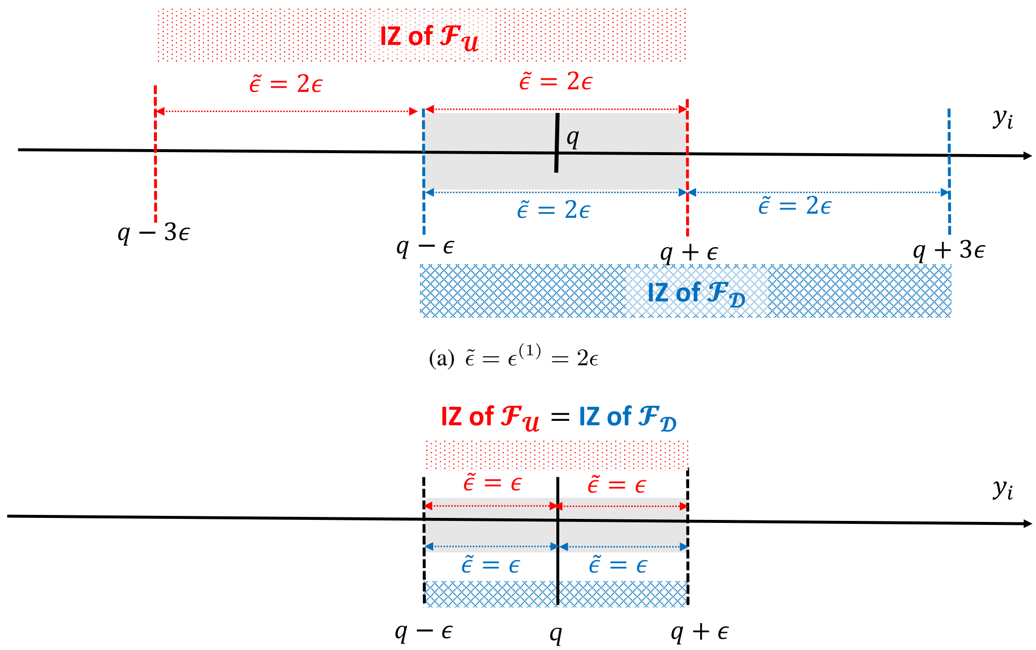
[**Stopping Condition**:] If $M = \emptyset$, return F. Otherwise, set $r = r + 1$, take one additional observation Y_{ir} for each system $i \in M$, then go to [**Feasibility Check**].

Remark 1. To implement Procedure $\mathcal{F}$, Kim and Nelson (2001) recommend the choice of $c = 1$. In this case, $g(\eta_B) = \frac{1}{2}\{1 + 2\eta_B\}^{-(n_0-1)/2}$ and the solution to equation (1) is $\eta_B = \frac{1}{2}\left[(2\beta_B)^{-2/(n_0-1)} - 1\right]$.

3 PROPOSED PROCEDURE

In this section, we provide an overall description of our procedure, denoted by $\mathcal{IZR}$. The $\mathcal{IZR}$ procedure introduces T relaxed tolerance levels, denoted by $\epsilon^{(\tau)}$ for $\tau = 1, 2, \ldots, T$, satisfying $\epsilon^{(1)} > \cdots > \epsilon^{(T)} = \epsilon$. For each relaxed tolerance level $\tilde{\epsilon} \in \{\epsilon^{(1)}, \epsilon^{(2)}, \ldots, \epsilon^{(T)}\}$, our procedure utilizes two subroutines: $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$. Subroutine $\mathcal{F}_{\mathcal{U}}$ uses threshold $q+\epsilon-\tilde{\epsilon}$ with tolerance level $\tilde{\epsilon}$ while subroutine $\mathcal{F}_{\mathcal{D}}$ uses threshold $q-\epsilon+\tilde{\epsilon}$ with tolerance level $\tilde{\epsilon}$. Subroutines $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ simultaneously check the feasibility of each system i for all tolerance levels using the monitoring statistics, $(\sum_{j=1}^{r} Y_{ij}) - r(q + \epsilon - \tilde{\epsilon})$ and $(\sum_{j=1}^{r} Y_{ij}) - r(q - \epsilon + \tilde{\epsilon})$, respectively. If both $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ result in the same feasibility decision (declaring system i as either feasible or infeasible) with some $\tilde{\epsilon} \in \{\epsilon^{(1)}, \epsilon^{(2)}, \ldots, \epsilon^{(T)}\}$, then the procedure stops taking more observations from system i and returns that feasibility decision for system i. Otherwise, the procedure keeps taking more observations from the system. Subroutines $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ maintain their own sets of relaxed tolerance levels, denoted by $\mathcal{E}_{\mathcal{U}i}$ and $\mathcal{E}_{\mathcal{D}i}$, respectively. For each system i, the sets $\mathcal{E}_{\mathcal{U}i}$ and $\mathcal{E}_{\mathcal{D}i}$ are initialized as $\{\epsilon^{(1)}, \epsilon^{(2)}, \ldots, \epsilon^{(T)}\}$ whereas they are updated during feasibility checks. These two sets may be different because (i) $\epsilon^{(\tau)}$ needs to be removed from $\mathcal{E}_{\mathcal{U}i}$ or $\mathcal{E}_{\mathcal{D}i}$ once the monitoring statistics of $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ first exit the continuation region with $\tilde{\epsilon} = \epsilon^{(\tau)}$ which is defined by $R(r; \tilde{\epsilon}, h^2, S_i^2)$, and (ii) their exit times may be different if $\tau < T$ (i.e., $\tilde{\epsilon} = \epsilon^{(\tau)} > \epsilon$).

We explain how the $\mathcal{IZR}$ procedure with subroutines $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ works using a simple example. One of the simplest ways for selecting relaxed tolerance levels $\epsilon^{(\tau)}$ is to multiply ϵ by powers of $\xi > 1$. More specifically, one can set $\epsilon^{(\tau-1)} = \xi\epsilon^{(\tau)}$ for $\tau = 2, 3, \ldots, T$ with $\epsilon^{(T)} = \epsilon$. When $\xi = 2$ and $T = 2$,



(a) $\tilde{\epsilon} = \epsilon^{(1)} = 2\epsilon$

(b) $\tilde{\epsilon} = \epsilon^{(2)} = \epsilon$

Figure 2: An example of thresholds and acceptable regions for $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ with $\xi = 2$ and $T = 2$.

Figure 2(a) presents the acceptable regions (i.e., the IZs) of $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ for $\epsilon^{(1)} = 2\epsilon$ with the red-dotted and blue-checkered intervals, respectively. Additionally, the original acceptable region with threshold q and tolerance level ϵ is shown as a gray-shaded interval. From the figure, one can see that the threshold of $\mathcal{F}_{\mathcal{U}}$ is $q + \epsilon - 2\epsilon = q - \epsilon$, and the upper bound of its red-dotted acceptable region of $\mathcal{F}_{\mathcal{U}}$ matches that of the original gray-shaded acceptable region. Similarly, the threshold of $\mathcal{F}_{\mathcal{D}}$ is $q - \epsilon + 2\epsilon = q + \epsilon$, and the lower bound of its (blue-checkered) acceptable region matches that of the original (gray-shaded) acceptable region. Figure 2(b) presents the acceptable regions of $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ when $\epsilon^{(2)} = \epsilon$. As $\epsilon^{(2)}$ is the same as the original tolerance level, $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ become identical and have the same acceptable region as $\mathcal{F}$.

Then we can consider a possible decision for system i with $\epsilon^{(1)} = 2\epsilon$ and $\epsilon^{(2)} = \epsilon$:

Case 1: If $y_i \leq q - 3\epsilon$, system i becomes desirable for both $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ under the relaxed tolerance level $\epsilon^{(1)} = 2\epsilon$ and is likely to be declared feasible by both subroutines, which is a correct decision for system i.

Case 2: If $q - 3\epsilon < y_i \leq q - \epsilon$, system i is acceptable for $\mathcal{F}_{\mathcal{U}}$ but desirable for $\mathcal{F}_{\mathcal{D}}$ under the relaxed tolerance level $\epsilon^{(1)}$. Therefore, $\mathcal{F}_{\mathcal{D}}$ is likely to declare system i feasible while $\mathcal{F}_{\mathcal{U}}$ can declare system i either feasible or infeasible. If they make the same decision (i.e., feasible decision), then the procedure stops and declares system i feasible, which is a correct decision. Otherwise, it proceeds with the subroutines $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ using the smaller tolerance level $\epsilon^{(2)}$.

Case 3: If $q - \epsilon < y_i < q + \epsilon$, system i is acceptable for both subroutines under the relaxed tolerance level $\epsilon^{(1)}$. Thus, if both subroutines make the same decision, any decision is a correct decision for system i. Otherwise, it proceeds with subroutines using the smaller tolerance level $\epsilon^{(2)}$.

Case 4: If $q + \epsilon \leq y_i < q + 3\epsilon$, system i is unacceptable for $\mathcal{F}_{\mathcal{U}}$ but acceptable for $\mathcal{F}_{\mathcal{D}}$ under the relaxed tolerance level $\epsilon^{(1)}$. By similar arguments as in Case 2, if the two subroutines make the same infeasible decision, then the procedure stops and declares system i infeasible, which is a correct decision. Otherwise, it proceeds with the subroutines using the smaller tolerance level $\epsilon^{(2)}$.

Case 5: If $y_i \geq q + 3\epsilon$, system i becomes unacceptable for both subroutines under the relaxed tolerance level $\epsilon^{(1)}$ and is likely to be declared infeasible by both subroutines, which is a correct decision for system i.

A detailed description of our proposed procedure is given in Algorithm 2.

Algorithm 2: Procedure $\mathcal{IZR}$

[**Setup**:] Choose initial sample size $n_0 \geq 2$, confidence level $0 < 1-\alpha < 1$, $c \in \mathbb{N}^+$, and $\Theta = \{1, 2, \ldots, k\}$. Set tolerance level $\epsilon > 0$, threshold q, and relaxed tolerance sets, $\mathcal{E}_{\mathcal{U}i} = \mathcal{E}_{\mathcal{D}i} = \{\epsilon^{(1)}, \ldots, \epsilon^{(T)}\}$ for each $i \in \Theta$ where $T \in \mathbb{N}^+$ and $\epsilon^{(1)} > \cdots > \epsilon^{(T)} = \epsilon$. Calculate $h^2 = 2c\eta(n_0 - 1)$, where $\eta > 0$ satisfies

$$g(\eta) = \beta = \begin{cases} \left[1 - (1-\alpha)^{1/k}\right]/T, & \text{if systems are independent;} \\ \alpha/(kT), & \text{otherwise.} \end{cases} \tag{2}$$

[**Initialization**:] For each $i \in \Theta$,
obtain initial observations, $Y_{i1}, Y_{i2}, \ldots, Y_{in_0}$, and compute S_i^2; and
set $Z_{\mathcal{U}i}^{(\tilde{\epsilon})} = Z_{\mathcal{D}i}^{(\tilde{\epsilon})} = 0$ for each $\tilde{\epsilon} \in \mathcal{E}_{\mathcal{U}i} (= \mathcal{E}_{\mathcal{D}i})$.
Set $r = n_0$, $M = \{1, 2, \ldots, k\}$, and $F = \emptyset$.
[**Feasibility Check**:] Set $M^{old} = M$.
for $i \in M^{old}$ **do**
  [**Subroutine** $\mathcal{F}_{\mathcal{U}}$:]
  for $\tilde{\epsilon} \in \mathcal{E}_{\mathcal{U}i}$ **do**,
    If $(\sum_{j=1}^{r} Y_{ij}) - r(q + \epsilon - \tilde{\epsilon}) \leq -R(r; \tilde{\epsilon}, h^2, S_i^2)$, set $Z_{\mathcal{U}i}^{(\tilde{\epsilon})} = 1$ and $\mathcal{E}_{\mathcal{U}i} = \mathcal{E}_{\mathcal{U}i} \setminus \{\tilde{\epsilon}\}$;
    Else if $(\sum_{j=1}^{r} Y_{ij}) - r(q + \epsilon - \tilde{\epsilon}) \geq R(r; \tilde{\epsilon}, h^2, S_i^2)$, set $Z_{\mathcal{U}i}^{(\tilde{\epsilon})} = -1$ and $\mathcal{E}_{\mathcal{U}i} = \mathcal{E}_{\mathcal{U}i} \setminus \{\tilde{\epsilon}\}$.
    If $Z_{\mathcal{U}i}^{(\tilde{\epsilon})} = Z_{\mathcal{D}i}^{(\tilde{\epsilon})} = 1$, move i from M to F. Else if $Z_{\mathcal{U}i}^{(\tilde{\epsilon})} = Z_{\mathcal{D}i}^{(\tilde{\epsilon})} = -1$, eliminate i from M.
  end for
  [**Subroutine** $\mathcal{F}_{\mathcal{D}}$:]
  for $\tilde{\epsilon} \in \mathcal{E}_{\mathcal{D}i}$ **do**,
    If $(\sum_{j=1}^{r} Y_{ij}) - r(q - \epsilon + \tilde{\epsilon}) \leq -R(r; \tilde{\epsilon}, h^2, S_i^2)$, set $Z_{\mathcal{D}i}^{(\tilde{\epsilon})} = 1$ and $\mathcal{E}_{\mathcal{D}i} = \mathcal{E}_{\mathcal{D}i} \setminus \{\tilde{\epsilon}\}$;
    Else if $(\sum_{j=1}^{r} Y_{ij}) - r(q - \epsilon + \tilde{\epsilon}) \geq R(r; \tilde{\epsilon}, h^2, S_i^2)$, set $Z_{\mathcal{D}i}^{(\tilde{\epsilon})} = -1$ and $\mathcal{E}_{\mathcal{D}i} = \mathcal{E}_{\mathcal{D}i} \setminus \{\tilde{\epsilon}\}$.
    If $Z_{\mathcal{U}i}^{(\tilde{\epsilon})} = Z_{\mathcal{D}i}^{(\tilde{\epsilon})} = 1$, move i from M to F. Else if $Z_{\mathcal{U}i}^{(\tilde{\epsilon})} = Z_{\mathcal{D}i}^{(\tilde{\epsilon})} = -1$, eliminate i from M.
  end for
end for
[**Stopping Condition**:] If $M = \emptyset$, return F. Otherwise, set $r = r + 1$, take one additional observation Y_{ir} for each system $i \in M$, then go to [**Feasibility Check**].

The statistical property of Algorithm 2 is provided in the following theorem whose proof is found in Zhou et al. (2024).

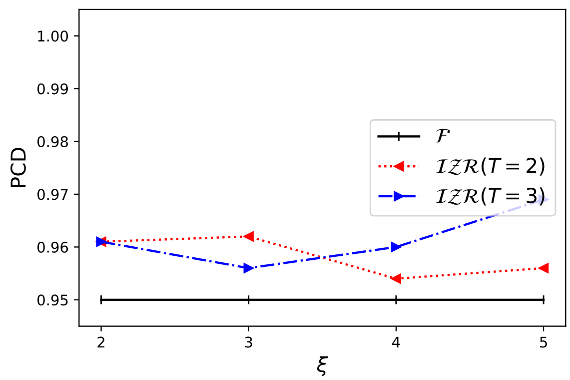
Theorem 1 Under Assumption 1, $\mathcal{IZR}$ guarantees

$$\mathbf{P}(\text{CD}) \geq 1 - \alpha.$$

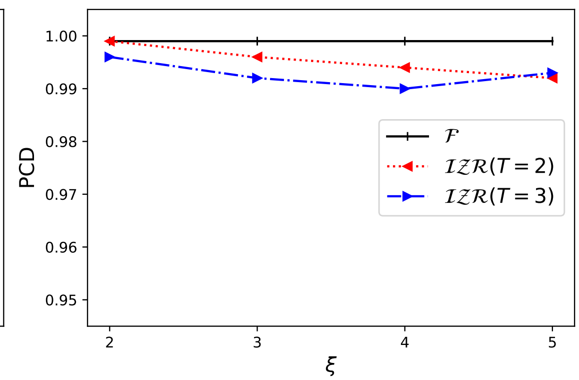
Remark 2. Similar to $\mathcal{F}$, the choice of $c = 1$ is recommended for $\mathcal{IZR}$, resulting in $\eta = \frac{1}{2}\left[(2\beta)^{-2/(n_0-1)} - 1\right]$ where β is given in equation (2). Large T may reduce the efficiency of $\mathcal{IZR}$ because of the value of η. We recommend to use $T \in \{2, 3\}$. It should be noted that $\mathcal{IZR}$ is identical to $\mathcal{F}$ if $T = 1$.

Table 1: Initial sets of $\mathcal{E}_{\mathcal{U}}$ and $\mathcal{E}_{\mathcal{D}}$ for $T \in \{2,3\}$ and $\xi \in \{2,3,4,5\}$.

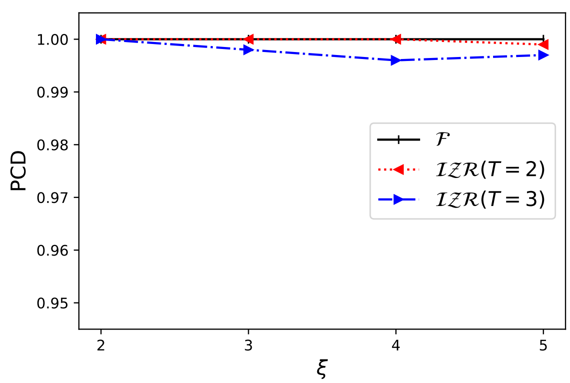
	$\xi = 2$	$\xi = 3$	$\xi = 4$	$\xi = 5$
$T = 2$	$\{0.04, 0.02\}$	$\{0.06, 0.02\}$	$\{0.08, 0.02\}$	$\{0.10, 0.02\}$
$T = 3$	$\{0.08, 0.04, 0.02\}$	$\{0.18, 0.06, 0.02\}$	$\{0.32, 0.08, 0.02\}$	$\{0.50, 0.10, 0.02\}$



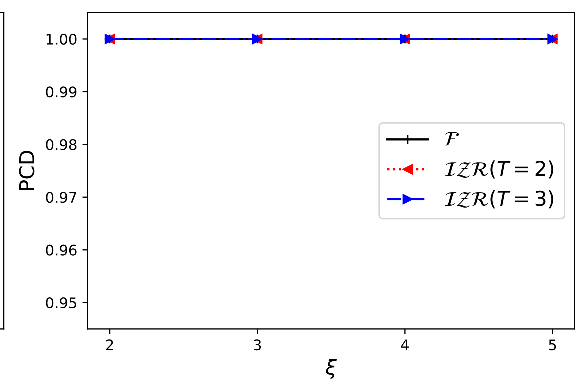
(a) $y = 0.02$



(b) $y = 0.05$



(c) $y = 0.1$



(d) $y = 0.5$

Figure 3: The PCD values of $\mathcal{F}$ and $\mathcal{IZR}$.

4 NUMERICAL EXPERIMENTS

In this section, we provide experimental results to compare $\mathcal{IZR}$ with $\mathcal{F}$. We consider one system (i.e., $k = 1$) and thus drop the subscript i. The mean of the system varies $y \in \{0.02, 0.03, 0.05, 0.1, 0.5\}$ and the variance is fixed to $\sigma^2 = 1$. The threshold is set to $q = 0$ and thus the constraint is $y \leq 0$. The tolerance level for $\mathcal{F}$ is $\epsilon = 0.02$ and thus we set the final tolerance level for $\mathcal{IZR}$ as $\epsilon^{(T)} = \epsilon = 0.02$ as well. For $\mathcal{IZR}$, we test $T \in \{2,3\}$ and $\xi \in \{2,3,4,5\}$. Therefore, there are a total of eight possible combinations of T and ξ, and the different sets $\mathcal{E}_{\mathcal{U}}$ and $\mathcal{E}_{\mathcal{D}}$ of relaxed tolerance levels for each combination are given in Table 1. We set $\alpha = 0.05$ and make 10,000 macro replications to report the estimated probability of CD (PCD) and the average total number of observations (OBS).

Figure 3 shows the values of PCD for procedures $\mathcal{F}$ and $\mathcal{IZR}$ with different T and ξ for various values of y. Both $\mathcal{F}$ and $\mathcal{IZR}$ result in PCD greater than or equal to the nominal level $1 - \alpha = 0.95$ for all settings.

Table 2 presents the values of OBS and the percentages (in the parenthesis) of time that $\mathcal{IZR}$ stopped with $(\epsilon^{(1)}, \epsilon^{(2)}, \ldots, \epsilon^{(T)})$. For example, when $T = 2$, $y = 0.02$, and $\xi = 2$, $(52.4\%, 47.6\%)$ means that in 52.4 and 47.6 percentages of macro replications the procedure $\mathcal{IZR}$ terminated with $\epsilon^{(1)} = 0.04$ and

Table 2: The OBS values of $\mathcal{F}$ and $\mathcal{IZR}$ and the percentages of times that $\mathcal{IZR}$ terminates with each $\epsilon^{(\tau)}$.

T	y	$\mathcal{F}$	$\mathcal{IZR}$			
			$\xi=2$	$\xi=3$	$\xi=4$	$\xi=5$
$T=2$	0.02	4129.98	4574.07	4886.70	5108.77	5227.03
			(52.4%, 47.6%)	(25.2%, 74.8%)	(17.0%, 83.0%)	(12.3%, 87.7%)
	0.05	2173.72	**1797.65**	**1855.28**	2174.93	2393.29
			(93.9%, 6.1%)	(64.2%, 35.8%)	(39.3%, 60.7%)	(27.0%, 73.0%)
	0.1	1184.28	**888.16**	**674.07**	**669.40**	**824.56**
			(99.9%, 0.1%)	(97.9%, 2.1%)	(86.0%, 14.0%)	(67.4%, 32.6%)
	0.5	258.91	**179.21**	**122.04**	**94.44**	**77.48**
			(100%, 0%)	(100%, 0%)	(100%, 0%)	(100%, 0%)
$T=3$	0.02	4129.98	5130.24	5725.80	6038.10	6268.79
			(13.8%, 41.5%, 44.7%)	(5.3%, 21.4%, 73.4%)	(3.4%, 12.9%, 83.7%)	(1.4%, 9.6%, 89.0%)
	0.05	2173.72	**1791.77**	**2068.85**	2498.42	2840.02
			(38.1%, 57.8%, 4.1%)	(8.6%, 58.8%, 32.7%)	(3.9%, 36.0%, 60.1%)	(1.7%, 23.7%, 74.6%)
	0.1	1184.28	**672.76**	**699.27**	**733.43**	**923.37**
			(88.5%, 11.5%, 0%)	(21.5%, 77.1%, 1.4%)	(7.2%, 81.8%, 11.0%)	(2.7%, 65.8%, 31.6%)
	0.5	258.91	**109.20**	**55.76**	**43.38**	**58.81**
			(100%, 0%, 0%)	(100%, 0%, 0%)	(93.2%, 6.8%, 0%)	(54.6%, 45.5%, 0%)

$\epsilon^{(2)} = 0.02$, respectively. In the table, we mark the OBS values in bold if $\mathcal{IZR}$ uses fewer observations than $\mathcal{F}$. When $y = 0.02$, $\mathcal{F}$ uses fewer observations than $\mathcal{IZR}$ under all settings. This is expected because (i) $\mathcal{IZR}$ has a high chance of terminating with a small tolerance level and (ii) the continuation region of $\mathcal{F}_{\mathcal{U}}$ and $\mathcal{F}_{\mathcal{D}}$ gets larger as the procedure proceeds with subroutines using a smaller tolerance level. It should be noted that $\mathcal{F}$ is designed for this scenario (i.e., $y = \epsilon = 0.02$) which unlikely happens in practice but $\mathcal{IZR}$ is designed to be less conservative than $\mathcal{F}$ in practical problems where at least some systems are not exactly the original tolerance level away from the threshold. Procedure $\mathcal{IZR}$ uses fewer observations than $\mathcal{F}$ for $\xi = 2, 3$ when $y = 0.05$ and for all ξ when $y = 0.1$ or 0.5. This happens because $\mathcal{IZR}$ terminates with a larger tolerance level than ϵ more often. For instance, when $y = 0.1$ or 0.5, $\mathcal{IZR}$ under the setting of $T = 2$ and $\xi = 2$ shows over 99.9% chance to terminate with $\epsilon^{(1)} = 0.04$. As discussed in Section 3, if either $y < q + \epsilon - 2\epsilon^{(\tau)}$ (similar to Case 1) or $y > q - \epsilon + 2\epsilon^{(\tau)}$ (similar to Case 5), then $\mathcal{IZR}$ has a high chance of terminating with $\epsilon^{(\tau)} > \epsilon$.

Unlike $\mathcal{F}$ with one fixed tolerance level, $\mathcal{IZR}$ is designed to detect feasibility early for a system whose mean is far from the threshold. Therefore, the best choice of T and ξ for $\mathcal{IZR}$ depends on the expected performance measure y, which is unknown. A thorough investigation on the selection of T or ξ is provided in Zhou et al. (2024).

5 CONCLUSION

We consider the problem of finding feasible systems with relaxed tolerance levels and propose a statistically valid procedure that returns a set that includes all desirable systems and excludes all unacceptable systems. Our experimental results show that the proposed procedure performs well in reducing the number of required observations especially for systems that are clearly desirable or clearly unacceptable.

ACKNOWLEDGMENTS

Sigrún Andradóttir was supported by NSF under grant CMMI-2127778 and Chuljin Park 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, S. and S.-H. Kim. 2010. “Fully sequential procedures for comparing constrained systems via simulation”. *Naval Research Logistics* 57(5):403–421.

Batur, D. and S.-H. Kim. 2010. “Finding feasible systems in the presence of constraints on multiple performance measures”. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 20(3):1–26.

Chen, C.-H., H. Lin, E. Yücesan, and S. E. Chick. 2000. “Simulation budget allocation for further enhancing the efficiency of ordinal optimization”. *Discrete Event Dynamic Systems* 10:251–270.

Chick, S. E. 2006. “Subjective probability and Bayesian methodology”. In *The Handbook of OR & MS: Simulation, Vol. 13*, edited by S. G. Henderson and B. L. Nelson, 225–257. Amsterdam: North-Holland Publishing.

Fabian, V. 1974. “Note on Anderson's sequential procedures with triangular boundary”. *The Annals of Statistics* 2(1):170–176.

Gao, S. and W. Chen. 2017. “Efficient feasibility determination with multiple performance measure constraints”. *IEEE Transactions on Automatic Control* 62(1):113–122.

Healey, M. C., S. Andradóttir, and S.-H. Kim. 2013. “Efficient comparison of constrained systems using dormancy”. *European Journal of Operations Research* 224:340–352.

Healey, M. C., S. Andradóttir, and S.-H. Kim. 2014. “Selection procedures for simulations with multiple constraints under independent and correlated sampling”. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 24:Article No. 14.

Hunter, S. R. and R. Pasupathy. 2013. “Optimal sampling laws for stochastically constrained simulation optimization on finite sets”. *INFORMS Journal on Computing* 25:527–542.

Kim, S.-H. and B. L. Nelson. 2001. “A fully sequential procedure for indifference-zone selection in simulation”. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 11:251–273.

Kim, S.-H. and B. L. Nelson. 2006. “Selecting the best system”. In *The Handbook of OR & MS: Simulation, Vol. 13*, edited by S. G. Henderson and B. L. Nelson, 501–534. Amsterdam: North-Holland Publishing.

Lee, L. H., N. A. Pujowidianto, L. W. Li, C.-H. Chen and C. M. Yap. 2012. “Approximate simulation budget allocation for selecting the best design in the presence of stochastic constraints”. *IEEE Transactions on Automatic Control* 57:2940–2945.

Lee, M. L., C. Park, and D. U. Park. 2018. “Self-adjusting the tolerance level in a fully sequential feasibility check procedure”. *European Journal of Operations Research* 271(2):733–745.

Pasupathy, R., S. R. Hunter, N. A. Pujowidianto, L. H. Lee and C. H. Chen. 2015. “Stochastically constrained ranking and selection via SCORE”. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 25(1).

Solow, D., R. Szechtman, and E. Yücesan. 2021. “Novel approaches to feasibility determination”. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 31(1).

Zhou, Y., S. Andradóttir, S.-H. Kim, and C. Park. 2024. “Indifference-zone relaxation procedure for ranking and selection”. Technical report, Department of Industrial Engineering, Hanyang University, Seoul, South Korea.

AUTHOR BIOGRAPHIES

CHULJIN PARK is an Associate Professor in the Department of Industrial Engineering at Hanyang University. He received his Ph.D. in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology in 2013. His research interests include simulation optimization, stochastic simulation, and applied statistics. His email address is parkcj@hanyang.ac.kr and his website is https://sites.google.com/site/stochsim/professor.

SIGRÚN ANDRADÓTTIR is a Professor in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. Prior to joining Georgia Tech, she was an Assistant Professor and later an Associate Professor in the Departments of Industrial Engineering, Mathematics, and Computer Sciences at the University of Wisconsin – Madison. She received her Ph.D. in Operations Research from Stanford University in 1990. Her research interests include simulation, applied probability, and stochastic optimization. She is a member of INFORMS and served as Editor of the Proceedings of the 1997 Winter Simulation Conference. She was the Simulation Area Editor of Operations Research Letters from 2002 to 2008, and has served as Associate Editor for various journals. Her email address is sa@gatech.edu and her website is http://isye.gatech.edu/faculty/sa.

SEONG-HEE KIM is a Professor in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. She received her Ph.D. in Industrial Engineering and Management Sciences from Northwestern University in 2001. Her email address is skim@isye.gatech.edu and her website is https://www2.isye.gatech.edu/~skim/.

YUWEI ZHOU is an Assistant Professor in the Kelley School of Business at Indiana University. Prior to joining Indiana University, she was a Postdoctoral Principal Researcher in the Booth School of Business at the University of Chicago from 2021 to 2024. She received her Ph.D. in Operations Research from the H. Milton Stewart School of Industrial and Systems

Engineering at Georgia Institute of Technology in 2021. Her research interest includes simulation optimization and data-driven decision making, with an emphasis on public sector applications. Her e-mail address is yuwzhou@iu.edu and her website is https://yuweizhou3.github.io.