Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

A NEW REWARD FUNCTION FOR BAYESIAN FEASIBILITY DETERMINATION

Junying He
Seong-Hee Kim

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology
755 Frest Drive, NW
Atlanta, GA 30332, USA

ABSTRACT

In Bayesian feasibility determination, a typical reward function is either the 0-1 or linear reward function. We propose a new type of reward function for Bayesian feasibility determination. Our proposed reward function emphasizes the importance of barely feasible/infeasible systems whose mean performance measures are close to the threshold. There are two main reasons why the barely feasible/infeasible systems are more important. First, the overall accuracy on solving a feasibility determination problem is heavily affected by those difficult systems. Second, if the decision maker wants to further find the best feasible system, it is likely that one of the barely feasible/infeasible systems is the best feasible. We derive a feasibility determination procedure with the new reward function in a Bayesian framework. Our experiments show that the Bayesian optimal procedure with the new reward function performs the best in making correct decisions on difficult systems when compared to existing procedures.

1. INTRODUCTION

We consider a problem of feasibility determination, where a decision maker wants to find a set of feasible systems among a finite number of simulated systems in the presence of a stochastic constraint. If a constraint is imposed on a performance measure whose value can only be estimated by stochastic simulation, we call it a stochastic constraint. In addition, we define a system as barely feasible/infeasible if its mean performance measure is close to a threshold value of the given constraint, and clearly feasible/infeasible if the mean performance measure is far from the threshold. Feasibility determination for a stochastic constraint occurs in many management and industrial applications. Some real-world examples are as follows:

1. The emergency department of a health-care unit has several shift arrangements for the staff. The decision maker wants to know which arrangements can keep patients' mean waiting time no more than 2 hours.
2. A manufacturing company has a few available production plans, and the decision maker wants to identify which plans can satisfy the production amount requirement of 10,000 units per month.
3. A facility management team is considering a number of cooling options on hand and wants to find out which options can keep the facility's temperature lower than or equal to 85°F.

Our problem is closely related to the study of constrained ranking-and-selection (R&S), where the goal is either to find a set of feasible systems or to find a feasible system with the largest or smallest mean performance measure in the presence of stochastic constraints. Three main approaches are usually used for the constrained R&S problems: the indifference-zone (IZ) approach, the optimal computing budget allocation (OCBA) approach, and the Bayesian approach. Among the IZ approach, Andradóttir and Kim

978-1-7281-3283-9/19/$31.00 ©2019 IEEE

(2010) consider a general form of a single stochastic constraint on a secondary performance measure and provide procedures for both feasibility determination and selection of the best feasible system. They introduce a parameter, namely error tolerance, that specifies how much a system's mean performance measure could be off from the threshold in the constraint but still acceptable to the decision maker. Batur and Kim (2010) focus on feasibility determination and provide IZ procedures for identifying a set of feasible systems in the presence of multiple stochastic constraints. Later, Healey et al. (2014) present IZ procedures that are more aggressive for feasibility determination and combine them with procedures to select the best feasible system. Among procedures with the OCBA approach, Lee et al. (2012) propose a budget allocation rule to maximize the probability of correct selection of the best feasible system under multiple stochastic constraints. Instead of selecting a single best feasible system, Gao and Chen (2017) develop an OCBA procedure that returns a set of feasible systems in the presence of multiple stochastic constraints. For the Bayesian approach, Xie and Frazier (2013) present sampling procedures that compare multiple systems to a known standard on a single performance measure. This is essentially feasibility determination under a single stochastic constraint because a system that has a mean performance measure better than the standard is equivalent to a feasible system. There are additional procedures that use the large deviation principle. While Hunter and Pasupathy (2013) and Pasupathy et al. (2014) focus on the selection of the best feasible system, Szechtman and Yücesan (2008) and Szechtman and Yücesan (2016) provide procedures to find a set of feasible systems. Gao and Chen (2017) also use the large deviation technique within the OCBA framework.

As mentioned above, procedures for constrained R&S involve either feasibility determination or feasibility determination and selection of the best. The main focus of this paper is on feasibility determination only and especially on the correct feasibility decision of barely feasible/infeasible systems. Even though the simulation is convenient for analyzing complex systems, it can be both time consuming and expensive. Therefore, it is crucial that the decision maker can allocate simulation effort wisely. Unlike IZ and OCBA procedures, Bayesian procedures incorporate the idea of value of information (VOI), which can be understood as the expected gain of reward from taking an additional simulated observation. The choice of a reward function is important as it determines when a Bayesian procedure stops taking observations and consequently, affects the overall performance of the procedure. More description about the Bayesian approach can be found in for example, Chick (2006), Chick and Gans (2009) and Chick and Frazier (2012).

In this paper, we propose a new form of a reward function that is more reasonable than the current popular linear and 0-1 reward functions for feasibility determination. Our motivation comes from the fact that, in many management or operation problems, barely feasible/infeasible systems are often more important in a sense that they are likely to be candidates for best feasible systems. For example, for the health-care unit example above, the decision maker may want to find the most cost-effective shift arrangement while keeping the patients' mean waiting time no more than 2 hours. Since adding more staff members would increase the cost of the shift arrangement while decreasing patients' mean waiting time, these two mean performance measures move in the opposite directions and the best arrangement is likely one of those that have the mean waiting time close to 2 hours. Thus, correct feasibility decisions on barely feasible/infeasible systems are more important than on clearly feasible/infeasible systems and it makes sense to assign a higher reward value to correct decision on these difficult systems. Unlike this intuition, the linear reward function gives a higher reward value to those located far from the threshold value. The 0-1 reward function assigns the same reward to all systems. Our proposed reward function puts a higher reward value to a system whose mean performance measure is closer to the threshold value of a constraint.

There are many functional shapes which assign higher reward values to systems whose means are close to the threshold such as a triangular, exponential, normal shape and so on. Under the assumption of normally distributed observations, we find that a reward function whose form is similar to the normal probability density function makes computation more tractable in deriving the expected reward. Thus, we propose a normal-shape reward function and derive a Bayesian optimal feasibility determination procedure with the reward function based on the method due to Xie and Frazier (2013).

The rest of the paper is organized as follows. Section 2 provides our notation and assumptions and defines the feasibility determination problem. In Section 3, we present our new reward function and a Bayesian optimal policy. Section 4 presents results from illustrative experiments to show the advantages of our proposed procedure compare to other existing procedures, followed by a conclusion in Section 5.

2. BACKGROUND

In this section, we introduce notation and necessary assumptions on simulation processes and define the feasibility determination problem.

2.1 Notation and Assumptions

For a simulation process, let $\mu_i = \mathrm{E}[Y_{ij}] \in \mathbb{R}$ and $\gamma_i = 1/\mathrm{Var}[Y_{ij}] \in (0,\infty)$, where Y_{ij} represents the jth simulation observation from system design i, for $i = 1,2,\ldots,k$ and $j = 1,2,\ldots$. We make the following assumptions on the simulation process:

Assumption 1 For any two systems $i,i' \in \{1,2,\ldots,k\}$ such that $i \neq i'$ and $j = 1,2,\ldots$ and $j' = 1,2,\ldots$, Y_{ij} and $Y_{i'j'}$ are independent.

Assumption 1 means that common random numbers are not used in the simulation process and observations from different systems are mutually independent.

Assumption 2 For each system $i = 1,2,\ldots k$, $Y_{ij} \overset{iid}{\sim} N(\mu_i, 1/\gamma_i), \quad j = 1,2,\ldots$.

Assumption 2 is plausible if $Y_{i1}, Y_{i2}, \ldots$ are within-replication averages across independent replications of system i, or if they are batch means from a large batch size within a single replication of a steady-state simulation after accounting for initialization effects. For more details, see Law and Kelton (2003).

The sampling precisions γ_i's are assumed to be known in this paper (and as in many other Bayesian R&S works). However, we consider μ_i's as the unknown mean performance measures of interest, for $i = 1,2,\ldots,k$. Using the Bayesian approach, we place a prior distribution on each μ_i. We suppose that these prior distributions come from the same distribution family ζ with parameter space Ω. To facilitate computation, we adopt independent conjugate priors. Specifically, we have the following assumption:

Assumption 3 For $i = 1,2,\ldots,k$, μ_i's are mutually independent and $\mu_i \sim N(\eta_i, 1/\lambda_i)$, where $\eta_i = \mathrm{E}[\mu_i] \in \mathbb{R}$ and $\lambda_i = 1/\mathrm{Var}[\mu_i] \in (0,\infty)$.

The assumption of known sampling precision γ_i is rarely true. The frequentist's approach, such as IZ approach, tends to deal with unknown γ_i's directly. However, the OCBA and Bayesian approaches often work on versions for known precisions. They then address the unknown variances by running a first-stage experiment that simulates a small number n_0 of replications and estimates γ_i by using its maximum likelihood estimator. In this paper, we consider known sampling precisions only.

2.2 Problem Formulation

In general, the goal of a feasibility determination problem is to find a set of systems among a finite number of simulated systems. We consider k available systems. Without loss of generality, we define that a system is feasible if and only if its mean performance measure of interest is less than or equal to the corresponding threshold. For simplicity, we consider situations where there is only one constraint with a threshold d. Therefore, a system i is feasible if and only if $\mu_i \leq d$. We define $\mathbb{F} = \{i : \mu_i \leq d, i \in \{1,2,\ldots,k\}\}$, which is the true set of feasible systems.

We formulate the feasibility determination as a dynamic program following Xie and Frazier (2013). The stage is indexed by $n = 0,1,2,\ldots$. At each stage n, we choose exactly one system $i_n \in \{1,2,\ldots,k\}$ to sample, and let $S_{n,i}$ be the parameters of the posterior distribution for μ_i for $i = 1,2,\ldots,k$. By convention, we denote $S_{0,i}$ as the parameters of prior distribution for μ_i. Since we choose conjugate priors to sampling distributions, the sampling process results in a sequence of posterior distributions, each of which resides in the same distribution family ζ parameterized by the same space Ω. Therefore, we have that $S_{n,i} = (\eta_{n,i}, \lambda_{n,i}) \in \Omega$, where $\Omega = \mathbb{R} \times (0,\infty)$, for all $n = 0,1,2,\dots$ and $i = 1,2,\dots,k$.

The reward function r is chosen by the decision maker. Specifically, the reward function r is defined as a two-piece function:

$$r(F;\boldsymbol{\mu},d) = \sum_{i\in F} r_0(\mu_i,d) + \sum_{i\notin F} r_1(\mu_i,d),$$

where r_0 and r_1 are known real-valued functions, F is any subset of $\{1,2,\dots,k\}$, and $\boldsymbol{\mu} = \{\mu_1,\dots,\mu_k\}$. At each stage n, the set $F_n \subset \{1,\dots,k\}$ is chosen to maximize the expected reward function, given the information of n observations. Specifically, for all $n \geq 0$,

$$\begin{aligned} F_n &= \operatorname*{arg\,max}_{F\subset\{1,2,\dots,k\}} \mathrm{E}_n\left[r(F;\boldsymbol{\mu},d)\right] \\ &= \operatorname*{arg\,max}_{F\subset\{1,2,\dots,k\}} \left\{ \sum_{i\in F} \mathrm{E}_n\left[r_0(\mu_i,d)\right] + \sum_{i\notin F} \mathrm{E}_n\left[r_1(\mu_i,d)\right] \right\}, \end{aligned}$$

where $\mathrm{E}_n[\cdot]$ denotes the conditional expectation given the information of observations at stage n. A policy π is composed of a decision rule for choosing the sequence of systems to be sampled (i.e., $(i_n)_{n\geq 1}$) and a termination rule for choosing a stopping stage τ so that no more observations are taken after stage τ. Eventually, the estimate of $\mathbb{F}$ returned by the procedure is F_τ. Although our problem formulation can adopt different unit costs for different systems, we assume a fixed unit cost c associated with simulating an observation for all systems. Our goal is to find a policy π that maximizes the expected total reward. That is, we want to solve the problem

$$\sup_{\pi} \mathrm{E}^{\pi}\left[r(F_\tau;\boldsymbol{\mu},d) - c\tau\right]. \tag{1}$$

where $\mathrm{E}^{\pi}[\cdot]$ denotes the unconditional expectation under policy π.

The decision maker needs to specify a reward function in order to find an optimal policy. Two common choices are 0-1 and linear reward functions:

- 0-1 reward function: $r_0(\mu_i,d) = I(\mu_i \leq d)$, $r_1(\mu_i,d) = I(\mu_i > d)$, where $I(\cdot)$ is the indicator function;
- linear reward function: $r_0(\mu_i,d) = d - \mu_i$, $r_1(\mu_i,d) = \mu_i - d$.

When a correct decision is made, the 0-1 reward function gives the same amount of reward to any system while the linear reward function gives a higher reward to a clearly feasible/infeasible system. Next we present a reward function that gives a higher reward to a barely feasible/infeasible system when a correct decision is made. Table 1 summarizes notation used throughout this paper.

3. NEW REWARD FUNCTION

In this section, we propose a new reward function, so called *the normal reward function* where the name comes from the fact that its functional form is similar to the normal probability density function. We provide a Bayesian optimal policy constructed with the new normal reward function using the framework of Xie and Frazier (2013).

3.1 Normal Reward Function

As discussed in section 1, barely feasible/infeasible systems are often more important in the feasibility determination problem. However, neither 0-1 nor linear can capture such importance of barely feasible/infeasible systems. Therefore, we propose the normal reward function as follows:

$$r(F;\boldsymbol{\mu},d) = \sum_{i\in F} r_0(\mu_i,d) + \sum_{i\notin F} r_1(\mu_i,d), \tag{2}$$

Table 1: Summary of notation used in the paper.

notation	meaning
k	total number of available systems
d	control requirement for each system i, $i = 1,\dots,k$
c	cost per simulation for each system i, $i = 1,\dots,k$
n	stage counter, $n = 0,1,2,\dots$
μ_i	mean performance for system i, $i = 1,\dots,k$
$\boldsymbol{\mu}$	vector of means $(\mu_1,\mu_2,\dots,\mu_k)$
γ_i	sampling precision for system i, $i = 1,\dots,k$
Ω	parameter space of the prior and posterior distributions
ζ	distribution family of the prior and posterior distributions with parameter space Ω
$\eta_{0,i}$	mean of prior distribution on μ_i, $i = 1,\dots,k$
$\eta_{n,i}$	mean of posterior distribution on μ_i at stage $n = 1,2,\dots$, $i = 1,\dots,k$
$\lambda_{0,i}$	precision of prior distribution on μ_i, $i = 1,\dots,k$
$\lambda_{n,i}$	precision of posterior distribution on μ_i at stage $n = 1,2,\dots$, $i = 1,\dots,k$
$S_{n,i}$	state of parameters of distribution on μ_i; $S_{n,i} = (\eta_{n,i},\lambda_{n,i})$
$\boldsymbol{S}_n$	vector of states $(S_{n,1},S_{n,2},\dots,S_{n,k})$
π	policy that governs the rules of sampling and termination
τ	stopping stage determined by the policy
$\mathbb{F}$	true set of feasible systems, $\mathbb{F} \subset \{1,2,\dots,k\}$
F_n	estimate of $\mathbb{F}$ at stage $n = 0,1,2,\dots$
F_τ	final estimate of $\mathbb{F}$ returned by the procedure

where r_0 and r_1 are

$$r_0(\mu_i,d) = \begin{cases} a\cdot\exp\left\{-\frac{1}{2}(d-\mu_i)^2\cdot b\right\}, & \text{if } \mu_i \le d; \\ 0, & \text{otherwise;} \end{cases}$$

$$r_1(\mu_i,d) = \begin{cases} 0, & \text{if } \mu_i \le d; \\ a\cdot\exp\left\{-\frac{1}{2}(d-\mu_i)^2\cdot b\right\}, & \text{otherwise.} \end{cases}$$

For each system, the normal reward function assigns reward values that follow the shape of a half-normal distribution, with its maximum at the threshold. As a result, barely feasible/infeasible systems tend to have larger rewards than clearly feasible/infeasible ones. There are two parameters the decision maker needs to choose before implementation. Generally speaking, the parameter a determines the maximum magnitude of the reward, and b determines the spread-out of the reward. Section 3.5 explains how to choose these parameters.

3.2 Conditions on Reward Functions

The framework for deriving a Bayesian optimal policy due to Xie and Frazier (2013) requires a reward function to satisfy some conditions. To state these conditions, we need some additional notation. For any generic $s \in \Omega$, define

$$
\begin{aligned}
h_{0i}(s) &= \mathrm{E}\left[r_0(\mu_i,d)|\mu_i \sim \zeta(s)\right]; \\
h_{1i}(s) &= \mathrm{E}\left[r_1(\mu_i,d)|\mu_i \sim \zeta(s)\right]; \\
h_i(s) &= \max\{h_{0i}(s), h_{1i}(s)\}; \\
R_i(s) &= \mathrm{E}[h_i(S_{1,i})|S_{0,i} = s, i_1 = i] - h_i(s) - c; \text{ and} \\
V_i(s) &= \sup_{\tau_i} \mathrm{E}^{\tau_i}\left[\sum_{n=1}^{\tau_i} R_i(S_{n-1,i}) \middle| S_{0,i} = s, i_1 = \ldots = i_{\tau_i} = i\right]. \qquad (3)
\end{aligned}
$$

Note that we use τ_i to represent the policy in which only system i can be sampled and it is sampled for τ_i times.

The function $h_i(s)$ can be interpreted as the expected reward to be obtained with respect to a distribution of the unknown mean performance measure, and $R_i(s)$ as the expected increment of expected reward given one more sample.

The required conditions from Xie and Frazier (2013) are then as follows:

Condition 1 For each system i, there exists a deterministic nonnegative function $H_i(s)$ on Ω such that for any $s \in \Omega$,

$$\mathrm{E}\left[h_i(S_{n,i})|S_{0,i} = s, i_1 = i_2 = \ldots = i_n = i\right] - h_i(s) \leq H_i(s).$$

Condition 2 For each system i, there exists a deterministic nonnegative function $\tilde{H}_i(s)$ on Ω such that for any $s \in \Omega$,

$$\mathrm{E}\left[h_i(S_{1,i})|S_{0,i} = s, i_1 = i\right] - h_i(s) \leq \tilde{H}_i(s) \quad \text{and} \quad \lim_{n\to\infty}\left[\sup_{s \in PS(i;n)} \tilde{H}_i(s)\right] = 0,$$

where $PS(i;n) := \{s \in \Omega : \exists s' \in \Omega \text{ s.t } \Pr\{S_{n,i} = s|S_{0,i} = s', i_1 = i_2 = \ldots = i_n = i\} > 0\}$.

Condition 3 For any system i and precision λ, there exists an interval $\left[\overline{\eta_i}(\lambda), \underline{\eta_i}(\lambda)\right]$ such that $\eta \notin \left[\overline{\eta_i}(\lambda), \underline{\eta_i}(\lambda)\right]$ implies $V_i(\eta, \lambda) = 0$.

Proofs that the normal reward function satisfies the above conditions are in He (2019).

3.3 Bayesian Optimal Sampling Policy

Problem (1) can be solved using dynamic programming techniques. Let Ω^k be the state space of $\boldsymbol{S}_n$ for all $n \geq 0$. For each $\boldsymbol{s} = (s_1, \ldots, s_k) \in \Omega^k$, we define $V(\boldsymbol{s})$ as the optimal expected total reward attainable when the initial state is $\boldsymbol{s}$. Specifically,

$$V(\boldsymbol{s}) = \sup_{\pi} \mathrm{E}^{\pi}\left[r(F_\tau; \boldsymbol{\mu}, d) - \tau c|\boldsymbol{S}_0 = \boldsymbol{s}\right]. \qquad (4)$$

Xie and Frazier (2013) prove that (4) is equivalent to

$$V(\boldsymbol{s}) = \sup_{\pi} \mathrm{E}^{\pi}\left[\sum_{n=1}^{\tau} R_{i_n}(S_{n-1,i_n}) \middle| \boldsymbol{S}_0 = \boldsymbol{s}\right]. \qquad (5)$$

Instead of solving (5) directly, consider the subproblem where only system i can be sampled. Specifically, the subproblem is (3).

Results from the dynamic programming literature (see, for example, Dynkin and Yushkevich (1979)) show that $V_i(s)$ satisfies Bellman's recursion:

$$V_i(s) = \max\left[0, L_i(s, V_i)\right], \tag{6}$$

where $L_i(s, V_i) = R_i(s) + \mathrm{E}\left[V_i(S_{1,i}) | S_{0,i} = s, i_1 = i\right]$.

Problem (6) is a standard optimal stopping problem that can be solved by specifying the so-called continuation set $\mathbb{C}_i$ (see, for example, Bertsekas (2007)). That is, we need to find $\mathbb{C}_i = \{s \in \Omega : V_i(s) > 0\}$. Then, an optimal solution to (6) is the stopping time τ_i^* given by $\tau_i^* = \inf\{n \geq 0, S_{n,i} \notin \mathbb{C}_i\}$. In general, τ_i^* can go to ∞. However, under Condition 2 we can provide a deterministic upper bound on τ_i^*, denoted as N_i, using the result from Xie and Frazier (2013):

$$N_i = \min\left\{n : \left[\sup_{s \in PS(i;n')} \tilde{H}_i(s)\right] \leq c, \forall n' \geq n\right\}. \tag{7}$$

We can now go back to the original problem (5). Let π^* be the Bayesian optimal policy with stopping stage τ^* such that $\mathrm{E}^{\pi^*}\left[r(F_{\tau^*}; \boldsymbol{\mu}, d) - \tau^* c | \boldsymbol{S}_0 = \boldsymbol{s}\right] = V(\boldsymbol{s})$. Given that all systems are mutually independent, it is straightforward that $V(\boldsymbol{s}) = \sum_{i=1}^k V_i(s_i)$. Since the value of information from each stage only depends on the system being sampled and the states of other systems remain unchanged, the order of the sequence of sampling decisions does not affect the total value of information. In fact, Xie and Frazier (2013) prove that the policy π^* with sampling decisions $(i_1^*, i_2^*, \ldots, i_n^*)$ and stopping stage τ^* is any policy that satisfies:

$$\begin{aligned} i_{n+1}^* &\in \{i, S_{n,i} \in \mathbb{C}_i\}, \forall i \geq 0; \\ \tau^* &= \inf\{n \geq 0 : S_{n,i} \notin \mathbb{C}_i, \forall i\}. \end{aligned}$$

Therefore, we can solve each subproblem (6) separately, and the final optimal policy π^* is the one that samples system i for τ_i^* stages sequentially for $i = 1, \ldots, k$, and has a stopping stage $\tau^* = \sum_{i=1}^k \tau_i^*$.

Chick and Gans (2009) and Chick and Frazier (2012) use a similar strategy of sampling only from system i before stopping. However, they consider the problem of selection of the best rather than feasibility decision, and their proposed procedures are not Bayesian optimal while the procedure presented here is Bayesian optimal. The Bayesian optimal feasibility determination procedure $\mathcal{BFD}$ is then stated in Algorithm 1.

Algorithm 1 Procedure $\mathcal{BFD}$

```
Setup: Let $F = \emptyset$. Specify number of systems $k$, threshold $d$ and unit cost $c$. Start with system $i = 1$.
Initialization: Specify prior distribution $N(\eta_{0,i}, 1/\lambda_{0,i})$ for the mean performance $\mu_i$ and sampling precision $\gamma_i$. Compute continuation region $\mathbb{C}_i$. Set $n_i = 0$.
Update: Let $n_i = n_i + 1$. Simulate one observation $y_i$ from system $i$. Compute
    $\eta_{n_i,i} = \frac{\lambda_{n_i-1,i}\eta_{n_i-1,i} + \gamma_i y_i}{\lambda_{n_i-1,i} + \gamma_i}$,
    $\lambda_{n_i,i} = \lambda_{n_i-1,i} + \gamma_i$.
Stopping Rule: If $(\eta_{n_i,i}, \lambda_{n_i,i}) \notin \mathbb{C}_i$, then stop sampling from system $i$ and go to Feasibility Check. Otherwise, go back to Update.
Feasibility Check: If $\eta_{n_i,i} \leq d$, then add $i$ in $F$.
Termination Rule: Set $i = i + 1$. If $i \leq k$, go to Initialization. Otherwise, return $F$ as the set of feasible systems.
```

Procedure $\mathcal{BFD}$ works for any reward function that satisfies conditions given in Section 3.2, but a different reward function results in a different continuation set $\mathbb{C}_i$. In the next subsection, we explain how to find $\mathbb{C}_i$.

3.4 Continuation Set

To find a continuation set, one has to solve (6). Under Assumptions 1 and 2, it can be shown that

$$\mathrm{E}\left[V_i(S_{1,i})|S_{0,i}=s, i_1=i\right] = \mathrm{E}\left[V_i(\eta+\tilde{\sigma}_i(\lambda)\cdot Z, \lambda+\gamma_i)\right],$$

where $s=(\eta,\lambda)\in\mathbb{R}\times(0,\infty)$, $\tilde{\sigma}_i(\lambda)=\sqrt{\frac{\gamma_i}{\lambda(\lambda+\gamma_i)}}$ and Z is a standard normal random variable. Therefore, (6) becomes

$$V_i(\eta,\lambda)=\max\{0, L_i(\eta,\lambda,V_i)\}, \tag{8}$$

where $L_i(\eta,\lambda,V_i)=R_i(\eta,\lambda)+\mathrm{E}\left[V_i(\eta+\tilde{\sigma}_i(\lambda)Z,\lambda+\gamma_i)\right]$.

To calculate $V_i(\eta,\lambda)$ for all possible $(\eta,\lambda)\in\Omega$, the main idea is to use a backward algorithm:

1. First, we start by considering a large number of stages N_i such that $V_i(\eta,\lambda_{0,i}+n\gamma_i)=0$, for all $n>N_i$ and all $\eta\in\mathbb{R}$. The number N_i can be found by using (7). For simplicity, we set $N_i=N=1000$ for our numerical experiments.
2. Starting from $\lambda=\lambda_{0,i}+N\gamma_i$, we compute $\left[\overline{\eta_i}(\lambda),\underline{\eta_i}(\lambda)\right]$ as the boundary of η such that $V_i(\eta,\lambda)=0$, if $\eta\notin\left[\overline{\eta_i}(\lambda),\underline{\eta_i}(\lambda)\right]$. Under Condition 3, we know such $\left[\overline{\eta_i}(\lambda),\underline{\eta_i}(\lambda)\right]$ exists.
3. Then, we discretize the range $\left[\overline{\eta_i}(\lambda),\underline{\eta_i}(\lambda)\right]$ into points $\left\{\eta_i(\lambda)^j\right\}$ with an interval of δ (in our experiments, we set $\delta=0.01$).
4. Using (8) and an approximation that

$$\begin{aligned}&\mathrm{E}\left[V_i\left(\eta+\tilde{\sigma}_i(\lambda)Z,\lambda+\gamma_i\right)\right]\\ &\approx\sum_j V_i\left(\eta_i^j(\lambda+\gamma_i),\lambda+\gamma_i\right)\cdot\left[\Phi\left(\frac{\eta_i^j(\lambda+\gamma_i)+\delta/2-\eta}{\tilde{\sigma}(\lambda)}\right)-\Phi\left(\frac{\eta_i^j(\lambda+\gamma_i)-\delta/2-\eta}{\tilde{\sigma}(\lambda)}\right)\right],\end{aligned}$$

   where $\Phi(\cdot)$ is the cumulative density function of a standard normal random variable, each $V_i(\eta_i^j(\lambda),\lambda)$ can be computed recursively for $\lambda\in\{\lambda_{0,i}+n\gamma_i : 0\le n\le N\}$.
5. Finally, for any arbitrary $(\eta,\lambda)\in\mathbb{R}\times\{\lambda_{0,i}+n\gamma_i : 0\le n\le N\}$, we set

$$V_i(\eta,\lambda)=\begin{cases}0, & \text{if } \eta\notin\left[\overline{\eta_i}(\lambda),\underline{\eta_i}(\lambda)\right];\\ V_i\left(\eta_i^{j^*}(\lambda),\lambda\right), & \text{otherwise}\end{cases}$$

   where $j^*=\arg\min\left\{|\eta-\eta_i^j(\lambda)|\right\}$.
6. As a result, we find $\mathbb{C}_i=\left\{[\overline{\eta_i}(\lambda),\underline{\eta_i}(\lambda)] : \lambda=\lambda_{0,i}+n\gamma_i, 0\le n\le N\right\}$.

The remaining work to complete the policy is to specify $h_i(\eta,\lambda)$ and $R_i(\eta,\lambda)$ functions, for each $i=1,2,\ldots,k$. We directly state the results here, and details of calculation can be found in He (2019).

Theorem 1 When the normal reward function in (2) is used for any $\eta\in\mathbb{R}$ and any $\lambda\in(0,\infty)$, the function $h_i(\eta,\lambda)$ is

$$h_i(\eta,\lambda)=\max\{h_{0i}(\eta,\lambda),h_{1i}(\eta,\lambda)\},$$

where

$$h_{0i}(\eta,\lambda)=\frac{a\sqrt{2\pi}}{\sqrt{b}}G(d,\eta,b,\lambda)\cdot\Phi\left((d-\frac{db+\eta\lambda}{b+\lambda})\sqrt{b+\lambda}\right),$$

$$h_{1i}(\eta,\lambda)=\frac{a\sqrt{2\pi}}{\sqrt{b}}G(d,\eta,b,\lambda)\cdot\left[1-\Phi\left((d-\frac{db+\eta\lambda}{b+\lambda})\sqrt{b+\lambda}\right)\right], \text{ and}$$

$$G(\eta_f,\eta_g,\lambda_f,\lambda_g) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{\lambda_f \lambda_g}{\lambda_f + \lambda_g}} \exp\left\{ -\frac{1}{2}(\eta_f - \eta_g)^2 \frac{\lambda_f \lambda_g}{\lambda_f + \lambda_g} \right\}.$$

Also, $R_i(\eta,\lambda)$ is defined as

$$R_i(\eta,\lambda) = \mathrm{E}\left[h_i(\eta + \tilde{\sigma}_i(\lambda) \cdot Z, \lambda + \gamma_i)\right] - h_i(\eta,\lambda) - c,$$

where

$$\begin{aligned} &\mathrm{E}\left[h_i(\eta + \tilde{\sigma}_i(\lambda) \cdot Z, \lambda + \gamma_i)\right] = \\ &\quad \frac{a\sqrt{2\pi}}{\sqrt{b}} \sqrt{\frac{\lambda(\lambda+\gamma_i)}{\gamma_i}} G\left(\frac{d-\eta}{\tilde{\sigma}_i(\lambda)}, 0, \frac{b\gamma_i}{\lambda(b+\lambda+\gamma_i)}, 1 \right) \cdot \Pr\left\{ Z_1 \le \frac{d-\eta}{\tilde{\sigma}_i(\lambda)}, Z_2 \le 0 \right\} \\ &\quad + \frac{a\sqrt{2\pi}}{\sqrt{b}} \sqrt{\frac{\lambda(\lambda+\gamma_i)}{\gamma_i}} G\left(\frac{d-\eta}{\tilde{\sigma}_i(\lambda)}, 0, \frac{b\gamma_i}{\lambda(b+\lambda+\gamma_i)}, 1 \right) \cdot \Pr\left\{ Z_1 \ge \frac{d-\eta}{\tilde{\sigma}_i(\lambda)}, Z_2 \ge 0 \right\}. \end{aligned}$$

The vector of variables $\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$ follows a multivariate normal distribution $MVN(\tilde{\boldsymbol{m}}, \tilde{\boldsymbol{\Sigma}})$ with mean $\tilde{\boldsymbol{m}}$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}$, where

$$\tilde{\boldsymbol{m}} = \begin{bmatrix} m_1 \\ -A + Bm_1 \end{bmatrix},$$

$$\tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} 1/p_1 & B/p_1 \\ B/p_1 & 1 + B^2/p_1 \end{bmatrix},$$

with $m_1 = \frac{\frac{d-\eta}{\tilde{\sigma}_i(\lambda)} b\gamma_i}{b\gamma_i + \lambda(b+\lambda+\gamma_i)}$, $p_1 = \frac{b\gamma_i}{\lambda(b+\lambda+\gamma_i)} + 1$, $A = -\frac{(d-\eta)(\lambda+\gamma_i)}{\sqrt{b+\lambda+\gamma_i}}$ and $B = \frac{\tilde{\sigma}_i(\lambda)(\lambda+\gamma_i)}{\sqrt{b+\lambda+\gamma_i}}$.

□

By the above theorem, $\mathrm{E}\left[h_i(\eta + \tilde{\sigma}_i(\lambda) \cdot Z, \lambda + \gamma_i)\right]$ can be calculated using the cumulative density function of a multivariate normal distribution.

3.5 Choices of Parameters

To finish up the procedure, we discuss how to specify the parameters a and b in the normal reward function. We define $\upsilon > 0$ as the importance parameter for a constraint. This parameter defines an important range where a correct feasibility decision is more desirable and the decision maker is willing to take more observations if needed. For systems outside the important range, making a correct decision is easier and a procedure still tries to make a correct feasibility decision due to positive rewards. We want to point out that the importance parameter and the error tolerance in the IZ approach have different interpretations. The error tolerance in the IZ approach defines a range around d where the decision maker does not care about the correct decision, which has the opposite meaning of our importance parameter.

The main motivation of the normal reward function is to increase the chance of a correct difficult systems whose means of performance are in $[d-\upsilon, d+\upsilon]$ and this requires assigning more observations to those systems than the 0-1 reward function. For a given unit cost c for simulating one observation, $\mathcal{BFD}$ gives an additional observation only when the expected additional reward is higher than the unit cost c. Thus, to make difficult systems receive more observations, we need to ensure that the range $[d-\upsilon, d+\upsilon]$ receives a higher reward than 1.

We set the reflection points of the normal reward function (points where the 2nd derivative of the reward function are equal to 0) to be $(d-\upsilon, 1)$ and $(d+\upsilon, 1)$. By this way, the normal reward function would give a reward greater than 1 and aggressively increase the reward on systems within $(d-\upsilon, d+\upsilon)$, while decrease the reward on those outside the range, which is consistent with the interpretation of the important range. Consequently, we choose to set $a = \sqrt{\exp(1)}$ and $b = \upsilon^{-2}$.

4. NUMERICAL EXPERIMENTS

In this section, we demonstrate the advantages of using the Bayesian approach with the normal reward function using simple examples.

We define CD_i as the event of making a correct feasibility decision for system i, and $\mathrm{CD} \equiv \cap_{i=1}^{k} \mathrm{CD}_i$ as the event of making correct decisions on feasibilities of all available systems. Furthermore, we define $\mathrm{PCD}_i \equiv \Pr\{\mathrm{CD}_i\}$ and $\mathrm{PCD} \equiv \Pr\{\mathrm{CD}\}$. In our experiments, we estimate each PCD_i and PCD empirically based on 10,000 replications. We also record the average number of observations spent on each system per replication (OBS_i), and find the average total number of observations per replication (OBS). Due to a limited space, we only report PCD and OBS but PCD_i and OBS_i are reported in He (2019).

The first comparison is among the three different reward functions to show that the normal reward function is more ideal for feasibility determination, especially on the barely feasible/infeasible systems. We implement the $\mathcal{BFD}$ procedure with each of the reward functions, and each version is denoted as $\mathcal{BFD}$ normal, $\mathcal{BFD}$ 0-1 and $\mathcal{BFD}$ linear, respectively.

The second comparison is among procedures from three different approaches for feasibility determination, namely Bayesian, IZ and OCBA approaches. To compare against the performance of our proposed Bayesian procedure, $\mathcal{BFD}$ normal, we choose the following procedures that best suit the problem of feasibility determination:

- The $\mathcal{BK}$ procedure (Batur and Kim 2010) which falls in the category of IZ procedures.
- The $\mathcal{GC}$ procedure (Gao and Chen 2017) which uses the OCBA framework with the large deviation principle.

More details about the $\mathcal{BK}$ and $\mathcal{GC}$ procedures can be found in He (2019).

4.1 Experimental Settings

We consider $k = 50$ systems. Without loss of generality, we set the threshold $d = 0$ for all systems. For simplicity, the unit cost of simulation is $c = 0.001$ for all systems. For each μ_i, we place a conjugate prior distribution $\mu_i \sim N(\eta_{0,i}, 1/\lambda_{0,i})$ with $\eta_{0,i} = 0$ and $\lambda_{0,i} = 0.01$. For the normal reward function, we set the importance parameter $\upsilon = 1$. The true mean performances μ_i of systems are $\mu_i = -2.5 + 0.1 \cdot (i-1)$ for $i = 1, 2, \ldots, 25$ and $\mu_i = 0.1 \cdot (i-25)$ for $i = 26, 27, \ldots, 50$. The set $\mathbb{F} = \{1, 2, \ldots, 25\}$ is the true set of feasible systems.

We consider three configurations for the systems' true precisions γ_i, $i = 1, 2, \ldots, 50$: constant precisions (CP), decreasing precisions (DP) and increasing precisions (IP). In CP, we set $\gamma_i = 1$ for all systems. As the true mean performances move away from the standard, the systems' true precisions decrease in DP, while they increase in IP. In particular, $\gamma_i = 1/[1 + (|i - 25.5| - 0.5) \cdot 0.1]^2$ for $i = 1, 2, \ldots, 50$ in DP; $\gamma_i = [1 + (|i - 25.5| - 0.5) \cdot 0.1]^2$ for $i = 1, 2, \ldots, 50$ in IP.

For the $\mathcal{BK}$ procedure, we use a simple grid search to explore different values for the confidence level $1-\alpha$ and error tolerance ϵ to find appropriate settings such that the procedure produces approximately the same average total number of observations as the $\mathcal{BFD}$ normal procedure. The reason for doing so is that we can compare the two procedures by comparing their PCD, while keeping the total cost roughly the same. Based on the grid search, the values of $(1-\alpha, \epsilon)$ for $\mathcal{BK}$ are $(0.90, 0.25)$, $(0.75, 0.45)$ and $(0.75, 0.15)$ for the CP, DP and IP configurations, respectively.

For the $\mathcal{GC}$ procedure, we set the total budget equal to the average total number of observations per replication of the $\mathcal{BFD}$ normal procedure in each configuration. In addition, we set the incremental budget at each stage $\Delta_0 = 5$.

4.2 Results

Table 2 shows that $\mathcal{BFD}$ with the normal reward function results in significantly higher PCD than the other two reward functions but spends more observations. By checking PCD$_i$'s and OBS$_i$'s (reported in He (2019)), we see that the improvement on PCD is due to more correct decisions on the barely feasible/infeasible systems (i.e., systems 16 to 35) and that the new procedure spends more observations only on the barely feasible/infeasible systems.

Table 3 shows PCD and OBS for the $\mathcal{BFD}$ normal, $\mathcal{BK}$ and $\mathcal{GC}$ procedures. The $\mathcal{BFD}$ normal procedure still performs the best among the three procedures in all three configurations in terms of PCD when spending a similar number of OBS. We see that the $\mathcal{BFD}$ normal procedure assigns simulation efforts more efficiently, in a sense that it spends less budget on the clearly feasible/infeasible systems, but more on the barely feasible/infeasible ones, compared to the $\mathcal{BK}$ and $\mathcal{GC}$ procedures. The performance of the $\mathcal{BK}$ procedure under the DP configuration is strange due to a large value of $\epsilon = 0.45$. Since the indifference zone parameter ϵ is large, the $\mathcal{BK}$ procedure tend to care less on many systems whose mean performance measures are close to d. In addition, the sampling precisions (sampling variances) of these systems are large (small) under the DP configuration, which causes the procedure to spend fewer observations on them. Consequently, we see that more observations are spent on systems far from d in the $\mathcal{BK}$ procedure under the DP configuration.

Table 2: Summary of PCD and average total number of observations (OBS) for $\mathcal{BFD}$ with the normal, 0-1 and linear reward functions under CP, DP and IP configurations.

	PCD			OBS		
	normal	0-1	linear	normal	0-1	linear
CP	0.942	0.876	0.477	1135	888	394
DP	0.914	0.829	0.429	1597	1267	765
IP	0.946	0.894	0.508	925	735	260

Table 3: Summary of PCD and average total number of observations (OBS) for $\mathcal{BFD}$ normal, $\mathcal{BK}$ and $\mathcal{GC}$ under CP, DP and IP configurations.

	PCD			OBS		
	$\mathcal{BFD}$ normal	$\mathcal{BK}$	$\mathcal{GC}$	$\mathcal{BFD}$ normal	$\mathcal{BK}$	$\mathcal{GC}$
CP	0.942	0.763	0.881	1135	1215	1135
DP	0.914	0.368	0.856	1597	1674	1597
IP	0.946	0.913	0.909	925	998	925

5. CONCLUSION

We introduce a new reward function, namely the normal reward function, that assigns a higher reward value on the barely feasible/infeasible systems than clearly feasible/infeasible ones. We demonstrate the advantages of the normal reward function using a Bayesian optimal feasibility determination procedure over popular 0-1 and linear reward functions. Then the Bayesian optimal feasibility determination procedure is compared with the existing IZ and OCBA procedures. From our experiments, we see that compared to the 0-1 and linear reward functions, the normal reward function does better in feasibility decisions on barely feasible/infeasible systems, while performing well on clearly feasible/infeasible systems.

REFERENCES

Andradóttir, S., and S. 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* 20(3):1–26.

Bertsekas, D. P. 2007. *Dynamic Programming and Optimal Control*. 3rd ed. Belmont, Massachusetts: Athena Scientific.

Chick, S. E. 2006. "Subjective Probability and Bayesian Methodology". In *Handbooks in Operations Research and Management Science*, edited by S.G. Henderson and B.L. Nelson, 225 – 257. Amsterdam: Elsevier.

Chick, S. E., and P. Frazier. 2012. "Sequential Sampling with Economics of Selection Procedures". *Management Science* 58(3):550–569.

Chick, S. E., and N. Gans. 2009. "Economic Analysis of Simulation Selection Problems". *Management Science* 55(3):421–437.

Dynkin, E. B., and A. A. Yushkevich. 1979. *Controlled Markov Processes*. New York: Springer.

Gao, S., and W. Chen. 2017. "Efficient Feasibility Determination With Multiple Performance Measure Constraints". *IEEE Transactions on Automatic Control* 62(1):113–122.

He, J. 2019. "Bayesian Approach for Feasibility Determination". Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

Healey, C. M., 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* 24(3):1–25.

Hunter, S. R., and R. Pasupathy. 2013. "Optimal Sampling Laws for Stochastically Constrained Simulation Optimization on Finite Sets". *INFORMS Journal on Computing* 25(3):527–542.

Law, A. M., and W. D. Kelton. 2000. *Simulation Modeling and Analysis*. 3rd ed. New York: McGraw-Hill, Inc.

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(11):2940–2945.

Pasupathy, R., S. R. Hunter, N. A. Pujowidianto, L. H. Lee, and C.-H. Chen. 2014. "Stochastically Constrained Ranking and Selection via SCORE". *ACM Transactions on Modeling and Computer Simulation* 25(1):1–26.

Szechtman, R., and E. Yücesan. 2008. "A New Perspective on Feasibility Determination". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mnch, O. Rose, T. Jefferson, and J. W. Fowler, 273–280. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Szechtman, R., and E. Yücesan. 2016. "A Bayesian Approach to Feasibility Determination". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 782–790. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Xie, J., and P. I. Frazier. 2013. "Sequential Bayes-Optimal Policies for Multiple Comparisons with a Known Standard". *Operations Research* 61(5):1174–1189.

AUTHOR BIOGRAPHIES

JUNYING HE is a PhD student in H. Milton Stewart School of Industrial and Systems Engineering at Georgia Insitute of Technology. His email address is junying.he@gatech.edu and his website is https://www.isye.gatech.edu/users/junying-he.

SEONG-HEE KIM is a Professor in H. Milton Stewart School of Industrial and Systems Engineering at Georgia Insitute 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/.