CLT and Bootstrap in High Dimensions

V. Chernozhukov (MIT) D. Chetverikov (UCLA) K. Kato (Tokyo)

October 27, 2016

Introduction

This presentation is based on:

1. “Central Limit Theorems and Bootstrap in High Dimensions," *ArXiv*, 2014, *Ann. Prob.*, 2016.

2. “Gaussian Approximation of Suprema of Empirical Processes," *Ann. Stat.*, 2014a

3. “Uniform Inference after Selection for LAD and Other Z-Estimation Problmes" *Biometrika*, 2014b (with A. Belloni). ArXiv 2013.

4. “Uniformly Valid Post-Regularization Confidence Regions for Many Functional Parameters in Z-Estimation Framework" (with A. Belloni and Y. Wei). ArXiv, 2015.

Introduction

Let $X_1, \ldots, X_n$ be a sequence of *centered* independent random vectors in $\mathbb{R}^p$, with each X_i having coordinates denoted by X_{ij}; that is,

$$X_i = (X_{ij})_{j=1}^p.$$

Define the normalized sum:

$$S_n^X := (S_{nj}^X)_{j=1}^p := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i. \tag{1}$$

Let $Y_1, \ldots, Y_n$ be independent Gaussian random vectors in $\mathbb{R}^p$:

$$Y_i \sim N(0, \mathrm{E}[X_i X_i']).$$

Define the Gaussian analog of S_n^X as:

$$S_n^Y := (S_{nj}^Y)_{j=1}^p := \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i. \tag{2}$$

Introduction

Define the Kolmorogorov distance between S_n^X and S_n^Y:

$$\rho_n := \sup_{A \in \mathcal{A}} \left| \mathrm{P}(S_n^X \in A) - \mathrm{P}(S_n^Y \in A) \right|$$

where $\mathcal{A}$ is some class of sets

Question: how fast can $p = p_n$ grow as $n \to \infty$ under the restriction that $\rho_n \to 0$?

Bentkus (2003): for i.i.d. X_i, if $\mathcal{A}$ is the class of all convex sets, then

$$\rho_n = O\left(\frac{p^{1/4} \mathrm{E}[\|X\|_2^3]}{\sqrt{n}} \right)$$

Typically $\mathrm{E}[\|X\|_2^3] = O(p^{3/2})$, so

$$\rho_n \to 0 \quad \text{if} \quad p = o(n^{2/7})$$

Nagaev (1976): this result is nearly optimal, $\rho_n \gtrsim \mathrm{E}[\|X\|_2^3]/\sqrt{n}$

Introduction

However, in modern statistics, often $p \gg n$

- high dimensional regression models
- multiple hypothesis testing problems

Question: can we find a non-trivial class of sets $\mathcal{A}$ such that

$$p = p_n \gg n \quad \text{but} \quad \rho_n \to 0$$

Our first main result(s):

Subject to some conditions, if $\mathcal{A}$ is the class of all rectangles (or sparsely convex sets), then

$$\rho_n \to 0 \quad \text{if} \quad \log p = o(n^{1/7})$$

Simulation Example

The example is motivated by the problem of removing the Gaussianity assumption in Dantzig/Lasso estimators of (very) high-dimensional sparse regression models. Let

$$S_n^X = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i, \quad X_{ij} = z_{ij}\varepsilon_i, \;\; \varepsilon_i \text{ i.i.d. } t(5)/c$$

z_{ij}'s are fixed bounded "regressors", $|z_{ij}| \leq B$, drawn from $U(0,1)$ distribution once, and

$$S_n^Y = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i, \quad Y_{ij} = z_{ij}e_i, \;\; e_i \text{ i.i.d. } N(0,1),$$

so that $\mathrm{E}[Y_i Y_i'] = \mathrm{E}[X_i X_i']$. Compare

$$\mathrm{P}\left(\|S_n^X\|_\infty \leq t\right) \text{ and } \mathrm{P}\left(\|S_n^Y\|_\infty \leq t\right).$$

(i.e. ρ_n for $\mathcal{A}$ = cubes in $\mathbb{R}^p$)

Simulation Example

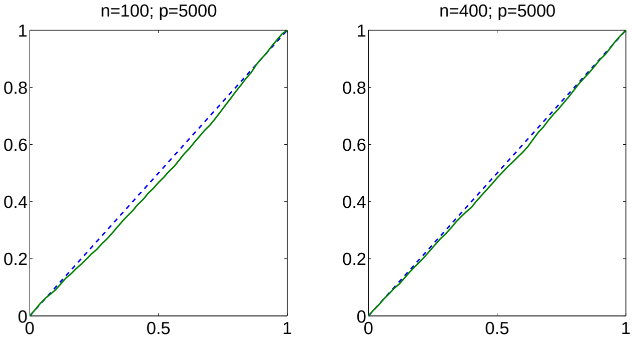

Figure: P-P plots comparing $\mathrm{P}\left(\|S_n^Y\|_\infty \leq t\right)$ and $\mathrm{P}\left(\|S_n^X\|_\infty \leq t\right)$. The dashed line is the 45° line.

Introduction – Bootstrap

Generally, $\mathrm{P}(S_n^Y \in A)$ is unknown since don't know covariance matrix $\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_i X_i']$. So the second result, is that under similar conditions

$$\rho_n^* = \sup_{A \in \mathcal{A}} \left| \mathrm{P}(S_n^{X*} \in A \mid \{X_i\}_{i=1}^n) - \mathrm{P}(S_n^Y \in A) \right| \to_{\mathrm{P}} 0$$

We prove this result for the Gaussian Bootstrap (*multiplier method* with Gaussian multipliers):

$$S_n^{X*} := \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \bar{X}) e_i, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \tag{3}$$

where $(e_i)_{i=1}^n$ are i.i.d. $N(0,1)$ multipliers; and the Empirical Bootstrap:

$$S_n^{X*} := \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \bar{X}) m_{i,n}, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \tag{4}$$

where $(m_{i,n})_{i=1}^n$ is n-dimensional multinomial variate based on n trials with success probabilities $1/n, \ldots, 1/n$.

Conditions

Let $b > 0$ and $q \geq 4$ be constants, and $(B_n)_{n=1}^{\infty}$ be a sequence of positive constants, possibly growing to ∞.

Consider the following conditions:

(M.1) $n^{-1} \sum_{i=1}^{n} \mathrm{E}[X_{ij}^2] \geq b$ *for all* $j = 1, \ldots, p$,

(M.2) $n^{-1} \sum_{i=1}^{n} \mathrm{E}[|X_{ij}|^{2+k}] \leq B_n^k$ *for all* $j = 1, \ldots, p$ *and* $k = 1, 2$.

and one of the following:

(E.1) $\mathrm{E}[\exp(|X_{ij}|/B_n)] \leq 2$ *for all* $i = 1, \ldots, n$ *and* $j = 1, \ldots, p$,

(E.2) $\mathrm{E}[(\max_{1 \leq j \leq p} |X_{ij}|/B_n)^q] \leq 2$ *for all* $i = 1, \ldots, n$,

Let $\mathcal{A} = \mathcal{A}^r$ be a the class of all rectangles:

$$A = \{z = (z_1, \ldots, z_p)' \in \mathbb{R}^p : z_j \in [a_j, b_j] \text{ for all } j = 1, \ldots, p\}$$

for some $-\infty \leq a_j \leq b_j \leq \infty$, $j = 1, \ldots, p$.

Formal Results, I

Theorem (Central Limit Theorem)

Recall that

$$\rho_n := \sup_{A \in \mathcal{A}^r} \left| \mathrm{P}(S_n^X \in A) - \mathrm{P}(S_n^Y \in A) \right|$$

Assume (M.1-2), then under (E.1)

$$\rho_n \leq C \left(\frac{B_n^2 \log^7 (pn)}{n} \right)^{1/6} \tag{5}$$

where the constant C depends only on b, and under (E.2)

$$\rho_n \leq C \left[\left(\frac{B_n^2 \log^7 (pn)}{n} \right)^{1/6} + \left(\frac{B_n^2 \log^3 p}{n^{1-2/q}} \right)^{1/3} \right] \tag{6}$$

where the constant C depends only on b and q.

Remark: Bentkus (1985) provides an example, with $(X_{ij}, 1 \leq j \leq p) \subset \mathcal{F}$, where $\mathcal{F}$ is P-Donsker, such that $\rho_n \gtrsim (1/n)^{1/6}$.

Formal Results, II

Theorem (Gaussian and Empirical Bootstrap Theorem)

Define

$$\rho_n^* := \sup_{A \in \mathcal{A}^r} \left| \mathrm{P}(S_n^{X*} \in A \mid \{X_i\}_{i=1}^n) - \mathrm{P}(S_n^Y \in A) \right|.$$

Assume (M.1-2), then under (E.1), with probability at least $1 - \alpha$*,*

$$\rho_n^* \leq C \left(\frac{B_n^2 \log^5(pn) \log^2(1/\alpha)}{n} \right)^{1/6}, \tag{7}$$

where the constant C depends only on b, and under (E.2), with probability at least $1 - \alpha$*,*

$$\rho_n^* \leq C \left[\left(\frac{B_n^2 \log^5(pn) \log^2(1/\alpha)}{n} \right)^{1/6} + \left(\frac{B_n^2 \log^3 p}{\alpha^{2/q} n^{1-2/q}} \right)^{1/3} \right] \tag{8}$$

where the constant C depends only on b and q.

Some ingredients behind the proofs, I

Focus on max rectangles for simplicity:

$$A = \left\{ z = (z_1, \ldots, z_p)' \in \mathbb{R}^p : \max_{1 \le j \le p} z_j \le s \right\}, \quad s \in \mathbb{R}$$

- **Slepian's interpolation:**

  Define

  $$Z(t) := \sqrt{t} S_n^X + \sqrt{1-t} S_n^Y, \quad t \in [0,1]$$

  Then

  $$\mathrm{P}(S_n^X \in A) - \mathrm{P}(S_n^Y \in A) = \mathrm{E}[1(Z(1) \in A)] - \mathrm{E}[1(Z(0) \in A)]$$

- **Smoothing:**

  Approximate the indicator map

  $$z \mapsto 1(z \in A) = 1\left(\max_{1 \le j \le p} z_j \le s \right)$$

  by some smooth map

  $$z \mapsto m(z)$$

  by smoothing the interval indicator $y \mapsto 1(y \le s)$ and smoothing the max operator $z \mapsto \max_{1 \le j \le p} z_j$.

Some ingredients behind the proofs, II

- Calculations:

$$\mathrm{E}[1(Z(1) \in A)] - \mathrm{E}[1(Z(0) \in A)] \overset{(1)}{\approx} \mathrm{E}[m(Z(1))] - \mathrm{E}[m(Z(0))]$$
$$= \int_0^1 \mathrm{E}\left[\frac{dm(Z(t))}{dt}\right] dt$$
$$\overset{(2)}{\approx} 0$$

by proving the (1) first and that

$$\mathrm{E}\left[\frac{dm(Z(t))}{dt}\right] \approx 0$$

- Approximation of max operator by a logistic potential:

$$\left| \max_{1 \le j \le p} z_j - \beta^{-1} \log\left(\sum_{j=1}^{p} \exp(\beta z_j)\right) \right| \le \frac{\log p}{\beta}$$

Some ingredients behind the proofs, III

- Anti-concentration of suprema of Gaussian processes: (needed to show negligibility of errors due to smoothing the indicator function)

$$\sup_{t\in\mathbb{R}} \mathrm{P}\left(t \le \max_{1\le j\le p} S_{n,j}^Y \le t+\epsilon\right) \le 4\epsilon\left(\mathrm{E}\left[\max_{1\le j\le p} S_{n,j}^Y\right]+1\right) \lesssim \epsilon\sqrt{\log p},$$

  stated for the case when $\mathrm{E}[(S_{n,j}^Y)^2] = 1$ for each j. This is opposite of the (super)-concentration.

  Ref: CCK, PTRF.

- Stein's leave-one-out method (needed to simplify computations of expectations)

  (stability property of third-order derivatives of the logistic potential over certain subsets of $\mathbb{R}^p$ play a crucial role)

Some ingredients behind the proofs, IV

- Double Slepian Interpolation: to improve the dependence of bounds on n (Inspired by Bolthausen's (1984) arguments for combinatorial CLTs)

Details on Double Slepian Interpolation

- By using single Slepian interpolant

$$Z(t) := \sqrt{t} S_n^X + \sqrt{1-t} S_n^Y, \quad t \in [0,1]$$

the argument gives

$$\rho_n \leq \rho_n' := \sup_{t\in[0,1], A\in\mathcal{A}^r} |\mathrm{P}(Z(t) \in A) - \mathrm{P}(Z(0) \in A)| \leq n^{-1/8} \times C(n,p).$$

- Define the double Slepian interpolation

$$D(v,t) := \sqrt{v} Z(t) + \sqrt{1-v} S_n^W, \quad v \in [0,1], \quad t \in [0,1]$$

where S_n^W is an independent copy of S_n^Y.

- By doing double interpolation and using other ingredients mentioned above, obtain

$$\rho_n' \leq \frac{1}{2}\rho_n' + n^{-1/6} \times C(n,p)' \implies \text{result}$$

Connections to Literature

Classical CLTs under expanding dimension:

- Senatov (1980), Asriev and Rotar (1985), Portnoy (1986), Götze (1991), Bentkus (2003), L.H.Y. Chen and Roellin (2011), and others

Bootstrap and Multiplier methods:

- Gine and Zinn (1990), Koltchinskii (1981), Pollard (1982)

Stein's Method and other modern invariance principles

- Chatterjee (2005), Roellin (2011).

Spin glasses

- Panchenko (2013), Talagrand (2003), and others.

Further Results

- (CCK, Ann. Stat. 2014a). The results presented extend to suprema of empirical processes:

$$\sup_{t \in \mathbb{R}} \left| \mathrm{P}\left(\sup_{f \in \mathcal{F}_n} \mathbb{G}_n(f) \leq t \right) - \mathrm{P}\left(\sup_{f \in \mathcal{F}_n} \mathbb{G}_P(f) \leq t \right) \right| \to 0$$

  provided the complexity of $\mathcal{F}_n$ does not grow too quickly. The approximations are more generally applicable than Hungarian couplings (e.g. Rio), and competitive when both apply.
- There is also an analogous result for Gaussian and Empirical bootstrap.

Results do extend beyond rectangles

Definition (Sparsely convex sets)

For integer $s > 0$, we say that $A \subset \mathbb{R}^p$ is an s-**sparsely convex set** if there exist an integer $Q \leq p^C$ and convex sets $A_q \subset \mathbb{R}^p$, $q = 1, \dots, Q$, such that

$$A = \cap_{q=1}^{Q} A_q$$

and the indicator function of each A_q,

$$w \mapsto 1\{w \in A_q\},$$

depends at most on s components of its argument $w = (w_1, \dots, w_p)$

Examples of Sparsely Convex Sets

- Example 1: Rectangle (1-sparse) – intersection of 1-d shells

$$A = \{z \in \mathbb{R}^p : z_j \in [a_j, b_j] \quad \text{for all } j = 1, \ldots, p\}$$

for some $-\infty \leq a_j \leq b_j \leq \infty$, $j = 1, \ldots, p$

- Example 2: Sparse Polytope (s-sparse) — intersection of sparsely defined hyperplanes.

$$A = \{z \in \mathbb{R}^p : v_j' z \leq a_j, \quad \text{for all } j = 1, \ldots, m\}$$

for some $a_j \in \mathbb{R}$ such that $v_j \in \mathcal{S}^{p-1}$ with $\|v_j\|_0 \leq s$, $j = 1, \ldots, m$

- Example 3: Sparse convex body (s-sparse) — generated by the intersection of cylinders with s-dimensional base:

$$A = \{z \in \mathbb{R}^p : \|(z_j)_{j \in J_k}\|_2^2 \leq a_k : \; k = 1, ..., m\},$$

for some $a_k > 0$ and J_k being a subset of $\{1, \ldots, p\}$ of fixed cardinality s, $k = 1, \ldots, m$

Conditions

Let $b > 0$ and $q \geq 4$ be constants, and $(B_n)_{n=1}^{\infty}$ be a sequence of positive constants, possibly growing to ∞.

Consider the following conditions:

(M.1′) $n^{-1} \sum_{i=1}^{n} \mathrm{E}[(v' X_i)^2] \geq b$ *for all* $v \in \mathcal{S}^{p-1}$ with $\|v\|_0 \leq s$,

(M.2) $n^{-1} \sum_{i=1}^{n} \mathrm{E}[|X_{ij}|^{2+k}] \leq B_n^k$ *for all* $j = 1, \ldots, p$ *and* $k = 1, 2$.

(E.1) $\mathrm{E}[\exp(|X_{ij}| / B_n)] \leq 2$ *for all* $i = 1, \ldots, n$ *and* $j = 1, \ldots, p$,

(E.2) $\mathrm{E}[(\max_{1 \leq j \leq p} |X_{ij}| / B_n)^q] \leq 2$ *for all* $i = 1, \ldots, n$,

Formal Results, III

Theorem (CLT for Sparsely Convex Sets)

For $\mathcal{A}^s$ denoting the class of all s-sparsely convex sets, let

$$\rho_n := \sup_{A \in \mathcal{A}^s} \left| \mathrm{P}(S_n^X \in A) - \mathrm{P}(S_n^Y \in A) \right|$$

Assume (M.1') and (M.2), then under (E.1)

$$\rho_n \leq C \left(\frac{B_n^2 \log^7(pn)}{n} \right)^{1/6} \quad (9)$$

where the constant C depends only on b and s, and under (E.2)

$$\rho_n \leq C \left[\left(\frac{B_n^2 \log^7(pn)}{n} \right)^{1/6} + \left(\frac{B_n^2 \log^3 p}{n^{1-2/q}} \right)^{1/3} \right] \quad (10)$$

where the constant C depends only on b, q, and s.

Formal Results, IV

Theorem (Gaussian Bootstrap Theorem)

Define

$$\rho_n^* := \sup_{A \in \mathcal{A}^s} \left| \mathrm{P}(S_n^{X*} \in A \mid \{X_i\}_{i=1}^n) - \mathrm{P}(S_n^Y \in A) \right|.$$

Assume (M.1-2), then under (E.1), with probability at least $1 - \alpha$,

$$\rho_n^* \leq C \left(\frac{B_n^2 \log^5(pn) \log^2(1/\alpha)}{n} \right)^{1/6}, \tag{11}$$

where the constant C depends only on b and s, and under (E.2), with probability at least $1 - \alpha$,

$$\rho_n^* \leq C \left[\left(\frac{B_n^2 \log^5(pn) \log^2(1/\alpha)}{n} \right)^{1/6} + \left(\frac{B_n^2 \log^3 p}{\alpha^{2/q} n^{1-2/q}} \right)^{1/3} \right] \tag{12}$$

where the constant C depends only on b, q, and s.

Applications to Confidence Bands

- In [3], we used the results to give confidence bands for a large collection of approximately linear estimators in high-dimensional (approximately) sparse linear median regression and Z-estimation problems
- In [4], we used the results to give confidence bands for a large collection of approximately linear *function-valued* estimators in high-dimensional (approximately) sparse functional regression and Z-estimation problems
- As explained in [3-4], approximately linear estimators gan be generated by solving Neyman-orthogonal estimating equations.

Example: High-Dimensional Logistic Regression

Suppose that we are interested in studying the dependence of a random variable Y on a p-dimensional vector of covariates, $X = (X_1, \ldots, X_p)'$, where p is large

One approach to study this dependence is the following:

- For $u \in \mathcal{U}$, define
$$Y_u = 1\Big\{Y \leq u\Big\}$$
- For each $u \in \mathcal{U}$, consider the logistic regression:
$$\mathrm{E}[Y_u \mid X] = \Lambda(X'\theta_u) + r_u(X),$$
where $\Lambda(\cdot)$ is the logistic link function;
- The map
$$u \mapsto \theta_u = (\theta_{u1}, \ldots, \theta_{up})'$$
is a function-valued parameter of interest, and $r_u(X)$ is an asymptotically vanishing approximation error (pretend that $r_u(X) \equiv 0$)

General Moment Conditions Setup and Z-estimators

For each $u \in \mathcal{U}$ and $j = 1, \ldots, p$:

- Assume that the parameter θ_{uj} satisfies the following moment condition:
$$\mathrm{E}[\psi_{uj}(W, \theta_{uj}, \eta_{uj})] = 0,$$
where W is a random vector, η_{uj} is a nuisance parameter, and ψ_{uj} is a known moment function
- Let $(W_i)_{i=1}^n$ be a random sample from the distribution of W
- Here both p and the dimension of η_{uj} can be larger than n; in fact, η_{uj} can be a function
- Let $\widehat{\eta}_{uj}$ be an ML-type estimator of η_{uj} (in the logistic regression example, we assume that $\widehat{\eta}_{uj}$ is either a Lasso-type or Post-Lasso-type estimator of η_{uj})
- Define a Z-estimator of θ_{uj}:
$$\widehat{\theta}_{uj} = \arg\min_{\theta \in \Theta_{uj}} \left| \frac{1}{n} \sum_{i=1}^{n} \psi_{uj}(W_i, \theta, \widehat{\eta}_{uj}) \right|,$$
where Θ_{uj} is a set where θ_{uj} is known to belong to

Neyman Orthogonality Condition

We assume that the moment functions ψ_{uj} satisfy the following Neyman orthogonality condition: for all $u \in \mathcal{U}$, $j = 1, \ldots, p$, and $\eta \in \mathcal{T}_{uj}$, the Gateaux derivative map with respect to the nuisance parameter vanishes:

$$\partial_r \mathrm{E}\psi_{uj}\{W, \theta_{uj}, \eta_{uj} + r(\eta - \eta_{uj})\}\Big|_{r=0} = 0$$

where $\mathcal{T}_{uj}$ is a set such that $\widehat{\eta}_{uj} \in \mathcal{T}_{uj}$ with probability approaching one uniformly over $u \in \mathcal{U}$ and $j = 1, \ldots, p$

> Neyman orthogonality condition: Heuristically, small deviations in nuisance parameters η_{uj} do not invalidate moment conditions.

Constructing Orthogonal Moment Conditions for Likelihood Models, I

In the likelihood models, we can generally construct moment functions satisfying the orthogonality condition building upon classic ideas of Neyman (1958, 1979)

Suppose log-likelihood function is given by $\ell(W, \theta, \beta)$ where

- θ is a scalar parameter of interest
- β is a *d*-dimensional nuisance parameter

Under regularity, true parameter values, θ_0 and β_0, satisfy

$$\mathrm{E}[\partial_\theta \ell(W, \theta_0, \beta_0)] = 0, \quad \mathrm{E}[\partial_\beta \ell(W, \theta_0, \beta_0)] = 0$$

However, the original moment function $\varphi(W, \theta, \beta) = \partial_\theta \ell(W, \theta, \beta)$ in general does not satisfy the orthogonality condition

Constructing Orthogonal Moment Conditions for Likelihood Models, II

We, therefore, construct new moment function satisfying the orthogonality condition:

$$\psi(W,\theta,\eta) = \partial_\theta \ell(W,\theta,\beta) - \mu \partial_\beta \ell(W,\theta,\beta),$$

- Nuisance parameter: $\eta = (\beta', \mu')'$
- μ is the d-dimensional orthogonalization parameter
  - True value μ_0 of the parameter μ is given by $\mu_0 = J_{\beta\beta}^{-1} J_{\beta\theta}$ for

$$J = \begin{pmatrix} J_{\theta\theta} & J_{\theta\beta} \\ J_{\beta\theta} & J_{\beta\beta} \end{pmatrix} = \partial_{(\theta',\beta')} \mathrm{E}\left[\partial_{(\theta',\beta')'} \ell(W,\theta,\beta)\right]\Big|_{\theta=\theta_0;\ \beta=\beta_0}$$

- Then $\mathrm{E}[\psi(W,\theta_0,\eta_0)] = 0$ for $\eta_0 = (\beta_0', \mu_0')'$ (provided μ_0 is well-defined)
- And ψ obeys the orthogonality condition: $\partial_\eta \mathrm{E}[\psi(W,\theta_0,\eta)]\Big|_{\eta=\eta_0} = 0$

This construction can be used to derive the moment functions ψ_{uj} satisfying the orthogonality condition in the logistic regression example with the log-likelihood function

$$\ell(X, Y_u, \theta_u) = Y_u \log \Lambda(X'\theta_u) + (1 - Y_u) \log(1 - \Lambda(X'\theta_u))$$

using θ_{uj} as a parameter of interest and $\eta = (\mu, \theta_{u,-j})$ as the nuisance parameter

Asymptotic Normality for Many Z-estimators

Theorem (Asymptotic normality)

Under certain regularity conditions, the estimators $\widehat{\theta}_{uj}$ *satisfy*

$$\frac{\sqrt{n}(\widehat{\theta}_{uj} - \theta_{uj})}{\sigma_{uj}} = \mathbb{G}_n \bar{\psi}_{uj} + o_P(1) \quad \textit{uniformly over } u \in \mathcal{U} \textit{ and } j = 1, \ldots, p,$$

where

$$\sigma_{uj}^2 = \frac{\mathrm{E}[\psi_{uj}^2(W, \theta_{uj}, \eta_{uj})]}{J_{uj}^2}, \quad J_{uj} = \partial_\theta \Big\{ \mathrm{E}[\psi_{uj}(W, \theta, \eta_{uj})] \Big\} \Big|_{\theta = \theta_{uj}},$$

and

$$\bar{\psi}_{uj}(W) = -\frac{\psi_{uj}(W, \theta_{uj}, \eta_{uj})}{\sigma_{uj} J_{uj}}.$$

As a consequence, under further regularity conditions,

$$\sup_{t \in \mathbb{R}} \left| \mathrm{P}\left(\sup_{u \in \mathcal{U}, 1 \leq j \leq p} \left| \frac{\sqrt{n}(\widehat{\theta}_{uj} - \theta_{uj})}{\sigma_{uj}} \right| \leq t \right) - \mathrm{P}\left(\sup_{u \in \mathcal{U}, 1 \leq j \leq p} |\mathcal{N}_{uj}| \leq t \right) \right| = o(1),$$

where $\{\mathcal{N}_{uj} \colon u \in \mathcal{U},\ j = 1, \ldots, p\}$ *is a certain Gaussian process.*

Simultaneous Confidence Bands via Bootstrap

The theorem above can be used to construct simultaneous confidence bands for the parameters θ_{uj}:

1. For each $u \in \mathcal{U}$ and $j = 1, \ldots, p$, let $\widehat{\sigma}_{uj}$ and $\widehat{J}_{uj}$ be estimators of σ_{uj} and J_{uj}, respectively
2. For each $u \in \mathcal{U}$ and $j = 1, \ldots, p$, let
$$\widehat{\psi}_{uj}(\cdot) = -\frac{\psi_{uj}(\cdot, \widehat{\theta}_{uj}, \widehat{\eta}_{uj})}{\widehat{\sigma}_{uj}\widehat{J}_{uj}}$$
be an estimator of $\bar{\psi}_{uj}(\cdot)$
3. Let $(\xi_i)_{i=1}^n$ be a random sample from the $N(0, 1)$ distribution that is independent of the data $(W_i)_{i=1}^n$
4. For $\alpha \in (0, 1)$, let c_α be the $(1 - \alpha)$ quantile of the conditional distribution of
$$T = \sup_{u \in \mathcal{U}, 1 \leq j \leq p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \widehat{\psi}_{uj}(W_i) \right| \quad \text{given the data } (W_i)_{i=1}^n$$
5. Define the simultaneous confidence bands:
$$\widehat{\Theta}_{uj} = \left(\widehat{\theta}_{uj} - \frac{c_\alpha \widehat{\sigma}_{uj}}{\sqrt{n}}, \widehat{\theta}_{uj} + \frac{c_\alpha \widehat{\sigma}_{uj}}{\sqrt{n}} \right), \quad u \in \mathcal{U}, \; j = 1, \ldots, p$$

Bootstrap Validity

Theorem

Under certain regularity conditions, the simultaneous confidence bands $\widehat{\Theta}_{uj}$ *satisfy*

$$\mathrm{P}\Big(\theta_{uj} \in \widehat{\Theta}_{uj} \textit{ for all } u \in \mathcal{U},\ j = 1, \ldots, p\Big) = 1 - \alpha + o(1).$$

In fact, the result holds uniformly over a large class of data-generating processes.

Conclusion

Thank you!