*Proceedings of the 2023 Winter Simulation Conference*
*C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.*

# STOCHASTIC CONSTRAINTS: HOW FEASIBLE IS FEASIBLE?

David J. Eckman

Industrial and Systems Engineering
Texas A&M University
TAMU 3131
College Station, TX 77843, USA

Shane G. Henderson

Operations Research and Information Engineering
Cornell University
Rhodes Hall
Ithaca, NY 14853, USA

Sara Shashaani

Industrial and Systems Engineering
North Carolina State University
915 Partners Way
Raleigh, NC 27695, USA

## ABSTRACT

Stochastic constraints, which constrain an expectation in the context of simulation optimization, can be hard to conceptualize and harder still to assess. As with a deterministic constraint, a solution is considered either feasible or infeasible with respect to a stochastic constraint. This perspective belies the subjective nature of stochastic constraints, which often arise when attempting to avoid alternative optimization formulations with multiple objectives or an aggregate objective with weights. Moreover, a solution's feasibility with respect to a stochastic constraint cannot, in general, be ascertained based on only a finite number of simulation replications. We introduce different means of estimating how "close" the expected performance of a given solution is to being feasible with respect to one or more stochastic constraints. We explore how these metrics and their bootstrapped error estimates can be incorporated into plots showing a solver's progress over time when solving a stochastically constrained problem.

## 1 INTRODUCTION

In the context of simulation optimization, *stochastic constraints* are inequality constraints involving the expected values of random variables. Typically, we cannot compute the expected values exactly, but we can estimate them from simulation replications. If we could compute the expected values exactly, then these constraints would really be deterministic constraints, since simulation is not needed to assess feasibility. Stochastic constraints arise quite naturally in practice, as we will see in a stylized call-center example where they limit the expected number of calls having long response times. But they introduce significant complexity relative to simulation-optimization problems with only deterministic constraints. Indeed, the feasible region of problems with stochastic constraints can be very complex. Moreover, the error in estimators of the expectations appearing in constraints makes it difficult to determine, or approximate, the feasible region.

In this paper, we focus on methods for quantifying the feasibility of a single solution in an optimization problem having one or more stochastic constraints. Our interest in this problem relates to the SimOpt testbed of simulation optimization problems (Eckman et al. 2021). SimOpt code automatically generates plots that shed light on the empirical performance of solvers operating on test problems, but only those with *deterministic* constraints. Our goal is to extend those plots to depict the feasibility, or lack thereof, of the solutions recommended during the course of a solver's search for an optimal solution.

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

The surveys Homem-de Mello and Bayraksan (2014, 2015) provide an excellent overview of, and many important references on, the role of constraints in stochastic optimization in general, not just in simulation, and on methods for tackling such problems. In particular, these surveys discuss the asymptotic (in sample size) convergence behavior of solutions to sample-average approximating problems and the use of variance-reduction techniques to reduce the error in sample-average approximations and, therefore, in the solutions obtained by solving those approximating problems. Given these surveys, we do not attempt to survey the literature on stochastic constraints, instead just highlighting some prominent areas, including multi-objective simulation optimization (Hunter et al. 2019; Cooper and Hunter 2020; Lee et al. 2010), ranking and selection with stochastic constraints (Healey et al. 2014; Hunter and Pasupathy 2013), and feasibility determination problems (Gao and Chen 2017; Shi et al. 2022; Solow et al. 2021). Stochastic constraints, as formulated here, are closely related to chance constraints, as initiated by Charnes et al. (1958) and Charnes and Cooper (1959). Miller and Wagner (1965) and Prékopa (1970) explored special cases where the resulting optimization problems are especially tractable. However, in the general case, chance-constrained problems possess little structure, making them challenging to solve exactly. Calafiore and Campi (2006) and Nemirovski and Shapiro (2007) develop approximation methods that are especially tractable. Dentcheva and Ruszczynski (2003) showed how to impose stochastic dominance constraints, which are in essence an infinite family of chance constraints, in a tractable manner. A related line of work centers on conditional value at risk (Rockafellar et al. 2000). Digabel and Wild (2015) provide a useful taxonomy for a very broad range of constraints.

The remainder of this paper is organized as follows. Section 2 explores a call-center example, in the context of which we discuss the advantages and disadvantages of alternative formulations to stochastic constraints. This example motivates our discussion of generic simulation-optimization problems with stochastic constraints in Section 3. Section 4 introduces some "feasibility scores" for quantifying the extent to which a solution is feasible or infeasible, and discusses the use of bootstrapping to provide confidence intervals for such measures. Section 5 explores a related metric based on hypothesis testing and Section 6 develops several additional metrics via a Bayesian perspective. Section 7 discusses how these metrics can be plotted over time when studying the performance of a simulation-optimization algorithm, and we conclude in Section 8.

## 2 CALL-CENTER EXAMPLE

The ideas we investigate are most concrete when explored through the lens of a practical example, namely call-center staffing. Call-center practice is complex (Koole and Li 2021), so we describe a greatly simplified model here. In this model, calls arrive according to a non-homogeneous Poisson process with rate function $(\lambda(t) : 0 \leq t \leq 16)$, where $t$ is measured in hours after the call center opens each day. We assume, unlike common practice, that all calls are of a single type. A single pool of agents answers the calls in first-in-first-out order. We assume no customer balking or abandonment, i.e., no callers hang up prior to receiving service. Time is divided into periods of equal length, with common choices being 15 minutes or 30 minutes; we consider 30-minute periods, so there are $J = 32$ periods over the 16-hour day. Agents can work one of a variety of shifts, which are assumed to consist of 4 hours of work, a 30-minute lunch break and 4 more hours of work. In reality, shifts can be of varying lengths with varying break times.

We wish to select the number of agents, $x_k$, who will work Shift $k$, $k = 1, 2, \ldots K$, so as to minimize staffing cost plus any overtime costs while maximizing the customer service experience. Staffing costs consist of a fixed amount $c_k$ per agent for Shift $k$ that covers wages, benefit costs and the like. Overtime arises when a call-center agent coming off shift needs to complete a call in progress, and also when agents must clear the queue of calls at the end of the day. If we let $O(x)$ denote the (random) cost associated with any overtime as a function of the staffing vector $x$, the total expected cost of staffing is $\sum_{k=1}^{K} c_k x_k + \mathbb{E}O(x)$.

We require a satisfactory customer experience, in that calls should be answered in a timely fashion in each period of the day. We measure the quality of service provided per period by the long-run (over many days of operations) fraction of customers that arrive in the given period whose time spent waiting prior

to initiating service exceeds some acceptable threshold, e.g., 20 seconds. For a single day's operation, let $L_j(x)$ be the number of so-called *late* calls and $N_j$ be the total number of calls arriving during period $j$. The long-run fraction of late calls can be expressed as $\mathbb{E}L_j(x)/\mathbb{E}N_j$. Notice that $\mathbb{E}N_j$ can be computed as the integral of $\lambda(t)$ over the $j$th period, which we express as the deterministic constant $n_j$.

Thus, our goal is to minimize $\sum_{k=1}^{K} c_k x_k + \mathbb{E}O(x)$ while also minimizing $\mathbb{E}L_j(x)/n_j$ for $j = 1, 2, \ldots, J$. We assume that both $\mathbb{E}O(x)$ and $\mathbb{E}L_j(x)$ must be estimated through simulation. This problem can be framed in several ways.

**Weighted objectives** We can combine the objectives into a single objective with linear weights, so that we elect to

$$\min_{x \in \mathbb{Z}_+^K} \sum_{k=1}^{K} c_k x_k + \mathbb{E}O(x) + \sum_{j=1}^{J} \beta_j \mathbb{E}L_j(x)/n_j,$$

where $\mathbb{Z}_+^K$ is the set of $K$-dimensional vectors consisting of nonnegative integers. Solvers that can solve problems with integer variables can then be used to tackle this formulation. Still, one must select the weights $(\beta_j : j = 1, 2, \ldots, J)$ and it is not obvious how to do that.

**Multi-objective** A multi-objective formulation attempts to identify the set of non-dominated solutions. (A dominated solution is one for which some other solution does at least as well on all objectives and strictly better on at least one objective.) This approach can be advantageous if one wants to consider additional characteristics of the non-dominated solutions when selecting a winner. However, this approach is problematic in our example because we have 33 objective functions, meaning there are likely a huge number of non-dominated solutions, and it could be difficult to identify them with confidence.

**Stochastic constraints** In this formulation, we attempt to minimize one objective while constraining the others. The $J = 32$ service-level requirements are expressed as constraints of the form $\mathbb{E}L_j(x)/n_j \leq p$, where common choices of $p$ are 0.2 or similar. Thus we arrive at the formulation

$$\begin{aligned} \min_{x \in \mathbb{Z}_+^K} \quad & \sum_{k=1}^{K} c_k x_k + \mathbb{E}O(x) \\ \text{s.t.} \quad & \mathbb{E}L_j(x) - p n_j \leq 0 \qquad\qquad j = 1, 2, \ldots, J. \end{aligned} \tag{1}$$

The constraints involve expectations that we cannot compute exactly, and so feasibility is always in question. Sometimes the value $p$ that appears in these constraints appears in legal contracts, and then these constraints must be carefully enforced. But in the absence of such contracts, it seems plausible that a decision maker would accept a solution that is barely infeasible yet has a low objective function value. In this latter setting, one might explore options by modifying the value $p$, or a separate value $p_j$ for each Period $j$, and re-solving, though how this can be done methodically is unclear. Still, an advantage of the stochastic-constraints formulation is that one can *choose* which objective to minimize and which ones to constrain, which enables a directed exploration of the set of non-dominated solutions.

**Beyond expected values** Many alternatives to the constraints in (1) are possible. For example, we might place constraints on the tail probabilities of the distribution of the observed (random) fraction of late calls, such as $\mathbb{P}(L_j(x) \leq pN_j) \geq 1 - \alpha_j$ for all periods $j$. Chance constraints such as these can induce complex feasible regions that make optimization very difficult, e.g., see Prékopa (1970).

Constraints can also be placed on tails or risk measures such as conditional value-at-risk, though tractability is a nontrivial concern. For example, if $\bar{W}_j(x)$ is defined as the mean waiting time over all customers that arrive in a single instance of Period $j$ (equaling 0 if there are no arrivals) as a function of the staffing vector $x$, then we might want to ensure that $\mathbb{E}(\bar{W}_j(x) | \bar{W}_j(x) > w) \leq \gamma_j$ for some fixed $w$. Another alternative that takes the form of a chance constraint is $\mathbb{P}(W_j(x) \leq w) \geq 1 - \alpha_j$, where $W_j(x)$ is the waiting time of a (uniformly at random) selected customer calling during Period $j$, with $W_j(x) = 0$ when there are no arrivals in Period $j$.

We elect to focus on the **stochastic constraints** approach in the remainder of this paper. Despite its weaknesses, e.g., we can never be certain (except in very special situations) that a given solution is feasible, it is readily communicated, it aligns with formulations used extensively in (deterministic) nonlinear programming, and it is relatively standard. Moreover, the disaggregated form of the constraints allows us to identify periods in which infeasibilities lie for a given solution, suggesting natural updates to the solution.

## 3 STOCHASTICALLY CONSTRAINED SIMULATION OPTIMIZATION

The generic simulation-optimization problem we consider is

$$\begin{aligned} \min_x \ & f(x) = \mathbb{E}F(x,\xi) \\ \text{s.t. } & g(x) = \mathbb{E}G(x,\xi) \le 0 \\ & h(x) \le 0 \\ & x \in \mathcal{D}. \end{aligned} \tag{2}$$

In the formulation (2), the domain $\mathcal{D}$ is non-random and may reflect, e.g., box constraints on the decision variables or a restriction that the decision variables be integers. (In this paper, we almost always consider a fixed solution $x$, so we need not distinguish between continuous, integer-ordered or mixed decision variables, though of course that distinction is important in developing solvers.) The vector-valued function $h(x)$ represents a set of deterministic constraints for which we can exactly (to numerical precision) evaluate the left-hand sides, e.g., lower and upper bounds on decision variables. The choice of whether to impose such bounds through $h$ or through the domain $\mathcal{D}$ is a matter of convenience and style.

The random element $\xi$ represents all random variables that are needed to generate a single replication of the simulation model, and the real-valued function $F$ represents the simulation logic needed to generate a random observation of the objective function, $F(x,\xi)$. The objective function $f(x)$ may not require simulation, in which case $F(x,\xi) = f(x)$ does not depend on $\xi$. But in general, we will need to estimate $f(x)$ through some estimator such as $\bar{F}(x) = n^{-1}\sum_{i=1}^{n} F(x,\xi_i)$, where $\xi_1,\ldots,\xi_n$ are independent and indentically distributed (iid) random elements with the same distribution as $\xi$.

The function $G$ is analogous to $F$, but is $\mathbb{R}^r$-valued because it represents a realization of the left-hand side of $r \ge 1$ so-called *stochastic* constraints. The term "stochastic constraint" refers to the fact that we cannot compute $g(x) = \mathbb{E}G(x,\xi)$ exactly, and instead use an estimator of the form $\bar{G}(x) = n^{-1}\sum_{i=1}^{n} G(x,\xi_i)$. Here, the iid random elements $\xi_1,\xi_2,\ldots,\xi_n$ will almost always be the same elements used in estimating $f(x)$, so that the estimators $\bar{F}(x)$ and $\bar{G}(x)$ are dependent, as are the components of the vector $\bar{G}(x)$.

In the context of the call-center example in the previous section, $\mathcal{D} = \mathbb{Z}_+^K$ and $r = J = 32$ because we have one constraint for each period. Also, $G(x,\xi) = (L_j(x,\xi) - pn_j \colon j = 1,2,\ldots,J)$ and $F(x,\xi) = \sum_{k=1}^{K} c_k x_k + O(x,\xi)$, where $\xi$ represents the arrival times of calls to the call center and their associated service times. There is dependence among the components of the random vector $G(x,\xi)$ because, e.g., many late calls in one period indicates congestion, which can carry over to the next period. Moreover, many late calls towards the end of the day suggests that there will be nontrivial overtime $O(x)$, so there is also dependence between the constraint random vector $G(x,\xi)$ and the objective function random variable $F(x,\xi)$.

In the formulation (2), we have chosen to express the left-hand side of the stochastic constraints as $g(x) = \mathbb{E}G(x,\xi)$. It is conceivable that one might want to further generalize the form of such constraint left-hand sides to nonlinear functions of expectations, e.g., $\psi(\mathbb{E}G(x,\xi))$, where $G(x,\xi)$ is, as before, a random vector and $\psi$ is a vector-valued nonlinear function. We have not encountered such problems in practice, so while there may be value in considering such a generalization, we do not pursue that here.

The stochastic constraints $\mathbb{E}G(x,\xi) \le 0$ are sufficiently general to encompass chance constraints, such as the constraints $\mathbb{P}(W_j(x) \le w) \ge 1-\alpha_j$, $j = 1,2,\ldots,J$ in the call-center example. In that setting, we would take $G(x,\xi) = (1-\alpha_j - \mathbb{I}(W_j(x) \le w) \colon j = 1,2,\ldots,J)$, with $\mathbb{I}(\cdot)$ equaling 1 if its

argument is true and 0 otherwise. Still with the call-center example, one could also express the constraints $\mathbb{E}(\bar{W}_j(x)|\bar{W}_j(x) > w) \leq \gamma_j,\ j = 1, 2, \ldots, J$ using $G(x, \xi) = (\bar{W}_j(x)\mathbb{I}(\bar{W}_j(x) > w) - \gamma_j \mathbb{I}(\bar{W}_j(x) > w) \colon j = 1, 2, \ldots, J)$, so that such constraints also fit within our formulation (2).

How can problems with stochastic constraints be solved? Space constraints preclude an in-depth discussion. Instead, we simply point to penalty methods, barrier methods and related approaches; see, e.g., Lan and Zhou (2016) and Zhang et al. (2022).

## 4 FEASIBILITY SCORES

We turn our focus to the task of assessing the feasibility (or infeasibility) of a given solution $x \in \mathcal{D}$ with respect to the stochastic constraints $g(x) \leq 0$. Recall that running $n$ replications at $x$ and averaging the outputs $G(x, \xi_1), G(x, \xi_2), \ldots, G(x, \xi_n)$ yields a point estimator $\bar{G}(x)$ for $g(x) = \mathbb{E}G(x, \xi)$. When no confusion can arise, we will suppress $x$ in the notation and simply write $\bar{G}, g$, etc. The estimator $\bar{G}$ offers a simple, yet crude, estimate of whether the simulated solution is feasible or not: if $\bar{G} \leq 0$, then $x$ looks feasible, while if $\bar{G} \nleq 0$, then $x$ looks infeasible. Although this approach of classifying a solution as either feasible or infeasible is straightforward and routinely used, it has two serious shortcomings: it does not indicate the *degree* to which the solution is feasible, and it fails to reflect the *uncertainty* one has about the solution's feasibility. To address the first issue, we introduce feasibility scores, which roughly measure the signed distance between the realized value of $\bar{G}$, denoted by $\bar{g}$, and the boundary of the nonpositive orthant. We address the second issue by supplementing feasibility scores with confidence intervals.

The feasibility scores we introduce are based on measures of how close $g$ is to satisfying or violating the stochastic constraints *if $g$ were known*. Since this is not typically the case, we present analogous metrics with the estimate $\bar{g}$ plugged in for $g$. In other words, though not otherwise explicitly stated hereafter, there is some true, but unknown, value $s(g)$ that quantifies how feasible a given solution is, and we settle for estimating this quantity with $s(\bar{g})$.

For problems with stochastic constraints, the geometry of the feasible region is, in general, unknown and cannot be ascertained from finite sampling. It is therefore impossible to determine a given solution's minimum distance to the boundary of the feasible region, as measured in the input space (that of $x$). We instead opt to calculate distances in the output space (that of $G(x, \xi)$) to assess how close a solution's expected performance is to satisfying or violating the stochastic constraints.

For a given solution $x$ and estimate $\bar{g}$, we define the *feasibility score*

$$s(\bar{g}) = \begin{cases} \inf\{d(y, \bar{g}) \colon y \nleq 0\} & \text{if } \bar{g} \leq 0, \\ -\inf\{d(y, \bar{g}) \colon y \leq 0\} & \text{if } \bar{g} \nleq 0, \end{cases} \tag{3}$$

where $d(\cdot, \cdot)$ is a function measuring distances in $\mathbb{R}^r$; specific examples will be given shortly. In words, the feasibility score is the minimum distance between $\bar{g}$ and the boundary of the nonpositive orthant $\{y \in \mathbb{R}^r \colon y \leq 0\}$. From (3), we see that the feasibility score can be expressed in terms of the solutions of two optimization problems: one that measures the minimum distance to the complement of the nonpositive orthant (to infeasibility) and another that measures the minimum distance to the nonpositive orthant (to feasibility). Since at least one of these distances is zero, we need to solve only one optimization problem, the identity of which depends on whether $\bar{g} \leq 0$ or $\bar{g} \nleq 0$.

By considering both the distance to feasibility and the distance to infeasibility, the feasibility score gives equal importance to how close a solution is to having feasible or infeasible performance. Furthermore, feasible solutions are not all regarded as being equally feasible, and likewise for infeasible solutions. The sign of the feasibility score indicates whether the solution looks feasible (positive) or infeasible (negative) based on the estimate $\bar{g}$. In addition, the magnitude of the feasibility score indicates the degree of feasibility or infeasibility, with more extreme values signifying that $g$ is believed to be far from the boundaries described by the constraints. A value of zero corresponds to a solution having expected performances that are on the boundary of the feasible region, i.e., the solution satisfies all of the stochastic constraints, with at least one constraint being binding.

Various choices of distance function can be used in the feasibility score. For instance, for the $L^p$ distance function $d(y, \bar{g}) = \|y - \bar{g}\|_p = (\sum_{i=1}^r |y_i - \bar{g}_i|^p)^{1/p}$ for $p \geq 1$,

$$s(\bar{g}) = \begin{cases} -\max_l \bar{g}_l & \text{if } \bar{g} \leq 0, \\ -\|\bar{g}^+\|_p & \text{if } \bar{g} \not\leq 0; \end{cases}$$

where $g = (g_1, g_2, \ldots, g_r)$ and $\bar{g}^+ = \max(\bar{g}, 0)$ with the max operator applied element-wise. The choice among various $L^p$ distance functions depends on whether the decision maker views the degree to which a solution is infeasible as being best represented by the maximum constraint violation ($L^\infty$), the sum of constraint violations ($L^1$), or the sum of squared constraint violations ($L^2$). A complicating factor is that the expected performances appearing in the constraints (the components of $g$) may be of different orders of magnitude or be measured in different units. The feasibility score defined in (3) is best used when the stochastic constraints have all been appropriately scaled to ensure that violations of the constraints can be aggregated in a meaningful way.

We can quantify the (simulation) error in the estimated feasibility score. One approach is to bootstrap the individual observations, $G(x, \xi_i)$ for $i = 1, 2, \ldots, n$, and compute, for each bootstrap instance, a feasibility score using the resulting $\bar{g}$. In particular, let $G(x, \xi_1^{*b}), G(x, \xi_2^{*b}), \ldots, G(x, \xi_n^{*b})$ denote the $n$ bootstrapped left-hand-side vectors for the $b$th bootstrap instance, and let $B$ be the number of bootstrap instances. For each bootstrap instance $b = 1, 2, \ldots, B$, compute the sample mean $\bar{g}^{*b} = n^{-1} \sum_{i=1}^n G(x, \xi_i^{*b})$ and the feasibility score $s(\bar{g}^{*b})$. The $\alpha/2$ and $1 - \alpha/2$ sample quantiles of $s(\bar{g}^{*1}), s(\bar{g}^{*2}), \ldots, s(\bar{g}^{*B})$ yield an approximate $100(1-\alpha)\%$ confidence interval for the true feasibility score $s(g)$. The width of the bootstrap confidence interval will decrease as the number of replications, $n$, increases because $\bar{g}$ becomes less variable.

## 5 A HYPOTHESIS-TESTING PERSPECTIVE

An alternative to feasibility scores can be motivated through hypothesis testing. Suppose we wish to test $H_0 \colon g_0 \leq 0$, the hypothesis that the population mean of the constraint left-hand side, $g_0 = \mathbb{E}G(x, \xi)$, lies in the nonpositive orthant and thus $x$ is feasible against the alternative hypothesis $H_1 \colon g_0 \not\leq 0$, i.e., that $x$ is infeasible. (As above, we fix $x$ and thus suppress it when no confusion can arise. Also, we are using standard notation from hypothesis testing where the hypothesised quantity receives a suffix 0. Thus, $g_0$ is a vector; the suffix does not refer to a component of a vector $g$.)

Generalized likelihood ratio tests are a standard approach to hypothesis testing with attractive optimality properties; see, e.g., § 9.5 of Rice (1988). In this setting, we compute the test statistic

$$-2 \ln \left( \frac{\sup_{g_0 \leq 0} \mathcal{L}(g_0)}{\sup_{g_0 \in \mathbb{R}^r} \mathcal{L}(g_0)} \right), \tag{4}$$

where $\mathcal{L}(g_0)$ is the likelihood of the observed mean, $\bar{g}$, as a function of the true mean $g_0$. If we assume that $\bar{G}$ is normally distributed with mean vector $g$ and positive definite variance-covariance matrix $\Sigma/n$, then the likelihood is simply the multivariate normal density with parameters $g_0$ and $\Sigma/n$. When $\bar{g} \not\leq 0$, (4) simplifies to

$$\inf_{g_0 \leq 0} (\bar{g} - g_0)^\intercal (\Sigma/n)^{-1} (\bar{g} - g_0). \tag{5}$$

The problem (5) is a convex quadratic minimization problem that can be easily solved numerically through, e.g., active-set methods.

When the sample mean $\bar{g} \leq 0$, the value of the test statistic is 0. We would prefer more information than this simple signal that the constraints are estimated to be feasible. Such information can be obtained by instead testing $H_0 \colon g_0 \not\leq 0$ ($x$ is not feasible) versus the alternative $H_1 \colon g_0 \leq 0$ ($x$ is feasible). The corresponding test statistic is

$$-2 \ln \left( \frac{\sup_{g_0 \not\leq 0} \mathcal{L}(g_0)}{\sup_{g_0 \in \mathbb{R}^r} \mathcal{L}(g_0)} \right),$$

which simplifies to

$$\inf_{g_0 \not\leq 0} (\bar{g} - g_0)^\intercal (\Sigma/n)^{-1} (\bar{g} - g_0). \tag{6}$$

The feasible region of the optimization problem in (6) is a union of $r$ half spaces and thus non-convex, but its geometry is exploitable. We can solve $r$ subproblems, with the $i$th subproblem having the same objective function but with feasible region $\{g_0 \in \mathbb{R}^r \colon g_0(i) = 0\}$, and take the minimum of the resulting optimal objective function values. (Here, $g_0(i)$ indicates the $i$th component of the vector $g_0$.) These subproblems are very easily solved.

We can combine the test statistics (5) and (6) into a single statistic by taking their difference which, since one test statistic is always 0, can be expressed as

$$\ell(\bar{g}) = \begin{cases} \inf_{g_0 \not\leq 0} (\bar{g} - g_0)^\intercal (\Sigma/n)^{-1} (\bar{g} - g_0) & \text{if } \bar{g} \leq 0, \\ -\inf_{g_0 \leq 0} (\bar{g} - g_0)^\intercal (\Sigma/n)^{-1} (\bar{g} - g_0) & \text{if } \bar{g} \not\leq 0. \end{cases} \tag{7}$$

This signed quantity is nonnegative when $\bar{g} \leq 0$, equal to 0 when $\bar{g}$ lies on the boundary of the nonpositive orthant, and negative when $\bar{g} \not\leq 0$.

Under a finite-second-moment assumption, the central limit theorem ensures that $\bar{G}$ is at least approximately normally distributed, so the use of the normal density above is plausible. Still, if $\bar{g}$ and $g_0$ are far apart, then the likelihood ratio computed under the normal density as above may be far from the correct ratio of likelihoods because we are then in the realm of large deviations. Thus, this measure using normal densities should be viewed as an imperfect, yet potentially useful, measure.

The statistic (7) is the difference between two squared Mahalanobis distances, where one of those distances is always equal to 0. As such, it is closely related to the feasibility scores we discussed in Section 4, except that we measure distances using the squared Mahalanobis norm instead of $L^p$ norms for some $p$. The Mahalanobis norm removes the effect of scalings and correlations in the components of the estimates of the constraint left-hand side. The Mahalanobis distance naturally reflects how certain "feasibility directions" may be easier to achieve through modification of the decision vector $x$ than others. For instance, in the call-center example, the number of lost calls in adjacent periods are positively correlated, so it is easier to simultaneously reduce infeasibility in multiple adjacent periods through modifications of the decision vector $x$ than to have opposite effects on infeasibility in adjacent periods.

We will rarely, if ever, know the exact value of $\Sigma$, so we will almost always replace it with the sample covariance matrix, $S$, say. Under moderate regularity conditions, e.g., $G(x, \xi)$ has a density, $S$ is invertible provided that $n \geq r + 1$, i.e., there are more data than constraints. There are potentially some practical situations where $S$ might not be invertible. For example, suppose our constraints include $a \leq \mathbb{E}Y(x) \leq b$ for constants $a < b$, where $Y(x)$ is some observable random variable. These constraints would be encoded in our framework using $G(x, \xi) = (-Y(x) + a, Y(x) - b)$. Because there is perfect negative correlation between the two elements of this random vector, $\Sigma$ and the sample covariance matrix $S$ will both be singular. (There is a straightforward solution to this difficulty, which for space reasons we omit here.)

An imperfection of the statistic (7) is that it depends on the assumed sample size $n$. Indeed, under the finite-second-moment assumption, $\bar{G}$ converges to $g$ and the sample covariance matrix $S$ converges to $\Sigma$ as $n \to \infty$, almost surely. Then, the objective function in the optimization problems defining the statistic (7) is asymptotically of the same order as $n(g - g_0)^\intercal \Sigma^{-1} (g - g_0)$. If $g(x)$ is not on the boundary of the feasible region, then the statistic (7) will converge at a linear rate to $\infty$ or $-\infty$ as $n \to \infty$, depending on whether $x$ is feasible or infeasible, respectively. This asymptotic (in $n$) behavior is problematic because the original constraint $g(x) \leq 0$ depends only on the mean of $G(x, \xi)$, which is somewhat at odds with a metric that depends on the sample size. This dependence could be problematic when trying to compare the feasibility of two solutions $x_1$ and $x_2$ that were evaluated with different numbers of replications. Moreover, even for a common sample size $n$ and the same left-hand sides, i.e., $g(x_1) = g(x_2)$, the respective values of (7) for solutions $x_1$ and $x_2$ can differ if their covariance matrices $\Sigma(x_1)$ and $\Sigma(x_2)$ differ. This too seems

undesirable, because from the perspective of a risk-neutral decision maker, both solutions are equally good (or bad); only their expected values matter.

An alternative definition of the statistic (7) that alleviates some of these shortcomings is to use the same optimization problems above, but to replace the scaled sample covariance matrix $S/n$ with just the sample covariance matrix $S$. This alternative can be interpreted as the squared distance from $\bar{g}$ to the boundary of the feasible region as measured in standard deviations, not standard errors. This quantity is unaffected by $n$ and instead fluctuates due to errors in the estimators $\bar{G}$ of $g$ and $S$ of $\Sigma$. This quantity is again dimensionless, though it still relies on the covariance matrix at the solution $x$ that is under consideration, which is at odds with a risk-neutral decision maker as discussed above.

## 6 BAYESIAN-INSPIRED METRICS

Feasibility scores (3) and the likelihood-ratio score (7) are real-valued quantities related to the position (in different senses) of $g$, the true mean of the constraint left-hand side, relative to the feasible region. A related, alternative measure can be found through a Bayesian perspective of the problem. In this perspective, the true mean $g$ and the true covariance matrix $\Sigma$ are considered to be random quantities with a specified prior distribution that is then updated, based on the simulation outputs we generate, to obtain a posterior distribution. We can then compute the expected feasibility score and the posterior probability of feasibility.

For simplicity and computational convenience, we will adopt the usual conjugate setup. To that end, we assume that $g|\Sigma \sim \mathcal{N}(0, \Sigma)$ and that $\Sigma$ is distributed according to an inverse Wishart distribution with appropriate parameters; see § 3.6 of Gelman et al. (2013). We further assume that, conditional on $g$ and $\Sigma$, the simulation replications $G(x, \xi_1), G(x, \xi_2), \ldots, G(x, \xi_n)$ are iid multivariate normal with mean vector $g$ and covariance matrix $\Sigma$. Let $\mathcal{F}$ be the sigma field generated by the simulation output data $G(x, \xi_1), G(x, \xi_2), \ldots, G(x, \xi_n)$. The (marginal) posterior distribution of $g$, conditional on $\mathcal{F}$, is then multivariate $t$ with parameters that depend on the prior distribution. For the choice of a non-informative prior, $g$ has a posterior multivariate $t$ distribution with location parameter $\bar{g}$, scale matrix $(n(n-r))^{-1} \sum_{i=1}^{n} (G(x, \xi_i) - \bar{g})(G(x, \xi_i) - \bar{g})^\intercal$ and degrees of freedom $n - r$; see § 3.6 of Gelman et al. (2013) for details, recalling that $r$ is the number of constraints and thus the dimension of $\bar{g}$ and $G$.

With this posterior distribution in hand, we can think about the *expected* feasibility score

$$\mathbb{E}(s(g)|\mathcal{F}), \tag{8}$$

where the expectation is over the posterior distribution of $g$. For the variations of $s(g)$ presented in Section 4, (8) involves integrating a multivariate $t$ density against a piecewise function. Although (8) may not be directly computable, Monte Carlo sampling can offer an approximation. Other penalty functions, such as those investigated in He and Kim (2019) and Chen et al. (2023), can be used in place of $s(g)$.

We can also consider the posterior probability of feasibility

$$\mathbb{P}(g \leq 0|\mathcal{F}) = \Phi_{n-r}(-\sqrt{n} S^{-1/2} \bar{g}), \tag{9}$$

where $\Phi_{n-r}$ is the cumulative distribution function of a multivariate $t$ random variable with location 0, scale matrix the identity and $n - r$ degrees of freedom, and $S^{-1/2}$ is a matrix square root of $S^{-1}$. The quantity in (9) will not usually be computable in closed form, but it can be evaluated using numerical integration techniques provided that the number of constraints $r$ is not too large (Matlab 2023), or approximated (rapidly) using Monte Carlo when $r$ is large. The posterior probability of feasibility may be easier to compute than the likelihood metric (7) because it does not involve any optimization.

A frequentist analog to the posterior probability that $g \leq 0$ is the (frequentist) probability $\mathbb{P}(\bar{G} \leq 0)$, where $\bar{G}$ is the sample mean of $G(x, \xi_1), G(x, \xi_2), \ldots, G(x, \xi_n)$. Under the assumption that these random vectors are iid normal random vectors with mean $g$ and covariance matrix $\Sigma$, $\bar{G}$ is normally distributed with mean vector $g$ and covariance matrix $\Sigma/n$. Thus, $\mathbb{P}(\bar{G} \leq 0) = \Phi(-\sqrt{n}\Sigma^{-1/2} g)$, where $\Phi$ is the cumulative distribution function of a standard normal random vector. We might approximate the latter expression by

the plug-in estimator $\Phi(-\sqrt{n}S^{-1/2}\bar{g})$; this estimator warrants some additional scrutiny, which we do not pursue here, due to the use of the plug-in estimators $S$ and $\bar{g}$.

**Remark 1** In both the Bayesian and frequentist explorations above, we assumed that $G(x,\xi)$ was normally distributed. This will typically hold at best only approximately. In that setting, we can still compute the diagnostics above, but they should not be viewed as accurate estimates of the quantities $\mathbb{P}(g \leq 0|\mathcal{F})$ and $\mathbb{P}(\bar{G} \leq 0)$ respectively. For example, when $g \not\leq 0$ in the frequentist setting, the event $\{\bar{G} \leq 0\}$ is a large deviation, and large deviation probabilities can be very complex to estimate accurately. In such a setting, we can still compute our diagnostics and use them as a rough sense of feasibility, or lack thereof, but they could be off by orders of magnitude relative to the true values.

An alternative to the plug-in estimator, using bootstrapping, is

$$\hat{p} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}(\bar{g}^{*b} \leq 0), \tag{10}$$

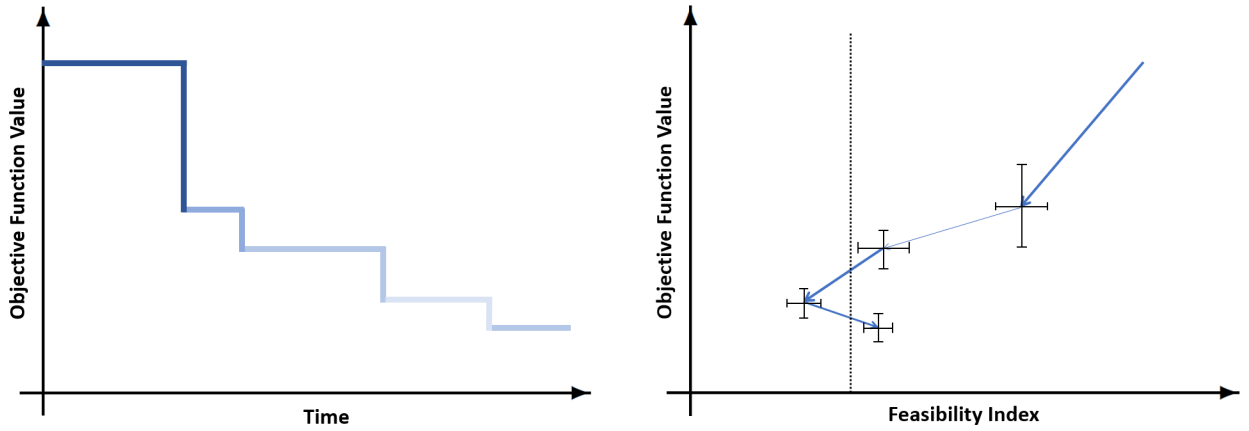
where $\bar{g}^*$ is as defined in Section 4, and the indicator function $\mathbb{I}(\bar{g}^{*b} \leq 0)$ equals 1 if $n^{-1}\sum_{i=1}^{n} G_j(x,\xi_i^{*b}) \leq 0$ for all $j = 1,2,\ldots,r$ and otherwise equals 0. This metric does not assume normality (or even that the joint distribution of the responses has elliptical contours). Bootstrapping also avoids numerical integration for large $r$, in contrast to the Bayesian metric $\mathbb{P}(g \leq 0|\mathcal{F})$. The bootstrapping estimator (10) estimates $\mathbb{P}(\bar{G} \leq 0)$ by the corresponding probability for the sample mean of bootstrap samples. We can expect this estimator to be reasonably accurate provided that $n$ is sufficiently large that the empirical (joint) distribution of $G$ is a good estimate of its true distribution and $\mathbb{P}(\bar{G} \leq 0)$ is neither too close to 0 nor too close to 1; see § 7.4 of Efron and Tibshirani (1994).

## 7 PLOTTING FEASIBILITY

We next consider how measures of feasibility might be incorporated into plots showing the performance of a simulation-optimization solver on a problem with stochastic constraints. We primarily consider a single run of a solver and the sequence of solutions it recommends as the best feasible solutions as it expends a fixed budget of simulation replications. At each recommended solution, we can calculate an associated feasibility metric. (For concreteness, we will focus our discussion on feasibility scores, but the proposed approach can be adapted to work with the other metrics.) It is natural to wonder why we should even concern ourselves with the feasibility of solutions recommended before exhausting the budget. Indeed, the terminal solution is the one most likely to be implemented by the decision maker, and thus its feasibility is of the greatest consequence. We contend that the feasibility of intermediate solutions is important for understanding *how* the algorithm arrives at its final recommendation: For example, does it approach the terminal solution from outside the feasible region? Or does the algorithm mostly explore within the feasible region and only gradually learn where its boundaries are? This level of insight into a solver's search behavior can be useful for researchers developing new solvers.

Suppose we were to compute the feasibility metrics at the recommended solutions based on the replications obtained by the solver over the course of its run. The resulting metrics would demonstrate how feasible the recommended solutions would have looked to the solver during its search. We might therefore choose to look unfavorably upon a solver that recommends solutions for which the data strongly suggested they were infeasible, e.g., those with very negative feasibility scores. This straightforward application of feasibility metrics is not without issue. For instance, some solvers might recommend solutions they have not simulated, such as the solution reported when applying Polyak-Ruppert averaging.

The approach we adopt is to have a post-processing stage in which we obtain $n$ fresh replications at each recommended solution. The outputs of these additional replications, referred to as post-replications, will be used to estimate both the objective function values and feasibility of the recommended solutions. Post-replications allow us to remove any bias coming from the solver's internal evaluation of the solutions' feasibility. For instance, the solver may tend to recommend a solution only if it clearly looks feasible when




(a) Augmented Progress Curve

(b) Objective-Feasibility Scatterplot

Figure 1: Proposed plots of estimated objective function value and feasibility metrics of solutions recommended by a solver over time on a single run. In (a), color depicts the feasibility metric, with darker shades indicating greater feasibility. In (b), line thickness depicts the length of time between recommended solutions, with thicker lines indicating longer spans of time. The vertical dotted line represents borderline feasibility and marginal confidence intervals are depicted for each recommended solution.

simulated. (An analogous form of optimization bias is present in a solver's evaluations of the objective function value of recommended solutions.) The variability of feasibility metrics can be controlled through the number of post-replications. As we have seen, for some feasibility metrics, namely the likelihood-ratio inspired metric (7), the posterior probability of feasibility (9), and the nonparametric estimator (10), their distribution (and realized values) will be heavily influenced by the number of post-replications.

We propose two approaches for simultaneously depicting the objective function estimates and feasibility metrics of the solutions recommended over time by a solver on a single run. The first approach is to augment the (unnormalized) progress curve plots of Eckman et al. (2023), which show the estimated objective function values of recommended solutions over time in a piecewise constant curve. To convey the estimated feasibility of recommended solutions in a progress curve, one can change the color or thickness of the line segments, depending on the feasibility metric of the corresponding recommended solution. Some metrics with unbounded ranges, like the deterministic feasibility score and likelihood-ratio inspired metric, must first be mapped onto a bounded interval. The variable-color idea is illustrated in Figure 1a.

The second approach, shown in Figure 1b, is to plot pairs of the objective function value and feasibility metric for each recommended solution and connect the points by arrows showing the ordering of the recommended solutions over time. Because the elapsed time between consecutive recommended solutions can vary, the arrows need to convey the length of the time between recommendations; this may be done by using different line thicknesses. This plot makes objective function values and feasibility metrics equally prominent, and therefore helps to show how a solver handles the trade-off between improving the objective while maintaining feasibility. As seen in Figure 1b, a vertical line can be added to indicate the value of the feasibility metric that signifies borderline feasibility. In the case of the feasibility score and likelihood scores, this line would be positioned at zero. Compared to the augmented progress curve, the objective-feasibility scatter plot can display confidence intervals around the estimated objective function value and the feasibility metric, e.g., Figure 1b depicts these confidence intervals around each point. A more sophisticated approach would be to construct and plot a joint confidence region, e.g., an ellipsoid, but we believe the simpler marginal confidence intervals can still convey the essence of the estimation error.

A critical, and so far unresolved, challenge is how to build upon these plots to display the performance of a solver over multiple runs, i.e., macroreplications, and compare multiple solvers. If the number

of macroreplications were small, say less than ten, then one could superimpose the curves from the individual macroreplications to get a sense of the run-to-run variability of the solver's behavior. For more macroreplications, and for comparing multiple solvers, the information shown in the individual curves must be summarized. Eckman et al. (2023) aggregate the objective function values of the solutions recommended at a given time $t$ for progress curves, reporting the mean, median, or quantiles. For feasibility metrics, however, averaging across replications is more fraught. If a solver recommends very feasible solutions on some macroreplications and very infeasible solutions on others, the mean feasibility metric will give the false impression that the solver is recommending solutions near the boundary of the feasible region. How best to summarize a collection of objective-feasibility curves remains a subject of ongoing research.

## 8 CONCLUSION

We explored how to measure the feasibility of a solution in the presence of stochastic constraints. Our goal is to stimulate discussion and to obtain advice from the research community. We have proposed feasibility scores, a likelihood-ratio score and the posterior probability of feasibility as potential metrics to indicate the degree of feasibility of a solution. Of these, feasibility scores are conceptually the most straightforward and consistent with the views of a risk-neutral decision maker. However, the other measures have their strengths and no metric should be viewed as dominant. The choice of which metric is best suited for a given problem will depend on the associated computational expense and the decision maker's statistical philosophy, tolerance toward risk, and regard for violating one or more constraints. Future research can shed light on how the metrics compare in terms of their theoretical and empirical performance, both for estimation and optimization. We have also discussed how such measures could be used in plots that depict solver progress. We are exploring these ideas with a view to application in SimOpt (Eckman et al. 2021).

## ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation Grants CMMI-2206972, CMMI-2035086, and CMMI-2226347.

## REFERENCES

Calafiore, G. C., and M. C. Campi. 2006. "The Scenario Approach to Robust Control Design". *IEEE Transactions on Automatic Control* 51(5):742–753.

Charnes, A., and W. W. Cooper. 1959. "Chance-Constrained Programming". *Management Science* 6(1):73–79.

Charnes, A., W. W. Cooper, and G. H. Symonds. 1958. "Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil". *Management Science* 4(3):235–263.

Chen, W., S. Gao, W. Chen, and J. Du. 2023. "Optimizing Resource Allocation in Service Systems via Simulation: A Bayesian Formulation". *Production and Operations Management* 32(1):65–81.

Cooper, K., and S. R. Hunter. 2020. "PyMOSO: Software for Multiobjective Simulation Optimization with R-PERLE and R-MinRLE". *INFORMS Journal on Computing* 32(4):1101–1108.

Dentcheva, D., and A. Ruszczynski. 2003. "Optimization with Stochastic Dominance Constraints". *SIAM Journal on Optimization* 14(2):548–566.

Digabel, S. L., and S. M. Wild. 2015. "A Taxonomy of Constraints in Simulation-Based Optimization". *arXiv preprint arXiv:1505.07881*.

Eckman, D. J., S. G. Henderson, and S. Shashaani. 2023. "Diagnostic Tools for Evaluating and Comparing Simulation-Optimization Algorithms". *INFORMS Journal on Computing* 35(2):495–508.

Eckman, D. J., S. G. Henderson, S. Shashaani, and R. Pasupathy. 2021. "SimOpt". *GitHub repository*. https://github.com/simopt-admin/simopt [Online; accessed March 21, 2023].

Efron, B., and R. J. Tibshirani. 1994. *An Introduction to the Bootstrap*. Boca Raton, Florida: CRC press.

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

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian Data Analysis*. Boca Raton, Florida: CRC press.

He, J., and S.-H. Kim. 2019. "A New Reward Function for Bayesian Feasibility Determination". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 3480–3491. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Healey, 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* 24(3):Article 14, 1–25.

Homem-de Mello, T., and G. Bayraksan. 2014. "Monte Carlo Sampling-Based Methods for Stochastic Optimization". *Surveys in Operations Research and Management Science* 19(1):56–85.

Homem-de Mello, T., and G. Bayraksan. 2015. "Stochastic Constraints and Variance Reduction Techniques". In *Handbook of Simulation Optimization*, edited by M. C. Fu, 245–276. New York, New York: Springer.

Hunter, S. R., E. A. Applegate, V. Arora, B. Chong, K. Cooper, O. Rincón-Guevara, and C. Vivas-Valencia. 2019. "An Introduction to Multiobjective Simulation Optimization". *ACM Transactions on Modeling and Computer Simulation* 29(1):Article 7, 1–36.

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.

Koole, G., and S. Li. 2021. "A Practice-Oriented Overview of Call Center Workforce Planning". https://arxiv.org/abs/2101.10122 [Online; accessed May 4, 2023].

Lan, G., and Z. Zhou. 2016. "Algorithms for Stochastic Optimization with Functional or Expectation Constraints". *arXiv preprint arXiv:1604.03887*.

Lee, L. H., E. P. Chew, S. Teng, and D. Goldsman. 2010. "Finding the Non-Dominated Pareto Set for Multi-Objective Simulation Models". *IIE Transactions* 42(9):656–674.

Matlab 2023. "Multivariate $t$ Cumulative Distribution Function". https://www.mathworks.com/help/stats/mvtcdf.html [Online; accessed May 3, 2023].

Miller, B. L., and H. M. Wagner. 1965. "Chance Constrained Programming with Joint Constraints". *Operations Research* 13(6):930–945.

Nemirovski, A., and A. Shapiro. 2007. "Convex Approximations of Chance Constrained Programs". *SIAM Journal on Optimization* 17(4):969–996.

Prékopa, A. 1970. "On Probabilistic Constrained Programming". In *Proceedings of the Princeton Symposium on Mathematical Programming*, Volume 113, 138. Princeton, New Jersey.

Rice, J. A. 1988. *Mathematical Statistics and Data Analysis*. Pacific Grove, California: Wadsworth & Brooks/Cole.

Rockafellar, R. T., S. Uryasev et al. 2000. "Optimization of Conditional Value-at-Risk". *Journal of Risk* 2:21–42.

Shi, Z., Y. Peng, L. Shi, C.-H. Chen, and M. C. Fu. 2022. "Dynamic Sampling Allocation under Finite Simulation Budget for Feasibility Determination". *INFORMS Journal on Computing* 34(1):557–568.

Solow, D., R. Szechtman, and E. Yücesan. 2021. "Novel Approaches to Feasibility Determination". *ACM Transactions on Modeling and Computer Simulation* 31(1):1–25.

Zhang, L., Y. Zhang, J. Wu, and X. Xiao. 2022. "Solving Stochastic Optimization with Expectation Constraints Efficiently by a Stochastic Augmented Lagrangian-Type Algorithm". *INFORMS Journal on Computing* 34(6):2989–3006.

## AUTHOR BIOGRAPHIES

**DAVID J. ECKMAN** is an Assistant Professor in the Wm Michael Barnes '64 Department of Industrial and Systems Engineering at Texas A&M University. His research interests deal with optimization and output analysis for stochastic simulation models. He is a co-creator of SimOpt, a testbed of simulation optimization problems and solvers. His e-mail address is eckman@tamu.edu.

**SHANE G. HENDERSON** holds the Charles W. Lake, Jr. Chair in Productivity in the School of Operations Research and Information Engineering at Cornell University. His research interests include simulation theory and a range of applications including emergency services. He is an INFORMS Fellow. He is a co-creator of SimOpt, a testbed of simulation optimization problems and solvers. His email address is sgh9@cornell.edu and his homepage is http://people.orie.cornell.edu/shane.

**SARA SHASHAANI** is an Assistant Professor in the Edward P. Fitts Department of Industrial and System Engineering at North Carolina State University. Her research interests are probabilistic data-driven models and simulation optimization. She is a co-creator of SimOpt. Her email address is sshasha2@ncsu.edu and her homepage is https://shashaani.wordpress.ncsu.edu/.