Using Functional Properties To Screen Out Simulated Solutions In Large Decision Spaces

David J. Eckman
Matthew Plumlee
Barry L. Nelson

Northwestern University, IEMS

INFORMS Annual Meeting
November 9, 2020

Supported by NSF grants DMS-1854562 and DMS-1953111.

Introduction

Stochastic simulation is used to understand complex systems.

- **Examples:** manufacturing, emergency services, ride-sharing.

Simulation models often have *large decision spaces*, $\mathcal{X}$.

- Thousands or even millions of **solutions**.
- Simulating all solutions can be impractical.

Screening

Infer whether the **performance** of a given solution is **acceptable** based on initial simulation experiments at a subset $\mathrm{X} \subseteq \mathcal{X}$.

- If unacceptable, remove from consideration (i.e., *screen out*).

Related Approaches

Subset Selection

- Guarantees to retain the optimal solution with high probability.
- **Requires simulating all solutions.**

Metamodeling

- Gaussian processes – predict performance at unsimulated solutions.
- Neural networks – **lack statistical guarantees.**

Our framework, called **Plausible Screening**,

- *exploits information* about the performance function
- to screen out *unsimulated* solutions
- in a *statistically controlled* way.

Acceptable Solutions

For a given performance function $\mu\colon \mathcal{X} \mapsto \mathbb{R}$, let $\mathcal{A}$ denote the set of **acceptable** solutions.

Examples:

- **Optimization:** Ex: layouts near the minimum-cost layout.

$$\mathcal{A} \equiv \{x \in \mathcal{X}\colon \mu(x) \leq \min_{x' \in \mathcal{X}} \mu(x') + \delta\}.$$

- **Feasibility Determination:** Ex: layouts with WIP < 1000.

$$\mathcal{A} \equiv \{x \in \mathcal{X}\colon \mu(x) \leq \mu_0\}.$$

- **Comparison to a Target:** Ex: layouts attaining a desired throughput.

$$\mathcal{A} \equiv \{x \in \mathcal{X}\colon |\mu(x) - \mu^\dagger| \leq \epsilon\}.$$

Statistical Guarantees for a Subset $\mathcal{S}_n$

Finite-sample confidence

For any $\mu \in \mathscr{M}$ and sufficiently large sample sizes,

$$\mathbb{P}(x_0 \in \mathcal{S}_n) \geq 1 - \alpha \text{ for all } x_0 \in \mathcal{A}.$$

Asymptotic confidence

For any $\mu \in \mathscr{M}$ and sufficiently large sample sizes,

$$\mathbb{P}(x_0 \in \mathcal{S}_n) \gtrsim 1 - \alpha \text{ for all } x_0 \in \mathcal{A}.$$

Consistency

For any $\mu \in \mathscr{M}$ and sufficiently large sample sizes,

$$\mathbb{P}(x_0 \in \mathcal{S}_n) \to 0 \text{ for all } x_0 \notin \mathcal{A}.$$

Spaces of Performance Functions

Assumption

Decision-maker possesses *known or assumed* properties of μ.

- Techniques: *sample-path arguments* and *stochastic orders*.

$\mathscr{M} \equiv$ set of functions possessing specified functional properties.

$\mathscr{M}(x_0) \equiv$ set of functions in $\mathscr{M}$ for which x_0 is acceptable.

The projection of $\mathscr{M}(x_0)$ onto $\mathbb{R}^k$ is given by

$$\mathsf{M}(x_0) \equiv \left\{ \mathsf{m} \in \mathbb{R}^k : \text{there exists } m \in \mathscr{M}(x_0) \text{ such that } m(\mathsf{X}) = \mathsf{m} \right\}.$$

Interpretation

$\mathsf{M}(x_0) \equiv$ set of performance vectors of the solutions $x_1, \ldots, x_k$ for which there exists an interpolating function belonging to $\mathscr{M}(x_0)$.

Example: Minimizing a Convex Function

$$\mathscr{M} \equiv \text{ set of convex functions mapping } \mathcal{X} \mapsto \mathbb{R}.$$

$$\mathscr{M}(x_0) \equiv \text{ set of } x_0\text{-optimal convex functions mapping } \mathcal{X} \mapsto \mathbb{R}.$$

$$\begin{aligned} \mathrm{M}(x_0) = \{\mathrm{m} \in \mathbb{R}^k : \text{ there exists } \mathrm{m}_0 \in \mathbb{R} \text{ and } \xi_1, \ldots, \xi_k \in \mathbb{R}^d \text{ such that} \\ \mathrm{m}_i - \mathrm{m}_j - (x_i - x_j)^\top \xi_i \leq 0 \text{ for all } i, j = 1, \ldots, k \\ \mathrm{m}_i - \mathrm{m}_0 - (x_i - x_0)^\top \xi_i \leq 0 \text{ for all } i = 1, \ldots, k \\ -\mathrm{m}_i + \mathrm{m}_0 \leq 0 \text{ for all } i = 1, \ldots, k\}. \end{aligned}$$

- m_0 represents the performance of x_0.
- ξ_i represents a subgradient at $x_i \in \mathrm{X}$, $i = 1, \ldots, k$.

Minimum Standardized Discrepancy

Retain solutions x_0 for which $\widehat{\mu}$ is *sufficiently close* to $\mathsf{M}(x_0)$.

The **minimum standardized discrepancy** between $\widehat{\mu}$ and $\mathsf{M}(x_0)$:

$$D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma}) = \min_{\mathsf{m} \in \mathsf{M}(x_0)} \sum_{i=1}^{k} \frac{n_i}{\widehat{\sigma}_i^2} (\widehat{\mu}_i - \mathsf{m}_i)^2,$$

where $\mathsf{m} = (\mathsf{m}_1, \ldots, \mathsf{m}_k)$.

Calculating $D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma})$ entails solving a **QP** for each $x_0 \in \mathcal{X}$.

Large $D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma}) \implies$ strong evidence that x_0 is **unacceptable**.

Plausible Screening

For a given cutoff D, Plausible Screening (PS) returns the subset

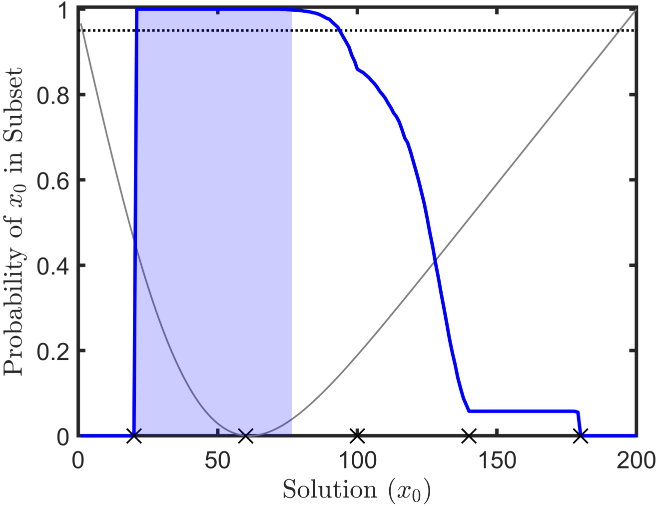
$$\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}} \equiv \left\{ x_0 \in \mathcal{X} \colon D_{\mathsf{n}}(x_0, \widehat{\mu}, \widehat{\Sigma}) \leq \mathsf{D} \right\}.$$

Theorem

For a suitable choice of D, $\mathcal{S}_{\mathsf{n}}^{\mathrm{PS}}$ achieves

- finite-sample confidence,
- asymptotic confidence, and
- S(X) consistency.

Newsvendor Problem



Average Subset Sizes: PS = **104/200** and STB = **199/200**.

Tandem Production Line Problem

Allocate 50 resources across 5 single-server stations in tandem.

- Infinite supply of products ready to process.
- Buffers in front of each station.
- Stations can become *blocked* or *starved*.

Objective

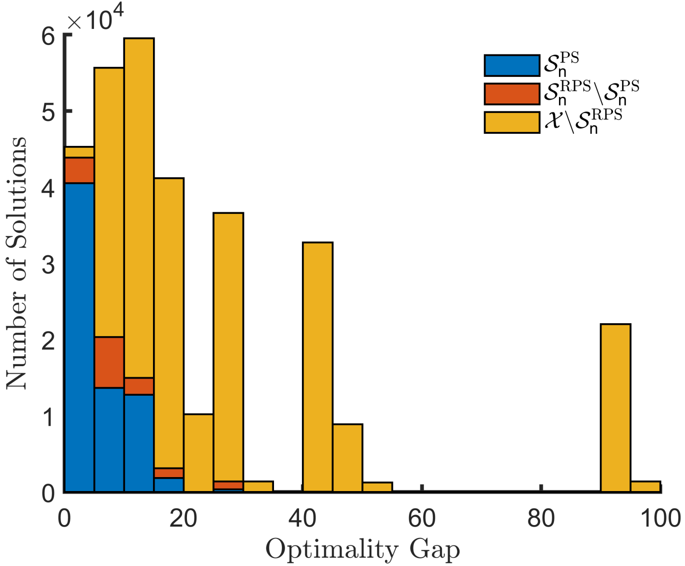
Minimize the expected completion time of 100th product.

- *Exponential* cycle times $\implies$ objective function is **convex**.

Simulated **100/316,251** feasible solutions, 100 replications each.

- PS screened out **247,053** solutions (**78%**).
- RPS screened out **232,503** solutions (**73%**), **20x** faster.

Tandem Production Line Problem



Conclusion

Takeaway

Plausible Screening can screen out **large swaths** of the feasible region, **without simulating all solutions**.

Future Work

- Incorporating other functional properties:
  - E.g., bounds, local information, stochastic gradients.
- How best to choose experimental set X and sample sizes?
- Sequential procedures with adaptive sampling.