Statistical Inference for Sample Average Approximation of Constrained Optimization and Variational Inequalities

Shu Lu
Department of Statistics and Operations Research
The University of North Carolina at Chapel Hill

The Finance for Energy Market Research Centre
May 19, 2017

Stochastic optimization and sample average approximation

$$\min_{x \in S} E[\Phi(x, \xi)]$$

- $S \subset \mathbb{R}^n$: the feasible set, assumed to be a convex polyhedron
- $\xi(\omega)$: a random vector taking values in a set $\Xi \subset \mathbb{R}^d$
- Φ: a function from $\mathbb{R}^n \times \Xi$ to $\mathbb{R}$

Evaluating $E[\Phi(x, \xi)]$ for a given x is often impractical. A common approach is to solve the sample average approximation (SAA) problem:

$$\min_{x \in S} N^{-1} \sum_{i=1}^{N} \Phi(x, \xi^i(\omega))$$

where $\xi^1, \cdots, \xi^N$ are i.i.d. random variables with distribution same as ξ

Example: Norm-constrained minimum variance portfolio selection

The true problem[1]

$$\min_x \frac{1}{2} x^T \Sigma x \quad s.t. \quad e^T x = 1, \|x\|_1 \leq c$$

The SAA problem:

$$\min_x \frac{1}{2} x^T \Sigma_N x \quad s.t. \quad e^T x = 1, \|x\|_1 \leq c$$

- $x \in \mathbb{R}^p$: portfolio allocations among p assets
- $e \in \mathbb{R}^p$: vector of all one's
- $c \geq 1$: a constant controlling the amount of short sales allowed
- $\Sigma \in \mathbb{R}^{p \times p}$: the true covariance matrix of the random return R
- $\Sigma_N \in \mathbb{R}^{p \times p}$: the sample covariance matrix of the random return, computed from independently and identically distributed sample data $\{r_{ij}\}_{j=1}^p, i = 1, \ldots N$

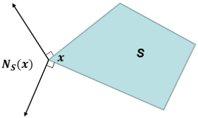
[1] It can be written as $\min_{x \in S} g(E[\Phi(x, R)])$

Stochastic variational inequalities[2] and sample average approximation

$$-E[F(x,\xi)] \in N_S(x) \qquad \text{(TRUE-VI)}$$

- $S \subset \mathbb{R}^n$: the feasible set, assumed to be a convex polyhedron
- ξ: a random vector taking values in a set $\Xi \subset \mathbb{R}^d$
- F: a function from $\mathbb{R}^n \times \Xi$ to $\mathbb{R}^n$
- $N_S(x)$: the normal cone to S at x

$$N_S(x) = \{v \in \mathbb{R}^n \mid \langle v, s-x \rangle \leq 0 \text{ for each } s \in S\}$$



Let $\xi^1, \cdots, \xi^N$ be i.i.d. random variables with distribution same as ξ. The SAA problem is

$$-N^{-1} \sum_{i=1}^{N} F(x, \xi^i(\omega)) \in N_S(x) \qquad \text{(SAA-VI)}$$

[2]See [Chen, Wets and Zhang 2012], [Rockafellar and Wets 2016] for alternative SVI formulations

Stochastic optimization related to stochastic variational inequalities

If the objective function of the stochastic optimization problem

$$\min_{x \in S} E[\Phi(x, \xi)]$$

is differentiable at a local minimizer x_0, then x_0 satisfies the first-order necessary condition

$$-\nabla_x E[\Phi(x_0, \xi)] \in N_S(x_0).$$

The above condition becomes a stochastic variational inequality, when

$$\nabla_x E[\Phi(x_0, \xi)] = E[\nabla_x \Phi(x_0, \xi)]$$

Example: stochastic equilibria in energy markets

- 4 gas producers ($i = 1, \cdots, 4$) decide the amount of gas (x_{ij}^t) to ship to 6 markets ($j = 1, \cdots, 6$) in 4 time periods ($t = 1, \cdots, 4$)
- Each producer i tries to maximize its own profit $E[\Phi_i(x, \xi)]$

$$x_i \in \underset{x_i \in \mathbb{R}_+^{24}}{\operatorname{argmax}} E[\Phi_i(x, \xi)]$$

  - $x_i = [x_{ij}^t]_{tj}$: variables of producer i
  - $x = [x_{ij}^t]_{tij}$: the vector of all variables
  - Φ_i: the profit function of producer i. It depends on x_j for $j \neq i$, since the total amount of production affects gas price
  - $\xi = [\xi^t]_t$: the random oil price

This Cournot-Nash equilibrium problem can be reformulated as a stochastic variational inequality:

$$0 \in -E \begin{bmatrix} \nabla_{x_1} \Phi_1(x, \xi) \\ \vdots \\ \nabla_{x_4} \Phi_4(x, \xi) \end{bmatrix} + N_{\mathbb{R}_+^{96}}(x)$$

The inference question

- In practice, we often solve the SAA problem to find the SAA solution, x_N
- How does data uncertainty affect the reliability of the SAA solution?
- One way to answer this question is by building confidence regions and intervals for the true solution, x_0, based on knowledge about x_N
- An asymptotically exact confidence region $C(x_N)$ is a set in $\mathbb{R}^n$ that depends on x_N and satisfies

$$\lim_{N\to\infty} P\left(x_0 \in C(x_N)\right) = 1 - \alpha$$

- We build confidence regions and intervals by utilizing the asymptotic distribution of SAA solutions

The normal map formulation of variational inequalities[3]

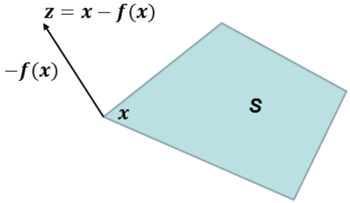
The normal map associated with a function $f : S \to \mathbb{R}^n$ and a set $S \subset \mathbb{R}^n$ is a function $f_S : \mathbb{R}^n \to \mathbb{R}^n$, defined as

$$f_S(z) = f(\Pi_S(z)) + z - \Pi_S(z) \text{ for each } z \in \mathbb{R}^n$$

where $\Pi_S(z)$ is the Euclidean projection of z on S

$$\boxed{-f(x) \in N_S(x)} \quad \underset{x=\Pi_S(z)}{\overset{z=x-f(x)}{\rightleftarrows}} \quad \boxed{f_S(z) \quad = \quad 0}$$

- Π_S is piecewise affine
- The normal manifold of S: the polyhedral subdivision of $\mathbb{R}^n$ corresponding to Π_S
- f_S is piecewise smooth if f is smooth, and is piecewise affine if f is affine



[3]Details about normal maps can be found in [Robinson 1992], [Ralph 1993], [Facchinei and Pang 2003], [Scholtes 2012] and references therein

The true problems and SAA problems

- Define the true function as $f_0(x) = E[F(x, \xi)]$ and the SAA function

$$f_N(x) = -N^{-1} \sum_{i=1}^{N} F(x, \xi^i(\omega))$$

- Write (TRUE-VI) as

$$-f_0(x) \in N_S(x)$$

and (SAA-VI) as

$$-f_N(x) \in N_S(x)$$

- Their corresponding normal map formulations are

$$(f_0)_S(z) = 0 \quad \text{(SVI-NM)} \qquad \text{and} \qquad (f_N)_S(z) = 0 \quad \text{(SAA-NM)}$$

- Let $z_0 = x_0 - f_0(x_0)$ and $z_N = x_N - f_N(x_N)$ be solutions to the normal map formulations

Convergence of SAA solutions to the true solution[4]

Under certain ▸ Assumptions

- For a.e. ω, (SAA-VI) has a locally unique solution x_N for N large enough, with $\lim_{N\to\infty} x_N = x_0$
- The corresponding solution z_N to (SAA-NM) is also locally unique, with $\lim_{N\to\infty} z_N = z_0$ almost surely
- Let Σ_0 be the covariance matrix of $F(x_0,\xi)$, and $\mathcal{N}(0,\Sigma_0)$ be a normal r.v. in $\mathbb{R}^n$ with zero mean and covariance matrix Σ_0. Then,

$$\sqrt{N} L_K(z_N - z_0) \Rightarrow \mathcal{N}(0,\Sigma_0) \qquad \text{(Conv-Dist-}z\text{)}$$

$$\sqrt{N}(x_N - x_0) \Rightarrow \Pi_K \circ (L_K)^{-1}(\mathcal{N}(0,\Sigma_0)) \qquad \text{(Conv-Dist-}x\text{)}$$

  where $L = \nabla_x E[F(x_0,\xi)]$, $K = T_S(x_0) \cap E[F(x_0,\xi)]^{\perp}$, L_K is the normal map associated with L and K, and $(L_K)^{-1}$ is its inverse
- L_K is a piecewise linear approximation of the normal map $(f_0)_S$ around z_0

[4] See related results in [Dupacova and Wets 1988], [King and Rockafellar 1993], [Gürkan, Özge and Robinson 1999], [Demir 2000], [Shapiro, Dentcheva and Ruszczyński 2009], [Gürkan and Pang 2009], [Xu 2010] etc.

Assumptions

Assumption 1: Implies the continuous differentiability of f_0 on O, the almost sure convergence $f_N \to f_0$ as an element of $C^1(X, \mathbb{R}^n)$ for any compact set $X \subset O$, and the weak convergence of $\sqrt{N}(f_N - f_0)$

(a) $E\|F(x,\xi)\|^2 < \infty$ for all $x \in O$, where O is an open set in $\mathbb{R}^n$.
(b) The map $x \mapsto F(x, \xi(\omega))$ is cont diff on O for a.e. $\omega \in \Omega$.
(c) There exists a square integrable random variable C such that $\|F(x,\xi(\omega)) - F(x',\xi(\omega))\| + \|dF(x,\xi(\omega)) - dF(x',\xi(\omega))\| \leq C(\omega)\|x - x'\|$, for all $x', x \in O$ and a.e. $\omega \in \Omega$.

Assumption 2: Guarantees the existence, local uniqueness, and stability of the true solution under small perturbation of f_0

Suppose that $x_0 \in O$ solves (SVI). Let $z_0 = x_0 - f_0(x_0)$, $L = df_0(x_0)$, $K = T_S(x_0) \cap \{z_0 - x_0\}^{\perp}$, and assume that the normal map L_K induced by L and K is a homeomorphism from $\mathbb{R}^n$ to $\mathbb{R}^n$

Properties of the limiting distributions

$$\underbrace{(L_K)^{-1}(\mathcal{N}(0,\Sigma_0))}_{\text{limiting distribution of } \sqrt{N}(z_N - z_0)} \quad \text{and} \quad \underbrace{\Pi_K \circ (L_K)^{-1}(\mathcal{N}(0,\Sigma_0))}_{\text{limiting distribution of } \sqrt{N}(x_N - x_0)}$$

- For a given $q \in \mathbb{R}^n$, $\Pi_K \circ (L_K)^{-1}(q)$ is the solution h of a linear VI:

$$-Lh + q \in N_K(h)$$

  and when L is symmetric it is the unique solution of the QP

$$\min_{h \in K} \frac{1}{2} h^T L h - q^T h$$

- $(L_K)^{-1}(q) = h - Lh + q$
- If K is a subspace, $\Pi_K \circ (L_K)^{-1}(q)$ and $(L_K)^{-1}(q)$ are linear functions of q, and x_N and z_N are asymptotically normal
- If K is a polyhedral convex cone but not a subspace, then $\Pi_K \circ (L_K)^{-1}(q)$ and $(L_K)^{-1}(q)$ are piecewise linear functions with multiple pieces, and x_N and z_N are not asymptotically normal

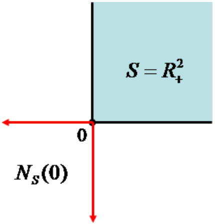
Example: a linear complementarity problem

- $F : \mathbb{R}^2 \times \mathbb{R}^6 \to \mathbb{R}^2$ given by $F(x,\xi) = \begin{bmatrix} \xi_1 & \xi_2 \\ \xi_3 & \xi_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \xi_5 \\ \xi_6 \end{bmatrix}$
- ξ uniformly distributed on $[0,2] \times [0,1] \times [0,2] \times [0,4] \times [-1,1] \times [-1,1]$
- Then $f_0(x) = E[F(x,\xi)] = \begin{bmatrix} 1 & 1/2 \\ 1 & 2 \end{bmatrix} x$

Let $S = \mathbb{R}^2_+$. The SVI is an LCP:

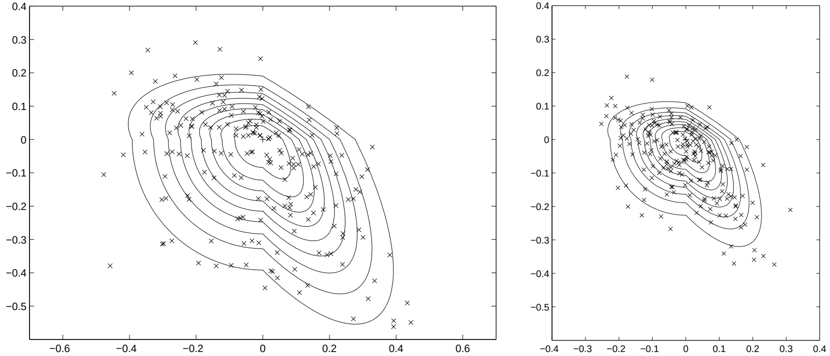
$$-f_0(x) \in N_{\mathbb{R}^2_+}(x),$$

which has a unique solution $x_0 = 0$, and $z_0 = x_0 - f_0(x_0) = 0$ is the unique solution for (SVI-NM)



Here, $L = \begin{bmatrix} 1 & 1/2 \\ 1 & 2 \end{bmatrix}$ and $K = S = \mathbb{R}^2_+$

In the example: scatter plots for z_N



- Left: solutions to 200 SAA problems with $N = 10$; Right: $N = 30$
- Curves are boundaries of sets

$$\{z \in \mathbb{R}^2 \mid N\big[L_K(z - z_0)\big]^T \Sigma_0^{-1}\big[L_K(z - z_0)\big] \leq \chi^2_N(\alpha)\}$$

which contain z_N with (approximately) probability $1 - \alpha$ for $\alpha = 0.1, \cdots, 0.9$

A computable, asymptotically exact confidence region for z_0

- From $\sqrt{N} L_K(z_N - z_0) \Rightarrow \mathcal{N}(0, \Sigma_0)$, an asymptotically exact $(1-\alpha)100\%$ confidence region for z_0[5] is

$$\left\{ z \in \mathbb{R}^n \mid N \left[L_K(z_N - z) \right]^T \Sigma_0^{-1} \left[L_K(z_N - z) \right] \le \chi_n^2(\alpha) \right\} \qquad \text{(CR0)}$$

- However, (CR0) is not computable as Σ_0 and L_K are unknown
- Interestingly, an asymptotically exact, and computable, confidence region is given by

$$\left\{ z \in \mathbb{R}^n \mid N \left[d(f_N)_S(z_N)(z - z_N) \right]^T \Sigma_N^{-1} \left[d(f_N)_S(z_N)(z - z_N) \right] \le \chi_n^2(\alpha) \right\} \qquad \text{(CR1)}$$

  - $d(f_N)_S(z_N)(z - z_N)$: the directional derivative of $(f_N)_S$ at z_N for the direction $z - z_N$
  - Σ_N: the sample covariance matrix of $F(x_N, \xi)$[6]

- With high probability, $d(f_N)_S(z_N)$ is linear and (CR1) is an ▸ ellipsoid

[5] $\chi_n^2(\alpha)$ satisfies $P(U > \chi_n^2(\alpha)) = \alpha$ for a χ^2 r.v. U with n deg of freedom
[6] Different from Σ_N in the portfolio selection example

In the example: Confidence regions for z_0 computed from z_{10}

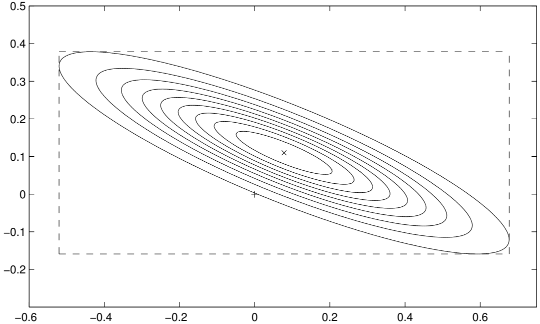
An SAA for the LCP (N=10): $$-\begin{bmatrix} 0.93 & 0.54 \\ 0.75 & 2.11 \end{bmatrix} x + \begin{bmatrix} 0.13 \\ 0.29 \end{bmatrix} \in N_{\mathbb{R}^2_+}(x)$$

A unique solution $x_{10} = (0.08, 0.11) = z_{10}$ marked as $\times$ $\quad +$: $z_0 = 0$

$$\underbrace{\begin{bmatrix} 0.42 & 0.01 \\ 0.01 & 0.19 \end{bmatrix}}_{\Sigma_{10}} \quad \underbrace{\begin{bmatrix} 0.93 & 0.54 \\ 0.75 & 2.11 \end{bmatrix}}_{d(f_{10})_{\mathbb{R}^2_+}(z_{10})} \quad \underbrace{\left\{ z \,\middle|\, 10(z - z_{10})^T \begin{bmatrix} 4.88 & 9.34 \\ 9.34 & 24.26 \end{bmatrix} (z - z_{10}) \le \chi^2_2(\alpha) \right\}}_{(1-\alpha)100\% \text{ confidence region for } z_0}$$

Shown in the figure: confidence regions for z_0 at levels $0.1, \cdots, 0.9$

90% simultaneous confidence intervals:
$(z_0)_1$: [-0.52, 0.68]
$(z_0)_2$: [-0.16, 0.38]
$(x_0)_1$: [0, 0.68]
$(x_0)_2$: [0, 0.38]



Individual confidence intervals for z_0 and x_0 (target level: 90%)

- Consider cells in the normal manifold of $\mathbb{R}^2_+$:
  $\{0\}, \{0\} \times \mathbb{R}_+, \mathbb{R}_+ \times \{0\}, \{0\} \times \mathbb{R}_-, \mathbb{R}_- \times \{0\}, \mathbb{R}^2_+, \mathbb{R}^2_-, \mathbb{R}_+ \times \mathbb{R}_-, \mathbb{R}_- \times \mathbb{R}_+,$
- C_{i_N}: the cell with the smallest dimension, among all cells that intersect the 95% region. Here it is $\{0\}$
- P_N: the 2-dim cell that contains z_N in its interior. Here it is $\mathbb{R}^2_+$
- Let $\tilde{z}_{i_N}$ be any point in ri C_{i_N}, and $K_N = \text{cone}(P_N - \tilde{z}_{i_N})$. Here it is $\mathbb{R}^2_+$
- With limiting probability $\geq 95\%$, K_N gives the cone that contains z_N in the polyhedral subdivision of $\mathbb{R}^2$ corresponding to L_K
- Let $M = (d(f_N)_S(z_N))^{-1}\Sigma_N^{1/2}$, and compute a number ℓ_N such that

$$\frac{\Pr\left(|(MZ)_j| \leq \ell_N, \ \text{ and } \ MZ \in K_N\right)}{\Pr\left(MZ \in K_N\ \right)} = 0.95, \quad \text{where } Z \sim \mathcal{N}(0, I)$$

- $\liminf_{N\to\infty} \text{Prob}\left(\sqrt{N}|(z_N - z_0)_j| \leq \ell_N\right) \geq 0.90$
- 90% individual confidence intervals for z_0 and x_0 (computation of intervals for x_0 is analogous)
  $(z_0)_1$: [-0.16, 0.32], $(z_0)_2$: [0, 0.22], $(x_0)_1$: [0, 0.32], $(x_0)_2$: [0, 0.22]

An alternative method

- Under additional assumptions, z_N converges to z_0 at an exponential rate
- At z_N we can define a function $\Phi_N(z_N) : \mathbb{R}^n \times \mathbb{R}^n$ so that

$$\lim_{N\to\infty} \text{Prob}\left[\sup_{h\in\mathbb{R}^n} \frac{\|\Phi_N(z_N)(h) - L_K(h)\|}{\|h\|} < \frac{\phi}{N^{1/3}}\right] = 1$$

- Replacing L_K by $\Phi_N(z_N)$ in the weak convergence results gives a different method for computing confidence regions and intervals

Portfolio selection example: Confidence intervals and coverage rates

- $\mathcal{A} = \{j : (x_0)_j \neq 0\}$
- 200 replications
- Avgcov: average coverage; Medcov: median coverage
- Avglen: average length

$1-\alpha = 90\%$		Our method			Normal estimation		
N and p		Avgcov $\mathcal{A}$	Medcov $\mathcal{A}$	Avglen $\mathcal{A}$	Avgcov $\mathcal{A}$	Medcov $\mathcal{A}$	Avglen $\mathcal{A}$
N=200	p=30	0.937	0.94	0.163	0.887	0.89	0.13
N=500	p=30	0.94	0.94	0.101	0.883	0.88	0.085
N=600	p=100	0.896	0.89	0.125	0.862	0.85	0.104
N=1000	p=100	0.928	0.94	0.099	0.89	0.89	0.083
$1-\alpha = 95\%$		Our method			Normal estimation		
N and p		Avgcov $\mathcal{A}$	Medcov $\mathcal{A}$	Avglen $\mathcal{A}$	Avgcov $\mathcal{A}$	Medcov $\mathcal{A}$	Avglen $\mathcal{A}$
N=200	p=30	0.965	0.965	0.210	0.94	0.94	0.183
N=500	p=30	0.97	0.96	0.129	0.945	0.91	0.112
N=600	p=100	0.97	0.96	0.083	0.945	0.89	0.073
N=1000	p=100	0.972	0.97	0.065	0.945	0.935	0.057

Energy market equilibrium example: Coverage rates ($\alpha = 0.05$)

- $\upsilon_j^{.05}$: Normal estimation
- $\tilde{h}_j^{.04}$: The presented method with $\alpha_1 = 0.01$, $\alpha_2 = 0.04$
- $\tilde{h}_j^{.025}$: The presented method with $\alpha_1 = 0.025$, $\alpha_2 = 0.025$
- 2000 replications

Percentile	$N = 200$			$N = 2,000$		
	$\upsilon_j^{.05}$	$\tilde{h}_j^{.04}$	$\tilde{h}_j^{.025}$	$\upsilon_j^{.05}$	$\tilde{h}_j^{.04}$	$\tilde{h}_j^{.025}$
MIN	88.20 %	88.70%	89.05 %	94.60 %	95.70 %	97.20 %
Q1	94.75 %	95.70 %	97.08%	94.85 %	96.00 %	97.50 %
MEDIAN	94.90 %	95.83 %	97.45%	95.30 %	96.25 %	97.95 %
Q3	95.05%	95.95 %	97.60%	95.40 %	96.5 %	98.35 %
MAX	100 %	100 %	100 %	100 %	100 %	100 %

Summary

- Development and justification of methods to build computable confidence regions and intervals for the true solutions of the expected-value formulation of stochastic variational inequalities
- Applied to stochastic Cournot-Nash production/distribution problems, sparse-penalized statistical regression and portfolio selection

This presentation is mainly based on the following papers:

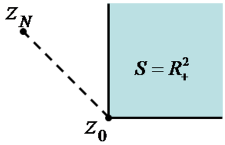
- Lamm, Lu. 2016. Generalized conditioning based approaches to compute confidence intervals for stochastic variational inequalities. Submitted
- Lamm, Lu and Budhiraja. 2016. Individual confidence intervals for solutions to expected value formulations of stochastic variational inequalities. *Mathematical Programming B*
- Lu. 2014. Symmetric confidence regions and confidence intervals for normal map formulations of stochastic variational inequalities. *SIAM Journal on Optimization*. Vol. 24, No. 3, pp. 1458-1484
- Lu and Budhiraja. Confidence regions for stochastic variational inequalities. *Mathematics of Operations Research*, 2013, Vol. 38, No. 3, pp. 545-568

Derivation of (CR1)

- A key observation: z_N in a neighborhood of z_0 satisfies

$$d\Pi_S(z_0)(z_N - z_0) + d\Pi_S(z_N)(z_0 - z_N) = 0$$

where $d\Pi_S(z_0)(z_N - z_0)$ is the directional derivative of Π_S at z_0 for the direction $z_N - z_0$



- This property holds, as long as z_0 and z_N are contained in a common n-cell

- With $\sqrt{N} L_K(z_N - z_0) \Rightarrow \mathcal{N}(0, \Sigma_0)$ and $L_K = d(f_0)_S(z_0)$, it can be shown

$$-\sqrt{N} d(f_N)_S(z_N)(z_0 - z_N) \Rightarrow \mathcal{N}(0, \Sigma_0)$$

which implies (CR1) is an asymptotically exact confidence region for z_0