CLT and Bootstrap in High Dimensions

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

May 18, 2015

Introduction

This presentation is based on:

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

2. “Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors," *Ann. Stat.*, 2013
3. “Comparison and Anti-Concentration Bounds for Maxima of Gaussian Vectors", *Prob. Theory Rel. Fields*, 2015+.
4. “Gaussian Approximation of Suprema of Empirical Processes," *Ann. Stat.*, 2014a
5. “Anti-Concentration and Adaptive, Honest Confidence Bands" *Ann. Stat.*, 2014b

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:

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

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:

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

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

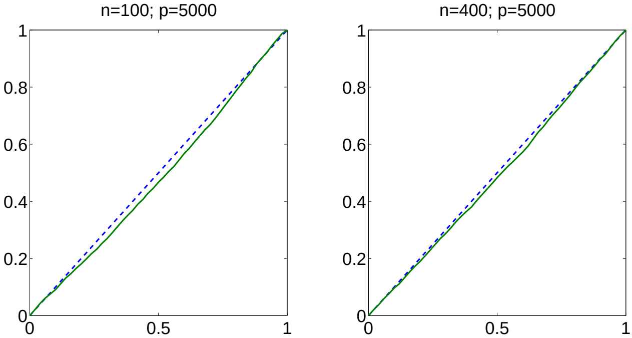

Figure: P-P plots comparing $\mathrm{P}\left(\|S_n^Y\|_\infty \leq t\right)$ and $\mathrm{P}\left(\|S_n^X\|_\infty \leq t\right)$. The dashed line is the 45° line.

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)

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

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

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

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:

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

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)

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

  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

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

- 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

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

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.
- (CCK, Ann. Stat. 2014b). Provide an application to the problem of uniform and uniform adaptive confidence bands in nonparametric problems, in particular providing a practical version of Gine-Nickl-type bands.

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 $Q > 0$ 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: (1-sparse)

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

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

- Example 2: (s-sparse)

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

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

- Example 3: (s-sparse)

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

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

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* $j = 1, \dots, p$ *and* $k = 1, 2$.

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

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

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)

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

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

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

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*.*

Conclusion

Thank you very much!