Statistical inference for sample average approximation of constrained optimization and variational inequalities

Shu Lu

Collaborators: Amarjit Budhiraja, Michael Lamm, Hongsheng Liu, Yufeng Liu, Liang Yin, Guan Yu, Kai Zhang

Department of Statistics and Operations Research
The University of North Carolina at Chapel Hill
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 \( \Xi \subset \mathbb{R}^d \)
- \( \Phi \): 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 \( \xi^1, \ldots, \xi^N \) are i.i.d. random variables with distribution same as \( \xi \)
Example: Norm-constrained minimum variance portfolio selection

The true problem:

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

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 returns
- $\Sigma_n \in \mathbb{R}^{p \times p}$: the sample covariance matrix of the random returns, computed from independently and identically distributed sample data $\{r_{ij}\}_{j=1}^p, i = 1, \ldots, n$
Stochastic variational inequalities\(^1\) and sample average approximation

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

- \(S \subset \mathbb{R}^n\): the feasible set, assumed to be a convex polyhedron
- \(\xi\): 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 \(\xi\). The SAA problem is

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

\(^1\)See [Chen, Wets and Zhang 2012], [Rockafellar and Wets 2016] for alternative SVI formulations
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 \(E[\Phi_i(x, \xi)]\)

\[
x_i \in \operatorname{argmax} E[\Phi_i(x, \xi)]
\]

- \(x_i = [x_{ij}^t]_{tij}\): variables of producer \(i\)
- \(x = [x_{ij}^t]_{tij}\): the vector of all variables
- \(\Phi_i\): 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

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

- We build confidence regions and intervals by utilizing the asymptotic distribution of SAA solutions
The normal map formulation of variational inequalities\(^2\)

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

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

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

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

\(^2\)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 \( f_N(x) = -\frac{1}{N-1} \sum_{i=1}^{N} F(x, \xi^i(\omega)) \)

- 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 \((f_0)_S(z) = 0 \quad (SVI-NM)\) and \((f_N)_S(z) = 0 \quad (SAA-NM)\)

- 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

Statistical inference for sample average approximation of constrained optimization and variational inequalities
Convergence of SAA solutions to the true solution

Under certain assumptions

- For a.e. $\omega$, (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 $\Sigma_0$ 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 $\Sigma_0$. Then,

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

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

  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 $z_0$

---

3See 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

\[(L_K)^{-1}(N(0, \Sigma_0))\] and \[\Pi_K \circ (L_K)^{-1}(N(0, \Sigma_0))\]

limiting distribution of \(\sqrt{N}(z_N - z_0)\)

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

Statistical inference for sample average approximation of constrained optimization and variational inequalities
Example: a linear complementarity problem

- \( F : \mathbb{R}^2 \times \mathbb{R}^6 \rightarrow \mathbb{R}^2 \) given by
  \[
  F(x, \xi) = \begin{bmatrix}
    \xi_1 & \xi_2 \\
    \xi_3 & \xi_4
  \end{bmatrix}
  \begin{bmatrix}
    x_1 \\
    x_2
  \end{bmatrix}
  + \begin{bmatrix}
    \xi_5 \\
    \xi_6
  \end{bmatrix}
  \]

- \( \xi \) 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

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

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

Statistical inference for sample average approximation of constrained optimization and variational inequalities
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 $z_0^4$ is

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

- However, (CR0) is not computable as $\Sigma_0$ and $L_K$ are unknown

- Interestingly, an asymptotically exact, and computable, confidence region is given by

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

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

- With high probability, $d(f_N)_S(z_N)$ is linear and (CR1) is an $\chi^2$ ellipsoid

$\chi^2_n(\alpha)$ satisfies $P(U > \chi^2_n(\alpha)) = \alpha$ for a $\chi^2$ r.v. $U$ with $n$ deg of freedom
In the example: Confidence regions for $z_0$ computed from $z_{10}$

An SAA for the LCP ($N=10$):

$[-0.93 \ 0.54] \times [0.13] \in N_{\mathbb{R}^2_+}(x)$

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

$\Sigma_{10} = \begin{bmatrix} 0.42 & 0.01 \\ 0.01 & 0.19 \end{bmatrix}$  
$d(f_{10})_{\mathbb{R}^2_+}(z_{10}) = \begin{bmatrix} 0.93 & 0.54 \\ 0.75 & 2.11 \end{bmatrix}$

$\left\{ z \mid 10(z - z_{10})^T \begin{bmatrix} 4.88 & 9.34 \\ 9.34 & 24.26 \end{bmatrix} (z - z_{10}) \leq \chi^2_2(\alpha) \right\}^{(1-\alpha)100\%}$ 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:

$(z_0)_1: [-0.52, 0.68]$
$(z_0)_2: [-0.16, 0.38]$
$(x_0)_1: [0, 0.68]$
$(x_0)_2: [0, 0.38]$

Statistical inference for sample average approximation of constrained optimization and variational inequalities
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 $r_i C_{i_N}$, 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 $M = \left((d(f_N)_{S}(z_N))^{-1}\Sigma_N^{1/2}\right)$, and compute a number $\ell_N$ such that

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

- $\lim inf_{N \to \infty} \Pr \left( \sqrt{N}|(z_N - z_0)_j| \leq \ell_N \right) \geq 0.90$

- 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; Medlen: median length

<table>
<thead>
<tr>
<th></th>
<th>Our method</th>
<th>Normal estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1 - \alpha = 90%$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n and p</td>
<td>Avgcov $\mathcal{A}$</td>
<td>Medcov $\mathcal{A}$</td>
</tr>
<tr>
<td>n=200 p=30</td>
<td>0.937</td>
<td>0.94</td>
</tr>
<tr>
<td>n=500 p=30</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>n=600 p=100</td>
<td>0.896</td>
<td>0.89</td>
</tr>
<tr>
<td>n=1000 p=100</td>
<td>0.928</td>
<td>0.94</td>
</tr>
<tr>
<td>$1 - \alpha = 95%$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n and p</td>
<td>Avgcov $\mathcal{A}$</td>
<td>Medcov $\mathcal{A}$</td>
</tr>
<tr>
<td>n=200 p=30</td>
<td>0.965</td>
<td>0.965</td>
</tr>
<tr>
<td>n=500 p=30</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>n=600 p=100</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>n=1000 p=100</td>
<td>0.972</td>
<td>0.97</td>
</tr>
</tbody>
</table>
Energy market equilibrium example: Coverage rates ($\alpha = 0.05$)

- $\nu_j^{05}$: 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

<table>
<thead>
<tr>
<th>Percentile</th>
<th>$N = 200$</th>
<th>\hspace{1cm}</th>
<th>$N = 2,000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\nu_j^{05}$</td>
<td>$h_j^{04}$</td>
<td>$h_j^{025}$</td>
</tr>
<tr>
<td>MIN</td>
<td>88.20 %</td>
<td>88.70 %</td>
<td>89.05 %</td>
</tr>
<tr>
<td>Q1</td>
<td>94.75 %</td>
<td>95.70 %</td>
<td>97.08%</td>
</tr>
<tr>
<td>MEDIAN</td>
<td>94.90 %</td>
<td>95.83 %</td>
<td>97.45%</td>
</tr>
<tr>
<td>Q3</td>
<td>95.05%</td>
<td>95.95 %</td>
<td>97.60%</td>
</tr>
<tr>
<td>MAX</td>
<td>100 %</td>
<td>100 %</td>
<td>100 %</td>
</tr>
</tbody>
</table>

Statistical inference for sample average approximation of constrained optimization and variational inequalities
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.

This presentation is mainly based on the following papers:

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 $\Pi_S$ 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 $z_0$. 

Statistical inference for sample average approximation of constrained optimization and variational inequalities