Robust Sample Average Approximation

Presenter: Ben-Zheng Li

January 11, 2026

Reference: D. Bertsimas, V. Gupta, N. Kallus, Robust Sample Average Approximation, *Mathematical Programming*, 2018.

Stochastic optimization with unknown distribution

Given a decision space $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ and an uncertain parameter $\xi \in \Xi \subseteq \mathbb{R}^d$ with unknown distribution F, we consider the stochastic optimization problem

$$\min_{x \in \mathcal{X}} \ \mathbb{E}_{\xi \sim F}\left[c(x;\xi)\right].$$

Here:

- x is the decision variable,
- $c(x;\xi)$ is the cost incurred when uncertainty ξ realizes,
- F is the **true but unknown** distribution of ξ.

Baseline: Sample Average Approximation (SAA)

- **Idea.** Replace the unknown distribution F by the empirical distribution constructed from observed samples $\xi^1, \ldots, \xi^N$.
- **Empirical distribution.**

$$\hat{F} = \frac{1}{N} \sum_{j=1}^{N} \delta_{\xi^j}.$$

Baseline: Sample Average Approximation (SAA)

- **Idea.** Replace the unknown distribution F by the empirical distribution constructed from observed samples $\xi^1, \ldots, \xi^N$.
- **Empirical distribution.**

$$\hat{F} = \frac{1}{N} \sum_{j=1}^{N} \delta_{\xi^j}.$$

- **SAA problem.**

$$\hat{z}_{\mathrm{SAA}} = \min_{x \in \mathcal{X}} \ \mathbb{E}_{\hat{F}}[c(x;\xi)] = \min_{x \in \mathcal{X}} \ \frac{1}{N} \sum_{j=1}^{N} c(x;\xi^j).$$

Remark. SAA is asymptotically consistent under mild conditions [6], i.e., $\hat{z}_{\mathrm{SAA}} \to z^\star$ as $N \to \infty$.

Issues of SAA

- **The issue is not inconsistency.** SAA converges asymptotically as $N \to \infty$.
- **The issue is finite-sample instability.** For finite N, the SAA solution can be highly sensitive to data perturbations, especially under heavy-tailed noise.
- **Lack of finite-sample guarantees.** There is generally no clean, non-asymptotic upper bound on the true out-of-sample performance

$$z(x_{\mathrm{SAA}}) = \mathbb{E}_F[c(x_{\mathrm{SAA}}; \xi)].$$

Common workaround: two-sample SAA (2-SAA).

- Split data into optimization and evaluation sets;
- Use classical concentration (e.g., Student-t).

However, this may fail under heavy tails and increases variance [3].

Distribution-first approach

- Construct a data-driven ambiguity set $\mathcal{F}_N$ such that

$$\Pr(F \in \mathcal{F}_N) \geq 1 - \alpha.$$

- Define the robust objective

$$\widehat{z}(x) := \sup_{F_0 \in \mathcal{F}_N} \mathbb{E}_{F_0}[c(x;\xi)].$$

Interpretation. Among all distributions that are statistically plausible given the data, evaluate the decision under the worst-case expected cost.

Key challenge: $\mathcal{F}_N$ lives in the space of probability distributions, so the inner problem is **infinite-dimensional** in general.

Robust SAA: the core objective

Goal. Construct an upper bound $\widehat{z}(x)$ such that, for confidence $1-\alpha$,

$$\Pr\Big(\mathbb{E}_F[c(x;\xi)] \le \widehat{z}(x),\ \forall x \in \mathcal{X}\Big) \ge 1-\alpha.$$

Key idea.

- Not to directly modify the objective;
- But to first construct a *confidence region* for the unknown distribution F.

General structure of the robust problem

- The decision variable in the inner problem is a *distribution*;
- A distribution is a function object (infinite-dimensional);
- The inner problem

$$\sup_{F_0 \in \mathcal{F}_N} \mathbb{E}_{F_0}[c(x;\xi)]$$

is therefore an infinite-dimensional optimization problem.

From GoF confidence regions to tractable formulations

General picture.

- goodness-of-fit (GoF) tests (KS, CvM, AD, χ^2, etc.) induce confidence regions in the space of probability distributions [4, 7];
- These regions define the ambiguity set $\mathcal{F}_N$;
- Under suitable structure, the worst-case expectation can be computed via convex optimization.

From GoF confidence regions to tractable formulations

General picture.

- goodness-of-fit (GoF) tests (KS, CvM, AD, χ^2, etc.) induce confidence regions in the space of probability distributions [4, 7];
- These regions define the ambiguity set $\mathcal{F}_N$;
- Under suitable structure, the worst-case expectation can be computed via convex optimization.

Remark. Depending on the test and cost structure, the resulting formulation may be an LP, SOCP, or other convex program.

A representative case: discrete / scenario-based formulation

Remark. To illustrate tractability, we consider a discrete (or discretized) formulation, which yields a representative finite-dimensional model.

Scenario representation. Let $\{\xi^1, \ldots, \xi^m\}$ be a finite set of scenarios. Any distribution is represented by a probability vector $p \in \Delta_m := \{p \geq 0 : \mathbf{1}^\top p = 1\}$.

Polyhedral ambiguity set (discrete case)

From SAA to Robust SAA. Recall that SAA replaces the true distribution by the empirical distribution constructed from data.

Empirical distribution. Let $\hat{p}_i$ denote the empirical frequency of scenario i.

Robustification via confidence bounds. Instead of trusting the empirical distribution exactly, we construct componentwise confidence bounds

$$\ell_i \leq p_i \leq u_i$$

such that

$$\Pr\big(\ell_i \leq p_i \leq u_i,\ \forall i\big) \geq 1 - \alpha.$$

Resulting ambiguity set.

$$\mathcal{F}_N = \big\{p \in \mathbb{R}^m :\ \ell \leq p \leq u,\ \mathbf{1}^\top p = 1\big\}.$$

Robust SAA: min–sup formulation

For a decision $x \in \mathcal{X}$, define scenario costs

$$c_i(x) := c(x;\xi^i), \qquad c(x) := (c_1(x), \ldots, c_m(x))^\top.$$

$$\widehat{z}(x) = \sup_{p \in \mathcal{F}_N} c(x)^\top p, \qquad \widehat{z}^\star = \min_{x \in \mathcal{X}} \widehat{z}(x).$$

Structure.

- Inner problem: worst-case expectation;
- Outer problem: decision optimization.

Inner problem: a linear program (discrete case)

Fix x. The inner problem is

$$\max_{p\in\mathbb{R}^m} \; c(x)^\top p \quad \text{s.t.} \quad \ell \le p \le u, \; \mathbf{1}^\top p = 1.$$

This is an LP:

- Linear objective;
- Linear constraints.

Interpretation. Find the statistically plausible distribution that maximizes expected cost.

Single-level reformulation via strong duality

By strong duality of linear programming [1], for each fixed x,

$$\sup_{p \in \mathcal{F}_N} c(x)^\top p = \min_{\lambda,\mu \geq 0,\ \nu \in \mathbb{R}} \ \nu + u^\top \lambda - \ell^\top \mu \quad \text{s.t.} \quad \lambda - \mu + \nu \mathbf{1} \geq c(x).$$

Thus the robust SAA problem becomes a single convex optimization problem.

If $\mathcal{X}$ is polyhedral and $c_i(x)$ is linear or piecewise-linear, this formulation reduces to an LP.

Finite-sample validity of robust SAA

Theorem. Finite-sample validity (uniform upper bound)

Let $\xi^1, \ldots, \xi^N \sim F$ be IID samples and let $\mathcal{F}_N = \mathcal{F}_N(\xi^1, \ldots, \xi^N)$ be a (random) ambiguity set constructed from the data such that

$$\Pr\big(F \in \mathcal{F}_N\big) \;\geq\; 1 - \alpha.$$

Define the robust SAA objective for any $x \in \mathcal{X}$ by

$$\hat{z}(x) \;:=\; \sup_{F_0 \in \mathcal{F}_N} \; \mathbb{E}_{F_0}[c(x;\xi)]\,.$$

Then, with probability at least $1 - \alpha$,

$$\sup_{x \in \mathcal{X}} \Big(\mathbb{E}_F[c(x;\xi)] - \hat{z}(x) \Big) \leq 0.$$

Implication. The optimal value of robust SAA is a high-confidence upper bound on the true optimal value of the stochastic program, without data splitting.

Asymptotic behavior of robust SAA

Theorem. Asymptotic consistency

Let $\{\xi^1, \ldots, \xi^N\}$ be IID samples drawn from the true distribution F. Assume that the ambiguity set $\mathcal{F}_N = \mathcal{F}_N(\xi^1, \ldots, \xi^N)$ is constructed such that

$$\Pr(F \in \mathcal{F}_N) \to 1 \quad \text{as } N \to \infty,$$

and that $\mathcal{F}_N$ shrinks to $\{F\}$ in the sense that

$$\sup_{F_0 \in \mathcal{F}_N} \left| \mathbb{E}_{F_0}[c(x;\xi)] - \mathbb{E}_F[c(x;\xi)] \right| \;\to\; 0 \quad \text{for each fixed } x \in \mathcal{X}.$$

Then, for any fixed $x \in \mathcal{X}$, the robust SAA objective

$$\hat{z}(x) := \sup_{F_0 \in \mathcal{F}_N} \mathbb{E}_{F_0}[c(x;\xi)]$$

converges to the true expected cost:

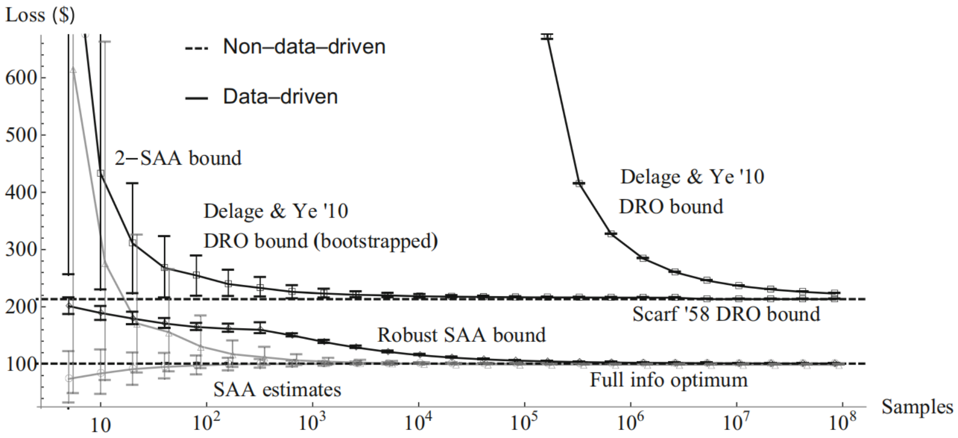
$$\hat{z}(x) \;\to\; \mathbb{E}_F[c(x;\xi)] \quad \text{as } N \to \infty.$$

Consistency properties across statistical tests

Table 1 Summary of convergence results

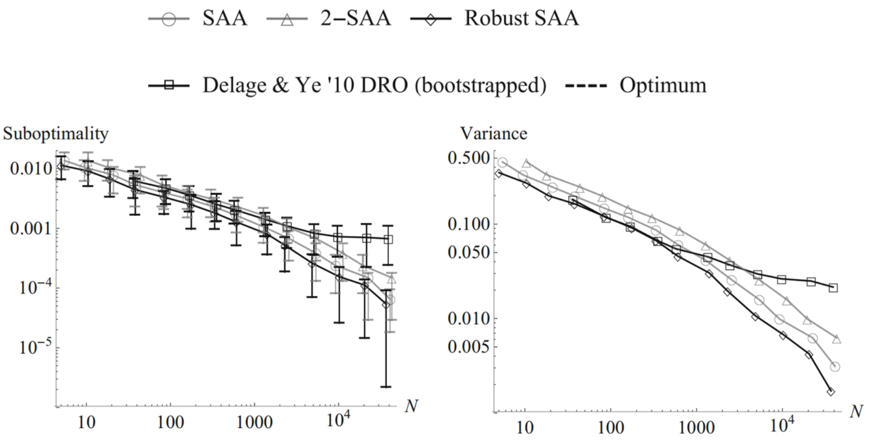
GoF test	Support	Consistent	Uniformly consistent
χ^2 and G-test	Finite	Yes	Yes
KS, Kuiper, CvM, Watson, and AD tests	Univariate	Yes	Yes
Test of marginals using the above tests	Multivariate	No	No
LCX-based test	Multivariate, bounded	Yes	Yes
LCX-based test	Multivariate, unbounded	Yes	?
Tests implied by DUSs of [13, 15]	Multivariate	No	No

Finite-sample behavior: comparison of upper bounds



Real-world application: portfolio allocation

- **Problem.** Portfolio allocation under distributional uncertainty, with a risk-sensitive objective (e.g., CVaR) [5].
- **Result.**



References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[2] Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[3] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall/CRC, New York, 1993.

[4] Erich L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses*. Springer, New York, 3 edition, 2005.

[5] R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–42, 2000.

[6] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 2009.

[7] Michael A. Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737, 1974.