Stochastic Constraints: How Feasible is Feasible?

David J. Eckman [1] Shane G. Henderson [2] Sara Shashaani [3]

[1]Texas A&M University, ISEN
[2]Cornell University, ORIE
[3]North Carolina State University, ISE

Winter Simulation Conference
December 13, 2023

This work is supported by the National Science Foundation under grants CMMI-2206972, CMMI-2035086, and CMMI-2226347.

Introduction

Stochastic constraints are inequality constraints involving expected values of rvs.

- Encompasses **chance constraints**, which involve probabilities.

Arise naturally in practice as bounds on acceptable secondary performance.

- E.g., mininum service requirements in a service system.

Past research has focused on solution methods (a.k.a. algorithms) for problems of

- stochastically constrained optimization (Hunter and Pasupathy, 2013; Homem-de Mello and Bayraksan, 2014, 2015) and
- feasibility determination (Gao and Chen, 2017; He and Kim, 2019; Shi et al., 2022).

This Talk

How feasible or infeasible is a given solution and how sure are we?

Call-Center Example

Problem of staffing a call center with a time-non-homogeneous arrival process.

- A 16-hour day is split into $J = 32$ 30-minute periods.
- Let x_k be the number of workers in Shift k, $k = 1, 2, \ldots, K$.

Expected staffing cost is $\sum_{k=1}^{K} c_k x_k + \mathbb{E}O(x)$ where $O(x)$ is *random* overtime cost.

Long-run fraction of late calls is $\mathbb{E}L_j(x)/n_j$ where n_j is the expected number of customers arriving during period j.

Want both *low* expected staffing costs and *low* expected fraction of late calls.

Formulations as an Optimization Problem

Weighted Objectives

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

Multiple Objectives

$$\min_{x\in\mathbb{Z}_+^K} \left\{ \sum_{k=1}^{K} c_k x_k + \mathbb{E}O(x) \text{ and } \mathbb{E}L_j(x)/n_j \text{ for } j = 1, 2, \ldots, J \right\}.$$

Stochastic Constraints

$$\min_{x\in\mathbb{Z}_+^K} \sum_{k=1}^{K} c_k x_k + \mathbb{E}O(x) \text{ subject to } \mathbb{E}L_j(x) - pn_j \leq 0 \text{ for } j = 1, 2, \ldots, J.$$

Stochastically Constrained Simulation Optimization

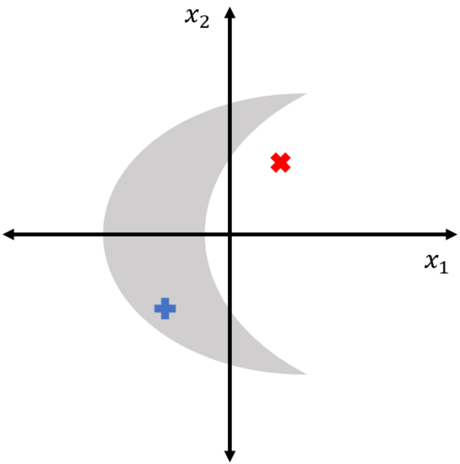
Generic Formulation

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

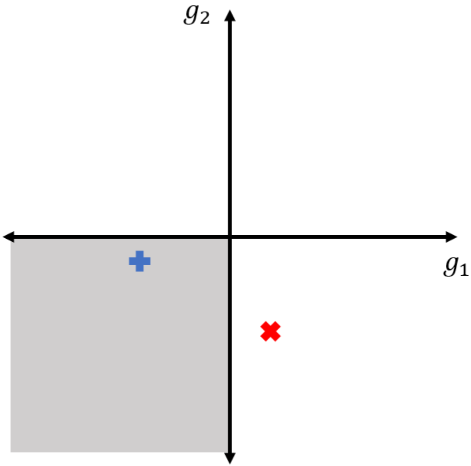
Estimate $f(x)$ by $\bar{F}(x) = \frac{1}{n}\sum_{i=1}^{n} F(x,\xi_i)$ and $g(x)$ by $\bar{G}(x) = \frac{1}{n}\sum_{i=1}^{n} G(x,\xi_i)$.

- Let $G(x,\xi)$ be vector valued, with r being the number of stochastic constraints.

Visualizing Stochastic Constraints



Input (x) space



Output (g) space

Setup

Fix a solution $x \in \mathcal{D}$ and sample size n.

Recall that $\bar{G}(x)$ is our estimator for $g(x) = \mathbb{E}G(x, \xi)$ based on n replications.

- We will use $\bar{G}$ and g for shorthand, and $\bar{g}$ for the realized value of $\bar{G}$.

Naive approach: If $\bar{G} \leq 0$, then x looks feasible. If $\bar{G} \not\leq 0$, then x looks infeasible.

Two shortcomings:

1. Does not indicate the **degree** to which x is feasible (or infeasible).
2. Fails to reflect the **uncertainty** about x's feasibility.

We will see a series of potential metrics aimed at overcoming these shortcomings.

Feasibility Scores

Feasibility scores are based on measures of how close g is to satisfying or violating the stochastic constraints, *were* g *known*.

We don't know g, so we use $\bar{g}$ as a plug-in:

$$s(\bar{g}) = \begin{cases} \inf\{d(y, \bar{g})\colon\, y \not\leq 0\} & \text{if } \bar{g} \leq 0, \\ -\inf\{d(y, \bar{g})\colon\, y \leq 0\} & \text{if } \bar{g} \not\leq 0, \end{cases}$$

where $d(\cdot, \cdot)$ is some distance function for $\mathbb{R}^r$.

In other words, $s(\bar{g})$ is the minimum distance between $\bar{g}$ and the boundary of the non-positive orthant.

Need to solve <u>one</u> optimization problem, based on whether $\bar{g} \leq 0$ or $\bar{g} \geq 0$.

Feasibility Scores

If $s(\bar{g}) \geq 0$, then x looks feasible. If $s(\bar{g}) < 0$, then x looks infeasible. If $s(\bar{g}) = 0$, then at least one constraint is binding.

Feasibility and infeasibility are given equal importance.

Feasible solutions are not regarded as equally feasible. Ditto for infeasible solutions.

For the L^p distance function $d(y, \bar{g}) = \|y - \bar{g}\|_p = \left(\sum_{i=1}^r |y_i - \bar{g}_i|^p\right)^{1/p}$ for $p \geq 1$,

$$s(\bar{g}) = \begin{cases} -\max_l \bar{g}_l & \text{if } \bar{g} \leq 0, \\ -\|\bar{g}^+\|_p & \text{if } \bar{g} \nleq 0; \end{cases}$$

where $\bar{g} = (\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_r)$ and $\bar{g}^+ = \max(\bar{g}, 0)$ element-wise.

Confidence Intervals for Feasibility Scores

We can quantify the (simulation) error in the estimated feasibility score via the **percentile bootstrap**.

Approach:

1. Bootstrap the individual observations $G(x, \xi_i)$ for $i = 1, 2, \ldots, n$.
2. For each bootstrap instance, compute the feasibility score for the realized $\bar{g}$.

The $\alpha/2$ and $1 - \alpha/2$ sample quantiles of these bootstrapped feasibility scores give an approximate $100(1 - \alpha)\%$ confidence interval for the true feasibility score $s(g)$.

A Hypothesis-Testing-Inspired Metric

Suppose we were to test the hypothesis

$$H_0\colon g_0 \leq 0 \quad \text{versus} \quad H_1\colon g_0 \nleq 0.$$

Generalized likelihood ratio tests would suggest the test statistic

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

where $\mathcal{L}(g_0)$ is the likelihood of $\bar{g}$ as a function of the g_0.

Assuming $\bar{G} \sim \mathcal{N}_r(g_0, \Sigma/n)$, the test statistic simplifies to

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

A similar test statistic arises from the test $H_0\colon g_0 \nleq 0$ versus $H_1\colon g_0 \leq 0$ for $\bar{g} \leq 0$.

A Hypothesis-Testing-Inspired Metric

Put the two cases together and replace unknown Σ with plug-in estimator S:

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

$\ell(\bar{g})$ is the difference of two squared **Mahalanobis distances** (one of which is 0).

- Removes the effects of scalings and correlations in components of $\bar{G}$.
- Reflects how some "feasibility directions" may be easier to achieve than others.

Downside: $\ell(\bar{g})$ *depends on n* and will in general converge to $\pm\infty$ as $n \to \infty$.

- Misaligned with motivation of a **risk-neutral** decision-maker.
- Replacing S/n with S does not resolve the risk-neutral issue.

Bayesian-Inspired Metrics

Bayesian perspective:

- True mean g and true covariance matrix Σ are *random*.
- Their posterior distribution depends on priors and observed simulation outputs.

Assumptions:

- Conjugate setup: $g \mid \Sigma \sim \mathcal{N}(0, \Sigma)$ and $\Sigma \sim \mathrm{InverseWishart}$.
- Conditional on g and Σ, $G(x, \xi_1), G(x, \xi_2), \ldots, G(x, \xi_n) \overset{i.i.d.}{\sim} \mathcal{N}_r(g, \Sigma)$.

For choice of non-informative prior, the *marginal* posterior distribution of g is given by

$$g \sim T_{n-r}\left(\bar{g}, \frac{1}{n(n-r)} \sum_{i=1}^{n} (G(x, \xi_i) - \bar{g})(G(x, \xi_i) - \bar{g})^\intercal\right) = T_{n-r}\left(\bar{g}, S/n\right).$$

Expected Feasibility Score

With the posterior distribution, we can calculate the **expected feasibility score**,

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

where $\mathbb{E}(\cdot \mid \mathcal{F})$ indicates expectation over the posterior distribution.

Entails integrating a multivariate t density against a piecewise function.

- If not directly computable, Monte Carlo can provide an approximation.

Notice

Degree of feasibility and its uncertainty are rolled into one metric.

Posterior Probability of Feasibility

Alternatively, we can compute the **posterior probability of feasibility**

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

where $\Phi_{n-r}(\cdot)$ is the cumulative distribution function of a standard MVT with $n - r$ df.

Entails integrating a multivariate t density over the non-positive orthant.

Not usually computable in closed form.

- Software can compute for modest r (say, ≤ 25).
- Or approximate via Monte Carlo if r is large.

A Bootstrap Estimator

An alternative to posterior probability of feasibility is

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

where $\bar{g}^* = n^{-1} \sum_{i=1}^{n} G(x, \xi_i^{*b})$ is the estimated LHS vector for bootstrap instance b.

Advantages:

- Does not rely on a normality assumption and avoids numerical integration.
- Should be reasonably accurate so long as n is sufficiently large.

Use Cases

How could these metrics be used?

1. Within a solver, to assess whether it has drifted into/out of the feasible region.
   - Navigation is less clear, but directional derivatives of these metrics (w.r.t. x) might be useful.

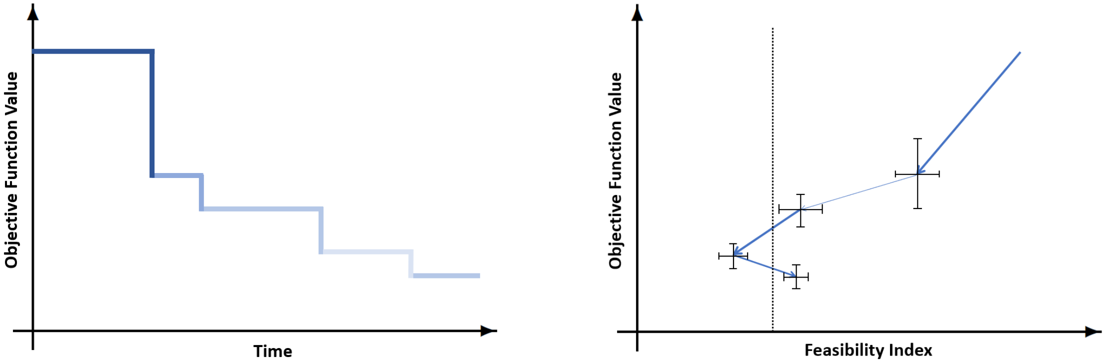
2. For studying a solver's search behavior.
   - How does it approach the terminal solution?
   - How aggressively does it enforce feasibility in the solutions it visits or recommends?

We are primarily interested in #2, as it pertains to experiments in SimOpt.

Goal: Augment plots of solver progress (in terms of objective function value) with some visualization of feasibility of recommended solutions.

Plotting Feasibility

Mock plots simultaneously depicting feasibility and optimality.



Aggregating feasibility metrics over multiple macroreplications remains a challenge.

Conclusion

Takeaway

There are many ways to quantify the (in)feasibility of solutions to stochastically constrained optimization problems; most (so far) are imperfect.

Future Work:

- Incorporating in plots in SimOpt.
- Design of solvers that employ these metrics to direct their search.
- Theoretical and empirical analysis on consistency and CI coverage.