Confidence regions and confidence intervals for stochastic variational inequalities via the normal map approach

Shu Lu

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

Based on joint work with Amarjit Budhiraja and Michael Lamm

Outline

Outline

The SVI problem

- $(\Omega, \mathcal{F}, P)$: a probability space
- ξ: a random vector from Ω to a closed set $\Xi \subset \mathbb{R}^d$
- Consider[1] a function $F : \mathbb{R}^n \times \Xi \to \mathbb{R}^n$, with $E\|F(x,\xi)\| < \infty$ for each x
- Let $f_0(x) = E[F(x,\xi)]$ for each x
- $S = \{x \in \mathbb{R}^n \mid Ax \leq b\} = \{x \in \mathbb{R}^n \mid \langle a_i, x\rangle \leq b_i, i = 1, \cdots, m\}$

The SVI problem we study is to find $x \in S$ such that

$$-f_0(x) \in N_S(x) \qquad \text{(TRUE-VI)}$$

where $N_S(x)$ is the normal cone to S at x and is defined as

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

[1] In papers we consider $F : O \times \Xi \to \mathbb{R}^n$ with O being an open set in $\mathbb{R}^n$

Relation with stochastic optimization

A local minimizer x_0 of the problem

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

with $\Phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ and $S \subset \mathbb{R}^n$ needs to satisfy

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

if $E[\nabla_x \Phi(x_0, \xi)] = \nabla_x E[\Phi(x_0, \xi)]$ exists, where ∇_x denotes the gradient w.r.t x

Example: a Nash equilibrium problem

- m players of the game
- x_i: decision variable for player i, which takes values in a set K_i
- θ_i: the profit of player i, a random function of $x = (x_1, \cdots, x_m)$
- Each player selects x_i to maximize the expected profit $E(\theta_i)$
- An equilibrium is attained, when no player can increase his expected profit by unilaterally changing his decision

Under mild conditions, the equilibrium problem can be formulated as

$$0 \in -E \begin{bmatrix} \nabla_{x_1}\theta_1(x) \\ \vdots \\ \nabla_{x_m}\theta_m(x) \end{bmatrix} + N_{K_1 \times \cdots \times K_N}(x)$$

Example: a linear complementarity problem

Let $F : \mathbb{R}^2 \times \mathbb{R}^6 \to \mathbb{R}^2$ be defined by

$$F(x,\xi) = \begin{bmatrix} \xi_1 & \xi_2 \\ \xi_3 & \xi_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \xi_5 \\ \xi_6 \end{bmatrix},$$

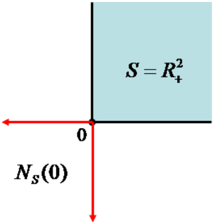
where ξ is a random vector uniformed distributed on

$$\{\xi \in \mathbb{R}^6 \mid (0,0,0,0,-1,-1) \le \xi \le (2,1,2,4,1,1)\}$$

Let $S = \mathbb{R}^2_+$. The SVI problem becomes the following LCP:

$$-\begin{bmatrix} 1 & 1/2 \\ 1 & 2 \end{bmatrix} x \in N_{\mathbb{R}^2_+}(x),$$

which has a unique solution $x_0 = 0$



The sample average approximation problem

- In most problems of interest, f_0 does not have a closed form expression and requires a numerical approximation
- Let $\xi^1, \cdots, \xi^N$ be i.i.d. random variables with distribution same as ξ
- Define $f_N : \mathbb{R}^n \times \Omega \to \mathbb{R}^n$ by $f_N(x, \omega) = N^{-1} \sum_{i=1}^N F(x, \xi^i(\omega))$
- The SAA problem is to find $x \in S$ such that

$$-f_N(x, \omega) \in N_S(x) \qquad \text{(SAA-VI)}$$

For the LCP example, an SAA problem with $N = 10$ is given by

$$-\begin{bmatrix} 0.9292 & 0.5400 \\ 0.7536 & 2.1111 \end{bmatrix} x + \begin{bmatrix} 0.1319 \\ 0.2906 \end{bmatrix} \in N_{\mathbb{R}^2_+}(x),$$

which has a unique solution $x_{10} = (0.0782, 0.1097)$

From SAA solutions to the true solution

Under certain conditions, solutions to the SAA problems[2]

- Almost surely converge to the solution of the true problem as $N \uparrow \infty$
- Follow certain asymptotic distribution around the true solution
- Converge to the true solution in probability at an exponential rate

An expression for confidence regions of the true solution is readily obtainable from the asymptotic distribution of SAA solutions. However, it is not directly computable

We consider the normal map transformation of variational inequalities, and propose methods to build asymptotically exact confidence regions/intervals for the true solution of the transformed problem, that are computable from SAA solutions

[2] See [King and Rockafellar 1993], [Gürkan, Özge and Robinson 1999], [Demir 2000], [Shapiro, Dentcheva and Ruszczyński 2009], [Xu 2010] and references therein

Outline

The normal map formulation of variational inequalities[3]

The normal map induced by the function f_0 and the set S is a function $(f_0)_S : \mathbb{R}^n \to \mathbb{R}^n$, defined as

$$(f_0)_S(z) = f_0(\Pi_S(z)) + z - \Pi_S(z) \text{ for each } z \in \mathbb{R}^n$$

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

$$\boxed{\begin{array}{c} -f_0(x) \in N_S(x) \\ z = x - f_0(x) \end{array}} \iff \boxed{\begin{array}{lcl} (f_0)_S(z) & = & 0 \qquad \text{(TRUE-NM)} \\ x = \Pi_S(z) & & \end{array}}$$

In the LCP example, $x_0 = 0$ is the unique solution for (TRUE-VI), so $z_0 = x_0 - f_0(x_0) = 0$ is the unique solution for (TRUE-NM)

$$\boxed{\begin{array}{c} -f_N(x) \in N_S(x) \\ z = x - f_N(x) \end{array}} \iff \boxed{\begin{array}{lcl} (f_N)_S(z) & = & 0 \qquad \text{(SAA-NM)} \\ x = \Pi_S(z) & & \end{array}}$$

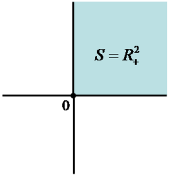
---

[3] For more on normal maps and normal manifolds, see [Robinson 1992], [Ralph 1993], [Facchinei and Pang 2003], [Scholtes 2012] and references therein

Two properties of the Euclidean projector

- Π_S is piecewise affine: it coincides with an affine function on each of finitely many full-dim polyhedrons whose union is $\mathbb{R}^n$
- Each such polyhedron is called an n-cell in the normal manifold of S; a k-dimensional face of it is a k-cell. The relative interiors of all cells form a partition of $\mathbb{R}^n$
- Π_S is B-differentiable: at each $z \in \mathbb{R}^n$, it has a B-derivative $d\Pi_S(z)$, which is a positively homogeneous function from $\mathbb{R}^n$ to $\mathbb{R}^n$ that approximates Π_S near z
- For all points z in the relative interior of a cell, the B-derivative $d\Pi_S(z)$ is the same; it changes abruptly across cells

Example: $S = \mathbb{R}^2_+$. Its normal manifold contains the four orthants, on each of which Π_S coincides with a distinct affine function. The B-derivative $d\Pi_S(z)$ is the identity map at $z \in \operatorname{int} S$, and is piecewise linear with four pieces at $z = 0$



Assumptions[5]

Assumption 1: [4] Implies the continuous differentiability of f_0, the almost sure convergence $f_N \to f_0$ as an element of $C^1(X, \mathbb{R}^n)$ for any compact set $X \subset \mathbb{R}^n$, and the weak convergence of $\sqrt{N}(f_N - f_0)$

(a) $E\|F(x,\xi)\|^2 < \infty$ for all $x \in \mathbb{R}^n$.
(b) The map $x \mapsto F(x, \xi(\omega))$ is continuously differentiable 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 $x', x \in \mathbb{R}^n$ 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 solves (TRUE-VI). 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 $\mathbb{R}^n$

[4] In papers the assumptions are limited to x or x' in a set $O \subset \mathbb{R}^n$

[5] Implications of the assumptions are obtained using tools from [Shapiro, Dentcheva and Ruszczyński 2009], [Robinson 1995] and a functional CLT

B-derivatives of the normal map

- Recall that the normal map $(f_0)_S$ is defined as

$$(f_0)_S(z) = f_0(\Pi_S(z)) + z - \Pi_S(z) \text{ for each } z \in \mathbb{R}^n$$

- Under Assumption 1, f_0 is continuously differentiable on $\mathbb{R}^n$
- By the chain rule, $(f_0)_S$ has a B-derivative at each $z \in \mathbb{R}^n$, with

$$d(f_0)_S(z)(h) = df_0(\Pi_S(z))(d\Pi_S(z)(h)) + h - d\Pi_S(z)(h) \text{ for each } h \in \mathbb{R}^n$$

- Discontinuity of $d\Pi_S(z)$ leads to discontinuity of $d(f_0)_S(z)$: when z moves from the interior of an n-cell to its boundary, $d(f_0)_S(z)$ changes abruptly
- By the definitions of L and K in Assumption 2, we have $d(f_0)_S(z_0) = L_K$

Asymptotic distribution of SAA solutions

Under Assumptions 1 and 2:

- For a.e. ω, (SAA-NM) has a locally unique solution z_N for N large enough, with $\lim_{N\to\infty} z_N = z_0$
- Accordingly, (SAA-VI) has a locally unique solution $x_N = \Pi_S(z_N)$ that almost surely converges to x_0

[6] This assumption can be relaxed for confidence regions

[7] $\chi_n^2(\alpha)$ satisfies $P(U > \chi_n^2(\alpha)) = \alpha$ for a χ^2 r.v. U with n deg of freedom

Asymptotic distribution of SAA solutions

Under Assumptions 1 and 2:

- For a.e. ω, (SAA-NM) has a locally unique solution z_N for N large enough, with $\lim_{N\to\infty} z_N = z_0$
- Accordingly, (SAA-VI) has a locally unique solution $x_N = \Pi_S(z_N)$ that almost surely converges to x_0
- Let Σ_0 be the covariance matrix of $F(x_0, \xi)$, and Y_0 be a normal r.v. in $\mathbb{R}^n$ with zero mean and covariance matrix Σ_0. Then,

$$\sqrt{N} d(f_0)_S(z_0)(z_N - z_0) \Rightarrow Y_0 \qquad \text{(Conv-Dist)}$$

- Assuming Σ_0 to be nonsingular[6], the following set

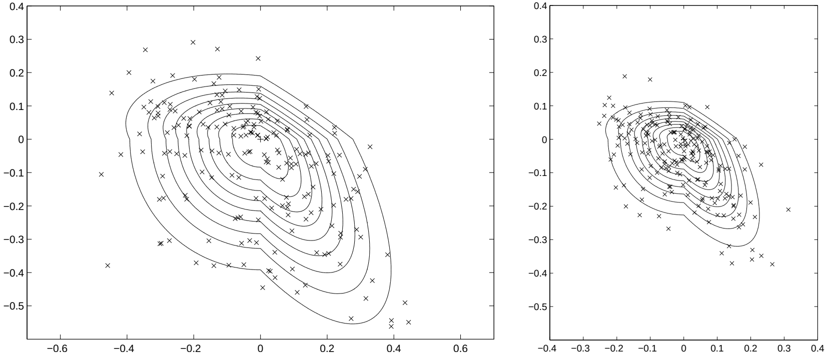
$$\{z \in \mathbb{R}^n \mid N\left[d(f_0)_S(z_0)(z_N - z)\right]^T \Sigma_0^{-1}\left[d(f_0)_S(z_0)(z_N - z)\right] \leq \chi_n^2(\alpha)\} \qquad \text{(CR0)}$$

defines an asymptotically exact $(1-\alpha)100\%$ confidence region for z_0[7]

[6] This assumption can be relaxed for confidence regions

[7] $\chi_n^2(\alpha)$ satisfies $P(U > \chi_n^2(\alpha)) = \alpha$ for a χ^2 r.v. U with n deg of freedom

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

$$\{z \in \mathbb{R}^2 \mid N\left[d(f_0)_{\mathbb{R}^2_+}(z_0)(z - z_0)\right]^T \Sigma_0^{-1}\left[d(f_0)_{\mathbb{R}^2_+}(z_0)(z - z_0)\right] \leq \chi^2_N(\alpha)\}$$

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

The challenge with computing confidence regions

- The expression (CR0) is not directly computable as Σ_0 and $d(f_0)_S(z_0)$ are unknown
- We can approximate Σ_0 by Σ_N, the sample covariance matrix of $\{F(x_N, \xi^i)\}_{i=1}^N$, which converges to Σ_0 almost surely
- Recall from the chain rule

$$d(f_0)_S(z_0)(h) = df_0(x_0)(d\Pi_S(z_0)(h)) + h - d\Pi_S(z_0)(h) \text{ for each } h \in \mathbb{R}^n$$

- We can approximate $df_0(x_0)$ by $df_N(x_N)$, which converges to $df_0(x_0)$ almost surely
- What to replace $d\Pi_S(z_0)$ with?

$d\Pi_S(z_N)$ does NOT converge to $d\Pi_S(z_0)$ in prob due to the discontinuity

Estimate $d\Pi_S(z_0)$ using exponential convergence rate: 1

Under the additional Assumption 3 below, there exist positive real numbers ϵ_0, β_0, μ_0, M_0 and σ_0, such that

$$\text{Prob}\left\{\|z_N - z_0\| < \epsilon\right\} \geq 1 - \beta_0 \exp\{-N\mu_0\} - \frac{M_0}{\epsilon^n} \exp\left\{-\frac{N\epsilon^2}{\sigma_0}\right\} \quad \text{(Conv-Prob)}$$

for each $\epsilon \in (0, \epsilon_0]$ and each N.

(a) For each $t \in \mathbb{R}^n$ and $x \in X$, let $M_x(t) = E\left[\exp\{\langle t, F(x,\xi) - f_0(x)\rangle\}\right]$ be the moment generating function of $F(x,\xi) - f_0(x)$. Assume

1. There exists $\zeta > 0$ such that $M_x(t) \leq \exp\{\zeta^2\|t\|^2/2\}$ for each (x, t).
2. There exists a nonnegative random variable κ such that $\|F(x,\xi(\omega)) - F(x',\xi(\omega))\| \leq \kappa(\omega)\|x - x'\|$ for all $x, x' \in O$ and almost every $\omega \in \Omega$.
3. The moment generating function of κ is finite valued in a neighborhood of zero.

(b) Similar conditions on $d_x F$.

Estimate $d\Pi_S(z_0)$ using exponential convergence rate: 2

- Method 1: substitute $d\Pi_S(z_0)$ by $\Lambda_N(z_N)$, defined as a weighted sum of all possible $d\Pi_S(z)$ for $z \in \mathbb{R}^n$ (there are only finitely many), with weights depending on z_N [8]
- Method 2: $\Lambda_N(z_N)$ is chosen as $d\Pi_S(z')$, where z' is a point in the relative interior of a cell that has the smallest dimension among all cells within a distance of $N^{-1/3}$ from z_N [9]

[8] Methods 1 and 2 are from [Lu and Budhiraja, 2013] and [Lu 2012] respectively

[9] The quantity $N^{-1/3}$ can be generalized as $1/g(N)$ with $g : \mathbb{N} \to \mathbb{R}_+$ satisfying $\lim_{N\to\infty} g(N) = \lim_{N\to\infty} \frac{N}{g(N)^2} = \infty$ and other conditions

Estimate $d\Pi_S(z_0)$ using exponential convergence rate: 2

- Method 1: substitute $d\Pi_S(z_0)$ by $\Lambda_N(z_N)$, defined as a weighted sum of all possible $d\Pi_S(z)$ for $z \in \mathbb{R}^n$ (there are only finitely many), with weights depending on z_N [8]
- Method 2: $\Lambda_N(z_N)$ is chosen as $d\Pi_S(z')$, where z' is a point in the relative interior of a cell that has the smallest dimension among all cells within a distance of $N^{-1/3}$ from z_N [9]
- $\Lambda_N(z_N)$ defined in either method 1 or method 2 is shown to be a good estimator of $d\Pi_S(z_0)$: there exists $\kappa > 0$ such that

$$\lim_{N\to\infty} \text{Prob}\left[\sup_{h\in\mathbb{R}^n} \frac{\|\Lambda_N(z_N)(h) - d\Pi_S(z_0)(h)\|}{\|h\|} < \frac{\kappa}{g(N)}\right] = 1$$

(For method 2, κ is actually 0.)

[8] Methods 1 and 2 are from [Lu and Budhiraja, 2013] and [Lu 2012] respectively

[9] The quantity $N^{-1/3}$ can be generalized as $1/g(N)$ with $g : \mathbb{N} \to \mathbb{R}_+$ satisfying $\lim_{N\to\infty} g(N) = \lim_{N\to\infty} \frac{N}{g(N)^2} = \infty$ and other conditions

Confidence regions based on methods 1 and 2

- For each $N \in \mathbb{N}$, define a function $\Phi_N(z_N) : \mathbb{R}^n \to \mathbb{R}^n$ by[10]

$$\Phi_N(z_N)(h) = df_N(\Pi_S(z_N))(\Lambda_N(z_N)(h)) + h - \Lambda_N(z_N)(h)$$

- $\Phi_N(z_N)$ is then a good approximation of $d(f_0)_S(z_0)$ and can be used as its substitute in (Conv-Dist) to get

$$\sqrt{N}\Phi_N(z_N)(z_N - z_0) \Rightarrow Y_0$$

- As a result, we have

$$\sqrt{N}\Sigma_N^{-1/2}\Phi_N(z_N)(z_N - z_0) \Rightarrow N(0, I_n)$$

- For each $\alpha \in (0,1)$ the following set defines an asymptotically exact $(1-\alpha)100\%$ confidence region for z_0

$$\{z \in \mathbb{R}^n \mid N\left[\Phi_N(z_N)(z_N - z)\right]^T \Sigma_N^{-1}\left[\Phi_N(z_N)(z_N - z)\right] \leq \chi_n^2(\alpha)\} \quad \text{(CR-1\&2)}$$

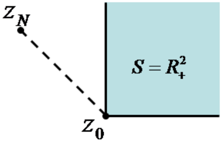
[10] We suppressed ω in the definition here

Confidence regions by the third method[11]

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

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

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

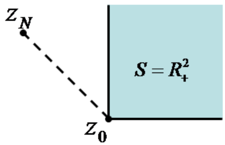


[11] Method 3 is from [Lu 2014]

Confidence regions by the third method[11]

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

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



- This property holds, as long as z_0 and z_N are contained in a common n-cell
- The consequence: $-\sqrt{N} d(f_N)_S(z_N)(z_0 - z_N) \Rightarrow Y_0$
- An asymptotically exact $(1-\alpha)100\%$ confidence region for z_0 (an ellipsoid when $d(f_N)_S(z_N)$ is an invertible linear map, which occurs with high probability):

$$\{z \in \mathbb{R}^n \mid N\left[d(f_N)_S(z_N)(z - z_N)\right]^T \Sigma_N^{-1}\left[d(f_N)_S(z_N)(z - z_N)\right] \leq \chi_n^2(\alpha)\} \quad \text{(CR-3)}$$

- This method does not need the exponential convergence rate of z_N

[11] Method 3 is from [Lu 2014]

Outline

Computation of simultaneous confidence intervals

- We compute simultaneous confidence intervals by finding edges of the minimum bounding box of the confidence region

$$\{z \in \mathbb{R}^n \mid N\big[\Phi_N(z_N)(z_N - z)\big]^T \Sigma_N^{-1} \big[\Phi_N(z_N)(z_N - z)\big] \leq \chi_n^2(\alpha)\} \quad \text{(CR-1\&2)}$$

$$\{z \in \mathbb{R}^n \mid N\big[d(f_N)_S(z_N)(z - z_N)\big]^T \Sigma_N^{-1} \big[d(f_N)_S(z_N)(z - z_N)\big] \leq \chi_n^2(\alpha)\} \quad \text{(CR-3)}$$

- If $d(f_0)_S(z_0)$ is piecewise linear, then its estimator $\Phi_N(z_N)$ is with high probability a piecewise linear function, in which case (CR-1&2) is the union of fractions of ellipses
- Even if $d(f_0)_S(z_0)$ is piecewise linear, $d(f_N)_S(z_N)$ is with high probability a linear function, with (CR-3) being a single ellipsoid
- In general, it is much more efficient to use (CR-3). If $d(f_0)_S(z_0)$ is linear, all three methods coincide for large N

Computation of individual confidence intervals: approach 1

- We present three approaches to compute individual confidence intervals that contain individual components with prescribed levels
- Recall from (Conv-Dist) that $\sqrt{N}(z_N - z_0) \Rightarrow d(f_0)_S(z_0)^{-1}(Y_0)$, where $Y_0 \sim \mathcal{N}(0, \Sigma_0)$
- Approach 1: use the following interval

$$(z_N)_i \pm \chi_1^2(\alpha)\sqrt{(d(f_N)_S(z_N)^{-1}\Sigma_N d(f_N)_S(z_N)^{-T})_{jj}}/\sqrt{N}$$

  as the $(1-\alpha)100\%$ confidence interval of $(z_0)_i$
  - Advantage: easy to compute, supported by current numerical tests, does not need the exponential convergence rate
  - Disadvantage: asymptotical exactness of the interval depends on a restrictive assumption
- The other two approaches use the estimator $\Phi_N(z_N)$ of $d(f_0)_S(z_0)$, obtained under the exponential convergence rate

Computation of individual confidence intervals: approach 2

- Recall that $\sqrt{N}(z_N - z_0) \Rightarrow d(f_0)_S(z_0)^{-1}(Y_0)$, where $Y_0 \sim \mathcal{N}(0, \Sigma_0)$
- Approach 2: let $Z \sim \mathcal{N}(0, I_n)$; find a number a such that[12]

$$\Pr\left(\left| \left(\Phi_N(z_N)^{-1}(\Sigma_N^{1/2} Z) \right)_j \right| \leq a \right) = 1 - \alpha,$$

and use $[(z_N)_j - aN^{-1/2}, (z_N)_j + aN^{-1/2}]$ as the $(1-\alpha)100\%$ confidence interval of $(z_0)_j$[13]

  - Advantage: asymptotical exactness of the interval is justified for general situations
  - Disadvantage: when $\Phi_N(z_N)$ is piecewise linear, computing a requires enumerating all of its pieces

[12] Approaches 2 and 3 are from [Lamm, Lu and Budhiraja, 2014]

[13] In the implementation we allow for a choice in where the interval is centered

Computation of individual confidence intervals: approach 3

- Recall $\sqrt{N}(z_N - z_0) \Rightarrow d(f_0)_S(z_0)^{-1}(Y_0)$; let $\{K_1, \ldots, K_l\}$ be a family of n-dim polyhedral convex cones whose union is $\mathbb{R}^n$, such that $d(f_0)_S(z_0)$ is represented by a different linear map on each K_i
- Given z_N, we can identify a cone $K(\omega)$ such that $K(\omega)$ belongs to $\{K_1, \ldots, K_l\}$ and contains $z_N - z_0$ in its interior with high probability

Computation of individual confidence intervals: approach 3

- Recall $\sqrt{N}(z_N - z_0) \Rightarrow d(f_0)_S(z_0)^{-1}(Y_0)$; let $\{K_1, \ldots, K_l\}$ be a family of n-dim polyhedral convex cones whose union is $\mathbb{R}^n$, such that $d(f_0)_S(z_0)$ is represented by a different linear map on each K_i
- Given z_N, we can identify a cone $K(\omega)$ such that $K(\omega)$ belongs to $\{K_1, \ldots, K_l\}$ and contains $z_N - z_0$ in its interior with high probability
- Next, find a number a such that the conditional probability

$$\frac{\Pr\left(\left| \left(\Phi_N(z_N)^{-1}(\Sigma_N^{1/2} Z) \right)_j \right| \leq a, \Phi_N(z_N)^{-1}(\Sigma_N^{1/2} Z) \in K(\omega) \right)}{\Pr\left(\Phi_N(z_N)^{-1}(\Sigma_N^{1/2} Z) \in K(\omega) \right)} = 1 - \alpha,$$

  and use $[(z_N)_j - aN^{-1/2}, (z_N)_j + aN^{-1/2}]$ as the $(1-\alpha)100\%$ confidence interval for $(z_0)_j$
- This approach focuses on a single $K(\omega)$ and avoids enumerating all pieces of $\Phi_N(z_N)$, is much more efficient than approach 2

Confidence regions/intervals for x_0

- Once a confidence region for z_0 is obtained, its projection onto the set S gives a (conservative) confidence region for x_0, since $x_0 = \Pi_S(z_0)$
- When S is a box, such a projection method is easy to implement to compute both simultaneous and individual confidence intervals for x_0
- When S is not a box, then the above projection method is hard to implement in general, and one needs to investigate the special structures in application problems to transform confidence regions/intervals for z_0 into those of x_0
  - For example, if each component of x_0 depends on only one component of z_0 then this transformation could be easy to do
- We provide a method to directly compute individual confidence intervals of components of x in [Lamm et al., 2014]

Outline

Estimators for the example

- SAA solutions: $x_{10} = (0.0782, 0.1097)$ and $z_{10} = (0.0782, 0.1097)$
- In method 1, $\Phi_{10}(z_{10})(\cdot)$ is a piecewise linear map represented by matrices

$$\begin{bmatrix} 0.9437 & 0.4333 \\ 0.5989 & 1.8915 \end{bmatrix}, \begin{bmatrix} 0.9437 & 0.1322 \\ 0.5989 & 1.2720 \end{bmatrix}, \begin{bmatrix} 0.9839 & 0.4333 \\ 0.1710 & 1.8915 \end{bmatrix}, \begin{bmatrix} 0.9839 & 0.1322 \\ 0.1710 & 1.2720 \end{bmatrix}$$

in orthants $\mathbb{R}^2_+$, $\mathbb{R}_+ \times \mathbb{R}_-$, $\mathbb{R}_- \times \mathbb{R}_+$ and $\mathbb{R}^2_-$ respectively

- In method 2, $\Phi_{10}(z_{10})(\cdot)$ is represented by

$$\begin{bmatrix} 0.9292 & 0.5400 \\ 0.7536 & 2.1111 \end{bmatrix}, \begin{bmatrix} 0.9292 & 0 \\ 0.7536 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.5400 \\ 0 & 2.1111 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

in corresponding orthants

- In method 3, the B-derivative $d(f_{10})_{\mathbb{R}^2_+}(z_{10})$ is the linear map represented by the matrix

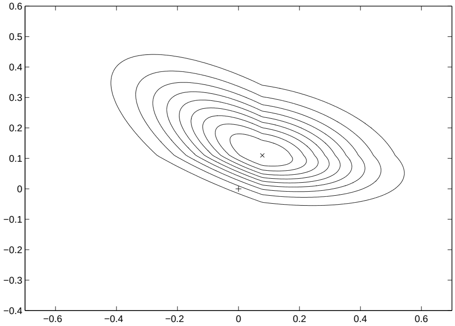
$$\begin{bmatrix} 0.9292 & 0.5400 \\ 0.7536 & 2.1111 \end{bmatrix}$$

Confidence regions for z_0 computed from z_{10}

$\times$: $z_{10} = (0.0782, 0.1097)$

$+$: $z_0 = 0$

From the innermost to outermost, the curves enclose confidence regions at levels $0.1, \cdots, 0.9$, computed from (CR-1&2) or (CR-3) with $N = 10$



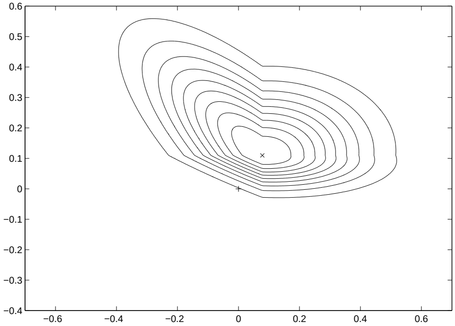
Confidence regions for z_0, $N = 10$, method 1

Confidence regions for z_0 computed from z_{10}

$\times$: $z_{10} = (0.0782, 0.1097)$

$+$: $z_0 = 0$

From the innermost to outermost, the curves enclose confidence regions at levels $0.1, \cdots, 0.9$, computed from (CR-1&2) or (CR-3) with $N = 10$



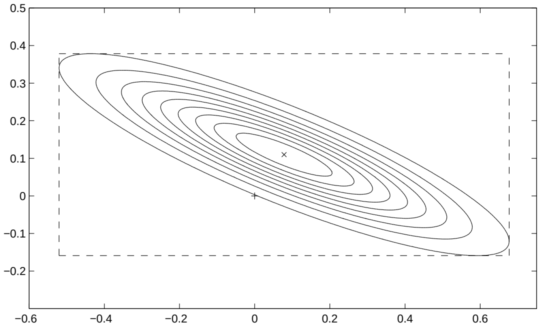
Confidence regions for z_0, $N = 10$, method 2

Confidence regions for z_0 computed from z_{10}

$\times$: $z_{10} = (0.0782, 0.1097)$

$+$: $z_0 = 0$

From the innermost to outermost, the curves enclose confidence regions at levels $0.1, \cdots, 0.9$, computed from (CR-1&2) or (CR-3) with $N = 10$



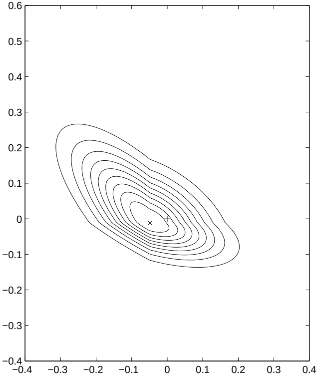
Confidence regions for z_0, $N = 10$, method 3

Confidence regions for z_0 computed from z_{30}

$\times$: $z_{30} = (-0.048, -0.011)$

$+$: $z_0 = 0$

From the innermost to outermost, the curves enclose confidence regions at levels $0.1, \cdots, 0.9$, computed from (CR-1&2) or (CR-3) with $N = 30$



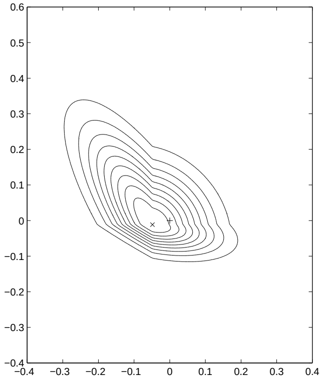
Confidence regions for z_0, $N = 30$, method 1

Confidence regions for z_0 computed from z_{30}

$\times$: $z_{30} = (-0.048, -0.011)$

$+$: $z_0 = 0$

From the innermost to outermost, the curves enclose confidence regions at levels $0.1, \cdots, 0.9$, computed from (CR-1&2) or (CR-3) with $N = 30$



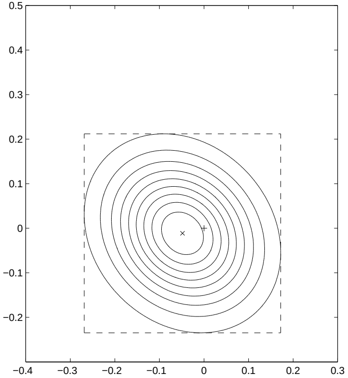
Confidence regions for z_0, $N = 30$, method 2

Confidence regions for z_0 computed from z_{30}

$\times$:
$z_{30} = (-0.048, -0.011)$

$+$: $z_0 = 0$

From the innermost to outermost, the curves enclose confidence regions at levels $0.1, \cdots, 0.9$, computed from (CR-1&2) or (CR-3) with $N = 30$



Confidence regions for z_0, $N = 30$, method 2

Confidence intervals

	z_{10}			z_{30}		
	Est	Sim CI	Ind CI	Est	Sim CI	Ind CI
$(z_0)_1$	0.08	[-0.52, 0.68]	[-0.38, 0.54]	-0.05	[-0.27, 0.17]	[-0.21, 0.12]
$(z_0)_2$	0.11	[-0.16, 0.38]	[-0.10, 0.32]	-0.01	[-0.23, 0.21]	[-0.18, 0.16]

Sim/ind confidence intervals for z_0 of level 90%, method 3

	$N = 10$			$N = 30$		
	$\alpha =$0.1	0.05	0.01	$\alpha =$0.1	0.05	0.01
Simultaneously for z_0	171	180	187	184	192	197

Coverage of sim confidence intervals from 200 SAA problems, method 3

Individual confidence intervals

Coverage of $(z_0)_1$ (left) and $(z_0)_2$ (right), $\alpha = .05$, 200 SAA problems

Approach	1	2	3
N=50	91.5%	92.5%	92.5%
N=100	93.5%	94.5%	94%
N=200	97%	97%	97.5%
N=2,000	94.5%	94.5%	94.5%

Approach	1	2	3
N=50	94%	96.5%	94.5%
N=100	93.5%	96.5%	92.5%
N=200	96.5%	98%	97%
N=2,000	93.5%	95.5 %	94.5%

Coverage of $(z_0)_2$ and half-width by cone, $N = 2,000$, $\alpha = .05$

Cone (samples in cone)	Coverage rate			Average half-width		
	1	2	3	1	2	3
$\mathbb{R}_- \times \mathbb{R}_-$(44)	95.45 %	97.73%	95.45%	.0253	.0246	.0253
$\mathbb{R}_- \times \mathbb{R}_+$ (59)	88.14%	98.31%	89.83%	.0127	.0246	.0133
$\mathbb{R}_+ \times \mathbb{R}_-$ (77)	94.81%	90.91%	96.10%	.0358	.0246	.0379
$\mathbb{R}_+ \times \mathbb{R}_+$ (20)	100 %	100%	100%	.0239	.0246	.0106

Evaluation of normality by χ^2 plots (N=10)

(CR-1&2) are based on the fact

$$\sqrt{N}\Sigma_N^{-1/2}\Phi_N(z_N)(z_N - z_0) \Rightarrow N(0, I_n)$$

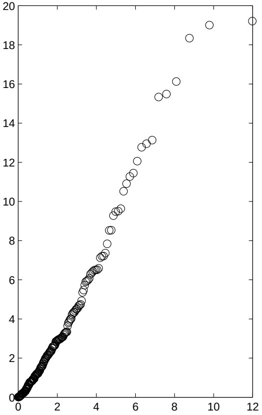
(CR-3) is based on the fact

$$-\sqrt{N}\Sigma_N^{-1/2}d(f_N)_S(z_N)(z_0 - z_N) \Rightarrow N(0, I_n)$$

Horizontal axis: $100(j - 1/2)/200$ quantiles of the χ^2 distribution with 2 degrees of freedom, $j = 1, \cdots, 200$

Vertical axis: 2-norms of the above random vectors for 200 SAA solutions with $N = 10$, ordered from smallest to largest

Slope > 1: the distances are too big



χ^2 plots, $N = 10$, method 1

Evaluation of normality by χ^2 plots (N=10)

(CR-1&2) are based on the fact

$$\sqrt{N}\Sigma_N^{-1/2}\Phi_N(z_N)(z_N - z_0) \Rightarrow N(0, I_n)$$

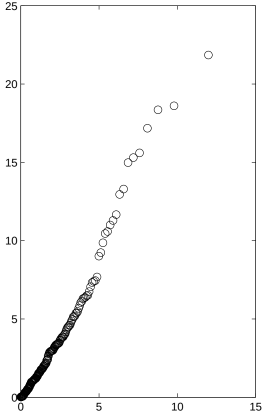
(CR-3) is based on the fact

$$-\sqrt{N}\Sigma_N^{-1/2}d(f_N)_S(z_N)(z_0 - z_N) \Rightarrow N(0, I_n)$$

Horizontal axis: $100(j - 1/2)/200$ quantiles of the χ^2 distribution with 2 degrees of freedom, $j = 1, \cdots, 200$

Vertical axis: 2-norms of the above random vectors for 200 SAA solutions with $N = 10$, ordered from smallest to largest

Slope > 1: the distances are too big



χ^2 plots, $N = 10$, method 2

Evaluation of normality by χ^2 plots (N=10)

(CR-1&2) are based on the fact

$$\sqrt{N}\Sigma_N^{-1/2}\Phi_N(z_N)(z_N - z_0) \Rightarrow N(0, I_n)$$

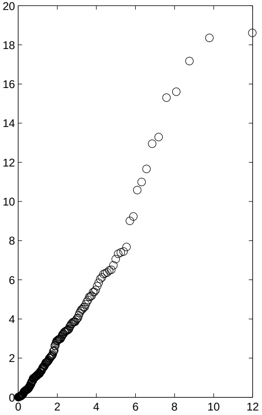
(CR-3) is based on the fact

$$-\sqrt{N}\Sigma_N^{-1/2} d(f_N)_S(z_N)(z_0 - z_N) \Rightarrow N(0, I_n)$$

Horizontal axis: $100(j - 1/2)/200$ quantiles of the χ^2 distribution with 2 degrees of freedom, $j = 1, \cdots, 200$

Vertical axis: 2-norms of the above random vectors for 200 SAA solutions with $N = 10$, ordered from smallest to largest

Slope > 1: the distances are too big



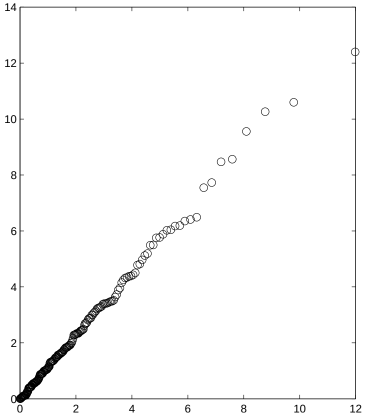
χ^2 plots, $N = 10$, method 3

Evaluation of normality by χ^2 plots (N=30)

Horizontal axis: the quantiles

Vertical axis: squared distances for 200 SAA problems with $N = 30$, ordered from smallest to largest

Slope ≈ 1: indicates near normality



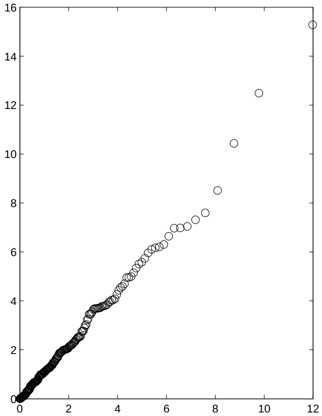
χ^2 plots, $N = 30$, method 1

Evaluation of normality by χ^2 plots (N=30)

Horizontal axis: the quantiles

Vertical axis: squared distances for 200 SAA problems with $N = 30$, ordered from smallest to largest

Slope $\approx$ 1: indicates near normality



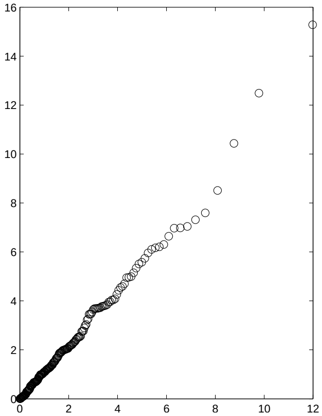
χ^2 plots, $N = 30$, method 2

Evaluation of normality by χ^2 plots (N=30)

Horizontal axis: the quantiles

Vertical axis: squared distances for 200 SAA problems with $N = 30$, ordered from smallest to largest

Slope $\approx$ 1: indicates near normality



χ^2 plots, $N = 30$, method 3

Outline

Summary

- Development and justification of methods to build asymptotically exact and computable confidence regions for the true solution of the normal map formulation of the SVI
- Derivation of formulas and methods to compute simultaneous and individual confidence intervals
- Applied to a stochastic Cournot-Nash production/distribution problem and statistical learning problems (the lasso and others)

Major assumptions:

- Continuous differentiability of $F(x, \xi)$ w.r.t. x
- Local uniqueness of the solution to the true problem

References

This presentation is based on the following papers:

- Lamm, Lu and Budhiraja. 2014. Individual confidence intervals for true solutions to stochastic variational inequalities. Submitted for publication
- Lu. 2014. Symmetric confidence regions and confidence intervals for normal map formulations of stochastic variational inequalities. SIAM Journal on Optimization. Forthcoming
- Lu. 2012. A new method to build confidence regions for solutions of stochastic variational inequalities. *Optimization*, published online at DOI10.1080/02331934.2012.727556
- Lu and Budhiraja. Confidence regions for stochastic variational inequalities. *Mathematics of Operations Research*, 2013, Vol. 38, No. 3, pp. 545-568

Thank you!

Estimate $d\Pi_S(z_0)$ using exponential convergence rate

- Write the k-cells as $C_1^k, \cdots, C_{j(k)}^k$. Define $d_i^k(z) = \min_{x \in C_i^k} \|z - x\|$
- Define $g : \mathbb{N} \to \mathbb{R}_+$, to satisfy

$$\lim_{N\to\infty} g(N) = \infty, \ \lim_{N\to\infty} \frac{N}{g(N)^2} = \infty, \text{ and other conditions}$$

- (Conv-Prob) implies the following for each N:

$$\text{Prob}\left\{\|z_N - z_0\| < \frac{1}{2g(N)}\right\} \geq 1 - \beta_0 \exp\{-N\mu_0\} - 2^n M_0 g(N)^n \exp\left\{-\frac{N}{4\sigma_0 g(N)^2}\right\}$$

- With high probability, for N large enough,

$$d_i^k(z_N) \begin{cases} > 1/g(N) & \text{if } C_i^k \text{ does not contain } z_0 \\ < 1/(2g(N)) & \text{if } C_i^k \text{ contains } z_0 \end{cases}$$

- Suppose that $z_0 \in \text{ri } C_{i(q)}^{k(q)}$; then $C_{i(q)}^{k(q)}$ has the smallest dimension among all cells containing z_0

Approximate $d\Pi_S(z_0)$: Single out $C_{i(q)}^{k(q)}$

- Write $\Psi_i^k = d\Pi_S(z)$ for $z \in \mathrm{ri}\, C_i^k$
- For each $N \in \mathbb{N}$ and $z \in \mathbb{R}^n$, define a function $\Lambda_N(z) : \mathbb{R}^n \to \mathbb{R}^n$
- Definition 1 of $\Lambda_N(z)$: A weighted sum of Ψ_i^k

$$\Lambda_N(z)(h) = \frac{\sum_{k=0}^{n} \sum_{i=1}^{j(k)} \left[1/g(N) - \min(d_i^k(z), 1/g(N))\right]^k \Psi_i^k(h)}{\sum_{k=0}^{n} \sum_{i=1}^{j(k)} \left[1/g(N) - \min(d_i^k(z), 1/g(N))\right]^k}$$

As $N \to \infty$ the weight of $\Psi_{i(q)}^{k(q)}$ in $\Lambda_N(z_N)$ becomes dominant

- Definition 2 of $\Lambda_N(z)$: Pick a single cell

$\Lambda_N(z)(h) = \Psi_{i_0}^{k_0}(h)$ where $C_{i_0}^{k_0}$ is a cell that has the smallest dimension among all cells within a distance of $1/g(N)$ from z