Central Limit Theorem and Bootstrap in High Dimensions

Zihang Liang

University of Science and Technology of China

October 2, 2024

Outline

Outline

Introduction

- Let $X_1, X_2, ..., X_n$ be a sequence of random vectors in $\mathbb{R}^p$
- Define the normalized sum of X:

$$S_{n,X} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

- Let covariance matrix $\Xi = \text{cov}(S_{n,X})$, and a Guassian random vector $G \sim N(0, \Xi)$
- Define

$$\rho_n(\mathcal{A}) = \sup_{A \in \mathcal{A}} |\mathbb{P}(S_n^X \in A) - \mathbb{P}(G \in A)| \tag{1}$$

where $\mathcal{A}$ is a class of Borel set of $\mathbb{R}^p$

Introduction

- **Question:** how fast the $p = p_n$ is allowed to grow as $n \to \infty$ while guaranteeing $\rho_n \to 0$?
- We hope to construct a convergence rate which allows p to grow sub-exponentially fast in the sample size n.
- The main result[1][2] suggests that under some conditions, we have

$$\rho_n \to 0 \text{ if } p_n = O(e^{Cn^c})$$

[1] Victor Chernozukov, Denis Chetverikov, and Kengo Kato. "Central Limit Theorems and Bootstrap in High Dimensions". In: *The Annals of Probability* 45 (Dec. 2014). DOI: 10.1214/16-AOP1113.

[2] Jinyuan Chang, Xiaohui Chen, and Mingcong Wu. "Central limit theorems for high dimensional dependent data". In: *Bernoulli* 30.1 (Feb. 2024). ISSN: 1350-7265. DOI: 10.3150/23-bej1614. URL: http://dx.doi.org/10.3150/23-BEJ1614.

Outline

1. Introduction
2. Central Limit Theorem for hyperrectangles
3. Generalization
4. Application for Bootstrap

Notation

- For $x, y \in \mathbb{R}^p$, $x \leq y \iff x_i \leq y_i$,$\forall i \in [p]$
- For $x \in \mathbb{R}^p$, $|x|_p$ means p-norm, and $|x|_0 = \#\{i : x_i \neq 0\}$
- $a_n \ll b_n \iff a_n = o(b_n) \iff \lim_{n\to\infty} \frac{a_n}{b_n} = 0$
- $a_n \lesssim b_n \iff a_n = O(b_n) \iff \limsup_{n\to\infty} \frac{a_n}{b_n} < \infty$
- $\psi_\alpha : [0, \infty) \to [0, \infty), \psi_\alpha(x) = \exp(x^\alpha) - 1$, where $\alpha \geq 1$

  Oclicz norm: $||X||_{\psi_\alpha} = inf\{\lambda : E[\psi_\alpha(|X|/\lambda)] \leq 1\}$

Outline

Central Limit Theorem for independent data

- In this subsection, we assume that $\{X_i\}$ is an independent sequence.

Central Limit Theorem for independent data

- In this subsection, we assume that $\{X_i\}$ is an independent sequence.
- Assume $\{X_i\}$ satisfy the following **conditions**:
  - **CI1:** (Non-degeneracy of convariance matrix) $\mathrm{Var}(S_{n,X,j}) \geq b$ for all $j \in [p]$
  - **CI2:** (Bound for moment) $n^{-1}\sum_{i=1}^{n} \mathbb{E}[|X_{i,j}|^{2+k}] \leq B_n^k$ for any $j \in [p]$, $k = 1, 2$

Central Limit Theorem for independent data

- In this subsection, we assume that $\{X_i\}$ is an independent sequence.
- Assume $\{X_i\}$ satisfy the following **conditions**:
  - **CI1:** (Non-degeneracy of convariance matrix) $\mathrm{Var}(S_{n,X,j}) \geq b$ for all $j \in [p]$
  - **CI2:** (Bound for moment) $n^{-1}\sum_{i=1}^{n} \mathbb{E}[|X_{i,j}|^{2+k}] \leq B_n^k$ for any $j \in [p]$, $k = 1, 2$
- smoothing: Let $\varrho_n = \sup_{y\in\mathbb{R}^p, v\in[0,1]} |\mathbb{P}(\sqrt{v}S_{n,X} + \sqrt{1-v}G \leq y) - \mathbb{P}(G \leq y)|$
- Obviously, to bound $\rho_n(\mathcal{A}^{re})$, we just need to bound ϱ_n

Central Limit Theorem for independent data

- By independency of $\{X_i\}$, let $Y_1, \ldots, Y_n$ be a sequence of Guassian random vectors in $\mathbb{R}^p$, and $Y_i \sim N(0, E[X_i X_i^T])$, and $\{W_i\}$ is a copy of $\{Y_i\}$, then $S_{n,Y} \overset{\mathrm{d}}{=} S_{n,W} \overset{\mathrm{d}}{=} G$, so

$$\varrho_n = \sup_{y \in \mathbb{R}^p, v \in [0,1]} |\mathbb{P}(\sqrt{v} S_{n,X} + \sqrt{1-v} S_{n,Y} \leq y) - \mathbb{P}(S_{n,W} \leq y)|$$

- This method is called Slepian-Stein interpolation[3]

[3] It's actually Stein method and Slepian interpolation, but in fact they are the same method, see (Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors". In: *The Annals of Statistics* 41.6 [Dec. 2013]. ISSN: 0090-5364. DOI: 10.1214/13-aos1161. URL: http://dx.doi.org/10.1214/13-AOS1161) in Appendix H

Key Lemma

- The Key Lemma indicates that ϱ_n satisfies the following inequality for all $\phi \geq 1$:

$$\varrho_n \lesssim \frac{\phi^2 \log^2 p}{n^{1/2}} \{\phi L_n \varrho_n + L_n \log^{1/2} p + \phi(M_{n,X}(\phi) + M_{n,Y}(\phi))\} + \frac{\log^{1/2} p}{\phi}$$

where

$$M_{n,X}(\phi) = n^{-1} \sum_{i=1}^{n} \mathbb{E}[\max_{1 \leq j \leq p} |X_{i,j}|^3 1\{\max_{1 \leq j \leq p} |X_{i,j}| > \sqrt{n}/4\phi \log p\}]$$

$$M_{n,Y}(\phi) = n^{-1} \sum_{i=1}^{n} \mathbb{E}[\max_{1 \leq j \leq p} |Y_{i,j}|^3 1\{\max_{1 \leq j \leq p} |Y_{i,j}| > \sqrt{n}/4\phi \log p\}]$$

$$L_n = \max_{j \in [p]} n^{-1} \sum_{i=1}^{n} \mathbb{E}[|X_{i,j}|^3]$$

High-dimensional CLT for hyperrectangles

- Select ϕ properly, and bound $M_{n,X}, M_{n,Y}, L_n$ respectively, we can get an explicit result
- (**Theorem**) Suppose that $||X_{i,j}||_{\psi_1} \leq B_n$ for all $i \in [n], j \in [p]$,then

$$\rho_n(\mathcal{A}^{re}) \lesssim \frac{B_n^{1/3}\log^{7/6}(pn)}{n^{1/6}}$$

- (**Theorem**) Suppose that $||X_{i,j}||_{\psi_\gamma} \leq B_n$ for all $i \in [n], j \in [p]$, and some $\gamma \geq 4$,then

$$\rho_n(\mathcal{A}^{re}) \lesssim \frac{B_n^{1/3}\log^{7/6}(pn)}{n^{1/6}} + \frac{B_n^{2/3}\log(p)}{n^{(\gamma-2)/3\gamma}}$$

Sketch of proof of Key Lemma

The following is some preparations, we hope to smooth the ϱ_n:

- Note that $x \le y \iff \max_{j\in[p]}(x_j - y_j) \le 0 \iff I_{(-\infty,0]}(\max_{j\in[p]}(x_j - y_j)) = 1$, so

$$\begin{aligned}
\varrho_n &= \sup_{y\in\mathbb{R}^p, v\in[0,1]} |\mathbb{P}(\sqrt{v}S_{n,X} + \sqrt{1-v}S_{n,Y} \le y) - \mathbb{P}(S_{n,W} \le y)| \\
&= \sup_{y\in\mathbb{R}^p, v\in[0,1]} |\mathbb{E}[I_{(-\infty,0]}(\max_{j\in[p]}(\sqrt{v}S_{n,X,j} + \sqrt{1-v}S_{n,Y,j} - y_j))] \\
&\quad - \mathbb{E}[I_{(-\infty,0]}(\max_{j\in[p]}(S_{n,W,j} - y_j))]|
\end{aligned}$$

- Smooth the function $I_{(-\infty,0]}$ and $\max_{j\in[p]}(\cdot_j - y_j)$

Sketch of proof of Key Lemma

- (Smooth indicate function)Let $g \in C^{\infty} : \mathbb{R} \to \mathbb{R}$ satisfy that $g(t) = 1$ for $t \leq 0$, and $g(t) = 0$ for $t \geq \phi^{-1}$
- (Smooth maximum function)Let $F_{\beta}(w) = \beta^{-1} \log(\sum_{j=1}^{p} \exp(\beta(w_j - y_j)))$, where $\beta = \phi \log p$
- $F_{\beta}(w) \approx \max_{j \in [p]}(w_j - y_j)$, $g \approx I_{(-\infty,0)}$.
- Let $m = g \circ F_{\beta}$, $\mathcal{I}_n = m(\sqrt{v} S_{n,X} + \sqrt{1-v} S_{n,Y}) - m(S_{n,W})$
- In fact,

$$\varrho_n \lesssim |\mathbb{E}[\mathcal{I}_n]| + \phi^{-1} \log^{1/2} p$$

Sketch of proof of Key Lemma

- double Slepian interpolant:

$$Z_i(t) = \frac{1}{\sqrt{n}}\{\sqrt{t}(\sqrt{v}X_i + \sqrt{1-v}Y_i) + \sqrt{1-t}W_i\}$$

$$Z(t) = \sum_{i=1}^{n} Z_i(t)$$

So

$$\mathcal{I}_n = \int_0^1 \frac{dm(Z(t))}{dt} dt$$

- Note that $\{Z_i\}$ is mutually independent

Sketch of proof of Key Lemma

Left-one-out:

- By Taylor expansion: let $Z^{(i)} = Z - Z_i \Longrightarrow Z^{(i)}$ is independent to Z_i

$$
\begin{aligned}
\mathbb{E}[\mathcal{I}_n] &= \frac{1}{2}\sum_{j\in[p]}\sum_{i\in[n]}\int_0^1 \mathbb{E}[m_j(Z)\dot{Z}_{i,j}]dt \\
&= \frac{1}{2}\sum_{j\in[p]}\sum_{i\in[n]}\int_0^1 \mathbb{E}[m_j(Z^{(i)})\dot{Z}_{i,j}]dt + \frac{1}{2}\sum_{j,k\in[p]}\sum_{i\in[n]}\int_0^1 \mathbb{E}[m_{jk}(Z^{(i)})\dot{Z}_{i,j}Z_{i,k}]dt + \\
&\frac{1}{2}\sum_{j,k,l\in[p]}\sum_{i\in[n]}\int_0^1\int_0^1 (1-\tau)\mathbb{E}[m_{jkl}(Z^{(i)}+\tau Z_i)\dot{Z}_{i,j}Z_{i,k}Z_{i,l}]d\tau dt
\end{aligned}
$$

Sketch of proof of Key Lemma

- The independency of $\{X_i\} \Longrightarrow$ The first and second term equal to 0, so we just need to bound the last term. Write the last term as III
- Divide III into small part and large part:

$$III_1 = \sum_{j,k,l} \sum_i \int_0^1 \int_0^1 (1-\tau) E[\chi_i m_{jkl}(Z^{(i)} + \tau Z_i) \dot{Z}_{ij} Z_{ik} Z_{il}] d\tau dt$$

$$III_2 = \sum_{j,k,l} \sum_i \int_0^1 \int_0^1 (1-\tau) E[(1-\chi_i) m_{jkl}(Z^{(i)} + \tau Z_i) \dot{Z}_{ij} Z_{ik} Z_{il}] d\tau dt$$

where $\chi_i = 1\{\max_{i \in [p], i \in [n]} \{|X_{i,j}|, |Y_{i,j}|, |W_{i,j}|\} \leq \sqrt{n}/4\beta\}$

Sketch of proof of Key Lemma

- Estimate III_1, III_2 respectively
- We have

$$\mathbb{E}[\mathcal{I}_n] \lesssim \frac{\phi^2 \log^2 p}{n^{1/2}}\{\phi L_n \varrho_n + L_n \log^{1/2} p + \phi(M_{n,X}(\phi) + M_{n,Y}(\phi))\}$$

Outline

Central Limit Theorem for dependent data

Now, we hope to omit the independency of $\{X_i\}$.

- **Difficulty 1**: without independency, **Slepian interpolation** is invalid.

  For example: Independent sequence $\{Y_i\}$ such that $Y_i \sim N(0, \Sigma_i)$, where $\Sigma_i = \text{cov}(X_i)$, however:

$$\text{cov}(G) = \text{cov}(S_{n,X}) = n^{-1} \sum_{t,s \in [n]} \mathbb{E}[X_t X_s^T] \neq n^{-1} \sum_{t \in [n]} \mathbb{E}[X_t X_t^T] = \text{cov}(S_{n,Y})$$

  In other word, interpolating for each X_i **cannot approximate** G.

Central Limit Theorem for dependent data

- **Difficulty 2**: without independency, **left-one-out method** is invalid.

  For example:

$$\mathbb{E}[m_j(Z^{(i)})\dot{Z}_{i,j}] = \mathbb{E}[m_j(Z^{(i)})\mathbb{E}[\dot{Z}_{i,j}|\mathcal{F}]] \neq 0$$

  In other words, we have to bound the **conditional expectation** respectively!

Central Limit Theorem for dependent data

- **Solution 1**: "Package" data so that we can approximate G applying interpolation:
  - Large-and-small-blocks

Central Limit Theorem for dependent data

- **Solution 1**: "Package" data so that we can approximate G applying interpolation:
  - Large-and-small-blocks
- **Solution 2**: Introduce some measures to measure the dependency (Do not be too corelated), which can help us to bound conditional expectation:
  - α-mixing sequence
  - Dependency graph
  - Sequence with physical dependence

Later, we always assume $p \geq n^{\kappa}$ since we only care about the case that $p \gg n$.

CLT for dependent data: α-mixing sequence

Concepts of $\alpha-$mixing:

- $\mathcal{A}, \mathcal{B}$ are σ-field, let $\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB)|$
- Let $\mathcal{F}_{-\infty}^{t} = \sigma(X_i, i \leq t)$, $\mathcal{F}_{t}^{\infty} = \sigma(X_i, i \geq t)$
- α-coefficient of $\{X_i\}$: $\alpha(n) = \sup_{t \geq 1} \alpha(\mathcal{F}_{-\infty}^{t}, \mathcal{F}_{t+n}^{+\infty})$
- $\{X_i\}$ is a α-mixing (or strong mixing) sequence, if $\lim_{n \to \infty} \alpha(n) = 0$

CLT for dependent data: α-mixing sequence

There are some conditions which we will assume later.

- **Condition 1** (Sub-exponential moment). There exist constants$\gamma_1 \geq 1, B_n \geq 1$, such that $||X_{i,j}||_{\psi_{\gamma_1}} \leq B_n$ ($\Rightarrow$ distribution function $\mathbb{P}(|X| > u)$ has an exponential decay)
- **Condition 2** (Decay of α-mixing coefficient). There exist some universal constants $K_1 > 1, K_2 > 0$ and $\gamma_2 > 0$, such that $\alpha_n(k) \leq K_1 \exp(-K_2 k^{\gamma_2})$ for any $k \geq 1$
- **Condition 3** (Non-degeneracy of covariance matrix). There exist a universal constant $K_3 > 0$ such that $\min_{j \in [p]} \mathrm{Var}(S_{n,X,j}) \geq K_3$

CLT for dependent data: α-mixing sequence

- Recall that $\varrho_n = \sup_{y \in \mathbb{R}^p, v \in [0,1]} |\mathbb{P}(\sqrt{v} S_{n,X} + \sqrt{1-v} G \leq y) - \mathbb{P}(G \leq y)|$
- (**Theorem**) Assume $\{X_i\}$ is an α-mixing sequence. Under **Conditions 1-3**:

$$\varrho_n \lesssim \frac{B_n^{2/3} (\log p)^{(1+2\gamma_2)/3\gamma_2}}{n^{1/9}} + \frac{B_n (\log p)^{7/6}}{n^{1/9}}$$

provided that $(\log p)^{3-\gamma_2} = o(n^{\gamma_2/3})$

Large-and-Small-Blocks Technique: Details

- Let $Q = o(n)$, and $Q = b + h$, where $h \ll b$

Large-and-Small-Blocks Technique: Details

- Let $Q = o(n)$, and $Q = b + h$, where $h \ll b$
- Divide the sequence into L blocks: $L = \lfloor \frac{n}{Q} \rfloor$

Large-and-Small-Blocks Technique: Details

- Let $Q = o(n)$, and $Q = b + h$, where $h \ll b$
- Divide the sequence into L blocks: $L = \lfloor \frac{n}{Q} \rfloor$
- Divide each blocks into large parts and small parts: For $\ell \in [L]$, Let $\mathcal{G}_\ell = \{\ell Q + 1, \ell Q + 2, ..., \ell Q + b, ...(\ell + 1)Q\}$, and $\mathcal{I}_\ell = \{\ell Q + 1, \ell Q + 2, ..., \ell Q + b\}$, $\mathcal{J}_\ell = \{\ell Q + b + 1, ..., (\ell + 1)Q\}$, and $\mathcal{J}_{L+1} = \{\text{The remaining } X_t\}$

Large-and-Small-Blocks Technique: Details

- Let $Q = o(n)$, and $Q = b + h$, where $h \ll b$
- Divide the sequence into L blocks: $L = \lfloor \frac{n}{Q} \rfloor$
- Divide each blocks into large parts and small parts: For $\ell \in [L]$, Let $\mathcal{G}_\ell = \{\ell Q + 1, \ell Q + 2, ..., \ell Q + b, ...(\ell + 1)Q\}$, and $\mathcal{I}_\ell = \{\ell Q + 1, \ell Q + 2, ..., \ell Q + b\}$, $\mathcal{J}_\ell = \{\ell Q + b + 1, ..., (\ell + 1)Q\}$, and $\mathcal{J}_{L+1} = \{\text{The remaining } X_t\}$
- packaging: Let $\tilde{X}_\ell = b^{-1/2} \sum\limits_{t \in \mathcal{I}_\ell} X_t$

Large-and-Small-Blocks Technique: Details

- A series of normal random vectors: Let independent sequence $\{Y_t\}_1^n$ such that $Y_t \sim N(0, \mathbb{E}[\tilde{X}_\ell \tilde{X}_\ell^\top])$ for $t \in \mathcal{I}_\ell$
- blocks of $\{Y_t\}$: Define $\tilde{Y}_\ell = b^{-1/2} \sum_{t \in \mathcal{I}_\ell} Y_t$. Corresponding: $\tilde{Y}_\ell \longleftrightarrow \tilde{X}_\ell$
- normalized sum: $S_{n,X} = \frac{1}{\sqrt{n}} \sum_{t=1}^n X_t$, $S_{n,X}^{(1)} = \frac{1}{\sqrt{L}} \sum_{\ell=1}^L \tilde{X}_\ell$, $S_{n,Y}^{(1)} = \frac{1}{\sqrt{L}} \sum_{\ell=1}^L \tilde{Y}_\ell$

Large-and-Small-Blocks Technique: Advantage

Advantages:

Large-and-Small-Blocks Technique: Advantage

Advantages:

- By dividing the sequence $\{X_i\}$ into a series of "large blocks" and "small blocks", we can construct a series of normal distribution $\{Y_\ell\}$, such that $G \approx n^{-1/2} \sum_{\ell} Y_\ell$, so we can apply **Slepian interpolation** for these "blocks".

Large-and-Small-Blocks Technique: Advantage

Advantages:

- By dividing the sequence $\{X_i\}$ into a series of "large blocks" and "small blocks", we can construct a series of normal distribution $\{Y_\ell\}$, such that $G \approx n^{-1/2} \sum_\ell Y_\ell$, so we can apply **Slepian interpolation** for these "blocks".
- By throwing out "small blocks", we can construct "time gap" between 2 "large blocks". As a result, we can bound the conditional expectation by $\alpha(n)$

Sketch of proof

- By triangle inequality:

$$\begin{aligned}
\varrho_n \leq & \underbrace{\sup_{y\in\mathbb{R}^p, v\in[0,1]} \left|\mathbb{P}(\sqrt{v}S_{n,X} + \sqrt{1-v}G \leq y) - \mathbb{P}(\sqrt{v}S_{n,X} + \sqrt{1-v}S^{(1)}_{n,Y} \leq y)\right|}_{I_1} \\
& + \underbrace{\sup_{y\in\mathbb{R}^p, v\in[0,1]} \left|\mathbb{P}(\sqrt{v}S_{n,X} + \sqrt{1-v}S^{(1)}_{n,Y} \leq y) - \mathbb{P}(S^{(1)}_{n,Y} \leq y)\right|}_{I_2} \\
& + \underbrace{\sup_{y\in\mathbb{R}^p} \left|\mathbb{P}(S^{(1)}_{n,Y} \leq y) - \mathbb{P}(G \leq y)\right|}_{I_3}
\end{aligned}$$

Sketch of proof

Bound I_1, I_3:

- **(Gaussian comparison)** $|\mathbb{P}(S_{n,Y}^{(1)} \leq y) - \mathbb{P}(G \leq y)| \lesssim |\mathbf{cov}(S_{n,Y}^{(1)}) - \mathbf{cov}(G)|_\infty (\log p)^{2/3}$
- **(Difference of covariance matrix)** $|\mathbf{cov}(S_{n,Y}^{(1)}) - \mathbf{cov}(G)|_\infty \lesssim B_n^2(hb^{-1} + bn^{-1})$
- $I_1 \leq I_3$

Sketch of proof

- $I_2 \approx \sup_{y\in\mathbb{R}^p, v\in[0,1]} |\mathbb{P}(\sqrt{v}S_{n,X}^{(1)} + \sqrt{1-v}S_{n,Y}^{(1)} \le y) - \mathbb{P}(S_{n,Y}^{(1)} \le y)| \xrightarrow{\text{Slepian-Stein Method}} \text{Bound}$
- The right-hand side just replaces $S_{n,X}$ in I_2 by $S_{n,X}^{(1)}$. What we do is throwing out the "small blocks" so that we can apply Slepian-Stein method.
- $\tilde{X}_\ell$ and $\tilde{X}_{\ell+1}$ have a time gap h(the size of small blocks).
- Bound conditional expectation by $\alpha(n)$: $\mathbb{E}\{|\mathbb{P}(\tilde{X}_{\ell,j} > u|\mathcal{F}_{-\ell}) - \mathbb{P}(\tilde{X}_{\ell,j} > u)|\} \lesssim \alpha(h)$, where $\mathcal{F}_{-\ell} = \sigma(\{\tilde{X}_s\}_{s\neq\ell})$.

CLT for dependent data: Dependent graph

- Let undirected graph $G_n = (V_n, E_n)$, where $V_n = [n]$ with $t \overset{denote}{\Longleftrightarrow} X_t$, $(t, s) \in E_n \Longleftrightarrow X_t$ and X_s are dependent. In particular, $(t, t) \in E_n$ for all t.

CLT for dependent data: Dependent graph

- Let undirected graph $G_n = (V_n, E_n)$, where $V_n = [n]$ with $t \overset{denote}{\Longleftrightarrow} X_t$, $(t, s) \in E_n \Longleftrightarrow X_t$ and X_s are dependent. In particular, $(t, t) \in E_n$ for all t.
- **(Packaging)** Let $\mathcal{N}_t := \{s : (t, s) \in E_n\}$ be the set of neighbor of X_t, and $\mathcal{N}_t^* := \cup_{s \in \mathcal{N}_t} \mathcal{N}_s$ be the set of second-degree neighbor of X_t.
- $D_n := \max_{t \in [n]} |\mathcal{N}_t|$, $D_n^* := \max_{t \in [n]} |\mathcal{N}_t^*|$
- Let $Z_{\mathcal{N}_t} := \sum_{t \in \mathcal{N}_t} Z_t$ be the sum of the random vector dependent to Z_t.

CLT for dependent data: Dependent graph

The proof coincides with the proof in the case of independent data. What difference is when left-one-out by Taylor expansion:

- **Independent** : $m(Z) = m(Z^{(-t)}) + \sum m_j(Z^{(-t)})Z_{t,j} + \dots$
- **Dependent** : $m(Z) = m(Z^{(-\mathcal{N}_t)}) + \sum m_j(Z^{(-\mathcal{N}_t)})Z_{\mathcal{N}_t,j} + \dots$

CLT for dependent data: Dependent graph

- **(Theorem)** Suppose **Condition** 1 and 3, it holds that

$$\varrho_n \lesssim \frac{B_n (D_n D_n^*)^{1/3} (\log p)^{7/6}}{n^{1/6}}$$

CLT for dependent data: physical dependence

In this section, we consider time series model.

- Time series model: $X_t = f_t(\epsilon_t, \epsilon_{t-1}, \ldots)$, where $t \geq 1$
- **(Coupling)**ϵ_t' independent copy of ϵ_t, let $X_{t,\{m\}}' := f_t(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-m}', \epsilon_{t-m-1}, \ldots)$
- $\theta_{m,q,j} = \sup_{t \geq 1} ||X_{t,j} - X_{t,j,\{m\}}'||_q$, $\Theta_{m,q,j} = \sum_{i=m}^{\infty} \theta_{i,q,j}$, where $q > 0$
- $||X_{\cdot,j}||_{q,\alpha} = \sup_{m \geq 0} (m+1)^{\alpha} \Theta_{m,q,j}$, and $||X_{\cdot,j}||_{\psi_\nu,\alpha} = \sup_{q \geq 2} q^{-\nu} ||X_{\cdot,j}||_{q,\alpha}$
- $\Psi_{q,\alpha} = \max_{j \in [p]} ||X_{\cdot,j}||_{q,\alpha}$, $\Phi_{\psi_\nu,\alpha} = \max_{j \in [p]} ||X_{\cdot,j}||_{\psi_\nu,\alpha}$

CLT for dependent data: physical dependence

- **(Theorem)** $\Phi_{\psi_\nu,\alpha} < \infty$ for some α, ν,then:

  (i) Under **Condition** 3,

$$\varrho_n \lesssim \frac{\Phi_{\psi_\nu,0}(\log p)^{7/6}}{n^{\alpha/(3+9\alpha)}} + \frac{\Psi_{2,\alpha}^{1/3}\Psi_{2,0}^{1/3}(\log p)^{2/3}}{n^{\alpha/(3+9\alpha)}} + \frac{\Phi_{\psi_\nu,\alpha}(\log p)^{1+\nu}}{n^{\alpha/(1+3\alpha)}}$$

  provided that $(\log p)^{\max\{6\nu-1,(5+6\nu)/4\}} = o\left\{n^{\alpha/(1+3\alpha)}\right\}$

  (ii)Under **Condition** 1 and 3,

$$\varrho_n \lesssim \frac{B_n(\log p)^{7/6}}{n^{\alpha/(12+6\alpha)}} + \frac{\Psi_{2,\alpha}^{1/3}\Psi_{2,0}^{1/3}(\log p)^{2/3}}{n^{\alpha/(12+6\alpha)}} + \frac{\Phi_{\psi_\nu,\alpha}(\log p)^{1+\nu}}{n^{\alpha/(4+2\alpha)}}$$

CLT for dependent data: physical dependence

- **Idea**: **Large-and-small-blocks method** to create "time gap". **Conditional expectation** and "time gap" to create independency.
- Apply **Large-and-small-blocks** with large size b, and small size m.
- Conditional expectation with time lag m: For each random vector X_t, define $X_t^{(m)} := \mathbb{E}[X_t|\epsilon_t, ..., \epsilon_{t-m}$
- $X_t^{(m)}$ is m-dependent sequence: $X_t^{(m)}$ is independent to $X_{t+m+1}^{(m)}$

CLT for dependent data: physical dependence. Part(ii)

- If $\{X_t\}$ satisfy **Condition 1 and 3**, since $\{X_t^{(m)}\}$ is m-dependent sequence with $m \ll b \ll n$, the result for dependent graph derives **part (ii)** of the theorem.
- If $\{X_t\}$ just satisfy **Condition 3**, without sub-exponential moment bounds, the conditions of the dependent graph theorem don't hold.

CLT for dependent data: physical dependence. Part(i)

Now we don't have exponential moment bound

- Let $\mathcal{F}_t = \sigma\{\epsilon_t, \epsilon_{t-1}, \dots\}$
- Define a projection operator $\mathcal{P}_t(\cdot) := \mathbb{E}[\cdot|\mathcal{F}_t] - \mathbb{E}[\cdot|\mathcal{F}_{t-1}]$. $\{\mathcal{P}_t\}$ is a family of mutually orthogonal operator: $\mathbb{E}[\mathcal{P}_s(X)\mathcal{P}_t(X)] = 0$ if $s \neq t$
- Then $X_t = \sum_{s=0}^{\infty} \mathcal{P}_{t-s}(X_t)$, and $||\mathcal{P}_{t-s}(X_{t,j})||_q \leq ||X_{t,j} - X'_{t,j,\{s\}}||_q \leq \theta_{s,q,j}$

CLT for dependent data: physical dependence. Part(i)

- As a result, though we don't have bounds for exponential moment, we can prove that the probability of "large part" decay in exponentially fast:

$$\mathbb{P}(|\tilde{X}_\ell| > u) \lesssim \exp(-C(u\Phi_{\psi_\nu,\alpha}^{-1})^\gamma)$$

- Large-and-small-blocks technique derives Part(i) of the theorem

Outline

CLT: For simple convex set

Some concepts of simple convex set:

- For closed convex set A, support function: $S_A(x) = \sup_{y \in A} x^T y$. Representation: $A = \cap_{v \in \mathbb{S}^{p-1}} \{x : x^T v \leq S_A(v)\}$
- Polyhedron: $A = \cap_{v \in \mathcal{V}(A)} \{x : x^T v \leq S_A(v)\}$, where $\mathcal{V}(A)$ is a finite subset of $\mathbb{S}^{p-1}$.
- $x \in A \iff v^T x \leq S_A(v)$ for all $v \in \mathcal{V}(A)$
- For each A, corresponding: Random vector $X \in \mathbb{R}^p \mapsto \tilde{X} := (v^T X)_{v \in \mathcal{V}(A)} \in \mathbb{R}^{|\mathcal{V}(A)|}$
- $\mathbb{P}(X \in A) = \mathbb{P}(\tilde{X} \leq (S_A(v))_v)$

CLT: For simple convex set

- fattening or shrinking a convex set: $A^{\epsilon} := \cap_{v \in \mathbb{S}^{p-1}} \{x : x^T v \leq S_A(v) + \epsilon\}$
- $A^K \subset A \subset A^{K,\epsilon}$: A can be approximated by a polyhedron A^K with $|\mathcal{V}(A)| = K$ with precision ϵ.
- A family of subset $\mathcal{A}^{si}(a, d) := \{A$ convex : A can be approximated by A^K satisfying $K \leq (pn)^d$ with precision $\epsilon = a/n\}$

CLT: For simple convex set

Let $A \in \mathcal{A}^{si}(a, d)$,
$\rho^K := |\mathbb{P}(S_{n,X} \in A^K) - \mathbb{P}(G \in A^K)|$, $\rho^{K,\epsilon} := |\mathbb{P}(S_{n,X} \in A^{K,\epsilon}) - \mathbb{P}(G \in A^{K,\epsilon})|$, then

$$|\mathbb{P}(S_{n,X} \in A) - \mathbb{P}(G \in A)| \leq \epsilon(\log K)^{1/2} + \rho^K + \rho^{K,\epsilon}$$

CLT: For simple convex set

$\mathcal{A} \subset \mathcal{A}^{si}(a, d)$ such that $\tilde{X}_i$ satisfies condition **CI1,2** uniformly for all $A \in \mathcal{A}$

- **(Independent)** (i)Suppose that $||v^T X_i||_{\psi_1} \leq B_n$ for all $i \in [n], v \in \mathcal{V}(A)$,then

$$\rho_n(\mathcal{A}) \lesssim \frac{B_n^{1/3} \log^{7/6}(pn)}{n^{1/6}}$$

(ii)Suppose that $||v^T X_i||_{\psi_\gamma} \leq B_n$ for all $i \in [n], v \in \mathcal{V}(A)$, and some $\gamma \geq 4$, then

$$\rho_n(\mathcal{A}^{re}) \lesssim \frac{B_n^{1/3} \log^{7/6}(pn)}{n^{1/6}} + \frac{B_n^{2/3} \log(p)}{n^{(\gamma-2)/3\gamma}}$$

CLT: For simple convex set

$\mathcal{A} \subset \mathcal{A}^{si}(a, d)$ such that $\tilde{X}_i$ satisfies **Condition 1,3** uniformly for all $A \in \mathcal{A}$

- **(Dependent, α-mixing)** If X_i also s.t. **Condition 2**($\Rightarrow \tilde{X}_i$ s.t.)

$$\rho_n(\mathcal{A}) \lesssim \frac{a(d \log p)^{1/2}}{n} + \frac{B_n^{2/3}(d \log p)^{(1+2\gamma_2)/(3\gamma_2)}}{n^{1/9}} + \frac{B_n(d \log p)^{7/6}}{n^{1/9}}$$

  provided that $(d \log p)^{3-\gamma_2} = o(n^{\gamma_2/3})$

- **(Dependent, Dependent graph)**

$$\rho_n(\mathcal{A}) \lesssim \frac{a(d \log p)^{1/2}}{n} + \frac{B_n\left(D_n D_n^*\right)^{1/3}(d \log p)^{7/6}}{n^{1/6}}$$

CLT: For simple convex set

For each polyhedron A , define $\Psi_{q,\alpha}(A)$, $\Phi_{\psi_\nu,\alpha}(A)$ with respect to $\{\tilde{X}_t\}$. Let $\Psi_{q,\alpha,\mathcal{A}} := \sup_{A\in\mathcal{A}} \Psi_{q,\alpha}(A^K)$, $\Phi_{\psi_\nu,\alpha,\mathcal{A}} := \sup_{A\in\mathcal{A}} \Phi_{\psi_\nu,\alpha}(A^K)$

- **(Dependent, Physical dependence)**(i) If $\mathcal{A}$ satisfies **Condition 3**, and $\Phi_{\psi_\nu,\alpha,\mathcal{A}} < \infty$,

$$\rho_n(\mathcal{A}) \lesssim \frac{a(d\log p)^{1/2}}{n} + \frac{\Phi_{\psi_\nu,0,\mathcal{A}}(d\log p)^{7/6}}{n^{\alpha/(3+9\alpha)}} + \frac{\Psi_{2,\alpha,\mathcal{A}}^{1/3}\Psi_{2,0,\mathcal{A}}^{1/3}(d\log p)^{2/3}}{n^{\alpha/(3+9\alpha)}} + \frac{\Phi_{\psi_\nu,\alpha,\mathcal{A}}(d\log p)^{1+\nu}}{n^{\alpha/(1+3\alpha)}}$$

  provided that $(d\log p)^{\max\{6\nu-1,(5+6\nu)/4\}} = o\left\{n^{\alpha/(1+3\alpha)}\right\}$

  (ii) If $\mathcal{A}$ satisfies **Condition 1,3**, and $\Phi_{\psi_\nu,\alpha,\mathcal{A}} < \infty$,

$$\rho_n(\mathcal{A}) \lesssim \frac{a(d\log p)^{1/2}}{n} + \frac{B_n(d\log p)^{7/6}}{n^{\alpha/(12+6\alpha)}} + \frac{\Psi_{2,\alpha,\mathcal{A}}^{1/3}\Psi_{2,0,\mathcal{A}}^{1/3}(d\log p)^{2/3}}{n^{\alpha/(12+6\alpha)}} + \frac{\Phi_{\psi_\nu,\alpha,\mathcal{A}}(d\log p)^{1+\nu}}{n^{\alpha/(4+2\alpha)}}$$

CLT: For sparsely convex set

- A is a s-sparsely convex set:
  - (i): Sparse representation $A = \cap_{q=1}^{K*} A_q$ for convex sets $A_1, ..., A_{K*}$.
  - (ii): Indicator function $I_{\{w \in A_q\}}$ of each A_q depends on at most s components of w.

CLT: For sparsely convex set

- A is a s-sparsely convex set:
  - (i): Sparse representation $A = \cap_{q=1}^{K*} A_q$ for convex sets $A_1, ..., A_{K*}$.
  - (ii): Indicator function $I_{\{w \in A_q\}}$ of each A_q depends on at most s components of w.
- **(Condition 6)** There exist a universal constant $K_5 > 0$ such that $\mathrm{Var}(v^T S_{n,X}) \geq K_5$ for each $v \in \mathbb{S}^{p-1}$ with $|v|_0 \leq s$
- $\mathcal{A}^{sp}(s) := \{A : A \text{ is s-sparsely convex set } \}$

CLT: For sparsely convex set

- Let A is a s-sparsely convex set, $B := \{w : ||w||_\infty \leq pn^{5/2}\}$, $A^* = A \cap B$.
- "small" far away from origin:

$$\mathbb{P}(S_{n,X} \in B^c) \lesssim B_n n^{-1}, \mathbb{P}(G \in B^c) \lesssim B_n n^{-1}$$

- $|\mathbb{P}(S_{n,X} \in A) - \mathbb{P}(G \in A)| \lesssim |\mathbb{P}(S_{n,X} \in A^*) - \mathbb{P}(G \in A^*)| + B_n n^{-1}$
- $A^* = A \cap B$ is also a s-sparsely convex set with $A^* = \cap_{q=1}^{K*} A_q^*$

CLT: For sparsely convex set

$\mathcal{A}_1^{sp}(s) := \{A \in \mathcal{A}^{sp}(s) : A \subset B, A$ contains a ball with radius $n^{-1}\}$,

- **Lemma D.1**[4]: $\mathcal{A}_1^{sp}(s) \subset \mathcal{A}^{si}(1, Cs^2)$. Besides, the approximating polyhedron A^K can be chosen to satisfy $|v|_0 \leq s$ for all $v \in \mathcal{V}(A^K)$.
- **Case 1**: If $A^* \in \mathcal{A}_1^{sp}(s)$, by **Lemma D.1**, we can apply the result for simple convex sets.

[4] Victor Chernozukov, Denis Chetverikov, and Kengo Kato. "Central Limit Theorems and Bootstrap in High Dimensions". In: *The Annals of Probability* 45 (Dec. 2014). DOI: 10.1214/16-AOP1113.

CLT: For sparsely convex set

- **Case 3**: $A^* \notin \mathcal{A}_1^{sp}(s)$, and each A_q^* contains a ball with radius n^{-1}, but they are not all the same.
- Note that A_q^* itself is also s-sparsely. We can apply **Lemma D.1** for each A_q^* such that: $A_q^{K_q} \subset A_q^* \subset A_q^{K_q, \frac{1}{n}}, \forall q \in [K_*]$
- Let $Q := \cap_{q=1}^{K_*} A_q^{K_q}$, then $Q^{\frac{1}{n}} := \cap_{q=1}^{K_*} A_q^{K_q, \frac{1}{n}}$, and

$$\varnothing = Q^{-\frac{1}{n}} \subset Q \subset A^* \subset Q^{\frac{1}{n}}$$

.

CLT: For sparsely convex set

- Nazarov's inequality gives the bound that

$$\mathbb{P}(G \in A^*) \leq \mathbb{P}(G \in Q^{\frac{1}{n}} \backslash Q^{-\frac{1}{n}}) \lesssim n^{-1}(s^2 \log p)^{1/2}$$

- By the result for simple convex set, we can get the estimation of $\mathbb{P}(S_{n,X} \in A^*)$.
- $|\mathbb{P}(S_{n,X} \in A^*) - \mathbb{P}(G \in A^*)| \leq \mathbb{P}(S_{n,X} \in A^*) + \mathbb{P}(G \in A^*)$

CLT: For sparsely convex set

Case 2(Difficult): $A^* \notin \mathcal{A}_1^{sp}(s)$, and at least 1 A_q don't contain a ball with radius n^{-1}.

- We need to derive the Berry-Esseen type estimate for convex sets:

$$\sup_{h\in\mathcal{H}} |\mathbb{E}[h(S_{n,X})] - \mathbb{E}[h(G)]| = \sup_{h\in\mathcal{H}} |\int h(x)d(F_n(x) - \Phi(x))|$$

  where $\mathcal{H}$ is the family of indicator function of convex sets.

- Detailed derivation can see[5], section 9. Motivation can see[6], chapter 7.

[5] Jinyuan Chang, Xiaohui Chen, and Mingcong Wu. "Central limit theorems for high dimensional dependent data". In: *Bernoulli* 30.1 (Feb. 2024). ISSN: 1350-7265. DOI: 10.3150/23-bej1614. URL: http://dx.doi.org/10.3150/23-BEJ1614.

[6] Rabi N. Bhattacharya and R. Ranga Rao. *Normal Approximation and Asymptotic Expansions*. Society for Industrial and Applied Mathematics, 2010. DOI: 10.1137/1.9780898719895. URL: https://epubs.siam.org/doi/abs/10.1137/1.9780898719895.

Outline

Parametric Bootstrap

In theory, $G \sim N(0, \Xi)$ depends on covariance matrix Ξ. However, covariance matrix Ξ is unknown in practice. As a result, we need to construct a series of matrix $\{\hat{\Xi}_n\}$ to approximate Ξ

- Let $\mathcal{X}_n = \{X_1, X_2, ..., X_n\}$ a data set, data-dependent Gaussian random vector $\hat{G}|\mathcal{X}_n \sim N(0, \hat{\Xi}_n)$.
- To establish theoretical validity, we have to bound

$$\hat{\rho}_n(\mathcal{A}) := \sup_{A \in \mathcal{A}} |\mathbb{P}(S_{n,X} \in A) - \mathbb{P}(\hat{G} \in A|\mathcal{X}_n)|$$

Parametric Bootstrap

- $\hat{\rho}_n(\mathcal{A}) \leq \rho_n(\mathcal{A}) + \sup_{A \in \mathcal{A}} |\mathbb{P}(\hat{G} \in A | \mathcal{X}_n) - \mathbb{P}(G \in A)|$
- The problem is boiled down to estimate the difference between covariance matrix.
- $\Delta_{n,r} := |\hat{\Xi}_n - \Xi|_\infty$
- By Gaussian comparison, we have

$$\hat{\rho}_n(\mathcal{A}^{\mathrm{re}}) \lesssim \rho_n(\mathcal{A}^{\mathrm{re}}) + \Delta_{n,r}^{1/3} (\log p)^{2/3}.$$

- Hope: Select proper approximation matrices $\hat{\Xi}_n$ such that $\Delta_{n,r} = o_{\mathrm{p}}\{(\log p)^{-2}\}$

Parametric Bootstrap

- we suggest to adopt the kernel-type estimator for its long-run covariance matrix, that is

$$\hat{\Xi}_n = \sum_{j=-n+1}^{n-1} \mathcal{K}\left(\frac{j}{b_n}\right) \hat{H}_j$$

where $\hat{H}_j = n^{-1} \sum_{t=j+1}^{n} (X_t - \bar{X})(X_{t-j} - \bar{X})^\top$ if $j \geq 0$ and $\hat{H}_j = n^{-1} \sum_{t=-j+1}^{n} (X_{t+j} - \bar{X})(X_t - \bar{X})^\top$ if $j < 0$ with optimal kernel

$$\mathcal{K}_{\mathrm{QS}}(x) = \frac{25}{12\pi^2 x^2} \left\{ \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right\}$$

Parametric Bootstrap

- With assumption above, we have

$$\Delta_{n,r} = O_{\mathrm{p}}\{B_n^2 n^{-c_1}(\log p)^{c_2}\} + O(B_n^2 n^{-\rho})$$

- we see that our proposed parametric bootstrap procedure is asymptotically valid even if the dimension p grows sub-exponentially fast in the sample size n

Reference

[1] Rabi N. Bhattacharya and R. Ranga Rao. *Normal Approximation and Asymptotic Expansions*. Society for Industrial and Applied Mathematics, 2010. DOI: 10.1137/1.9780898719895. URL: https://epubs.siam.org/doi/abs/10.1137/1.9780898719895.

[2] Jinyuan Chang, Xiaohui Chen, and Mingcong Wu. "Central limit theorems for high dimensional dependent data". In: *Bernoulli* 30.1 (Feb. 2024). ISSN: 1350-7265. DOI: 10.3150/23-bej1614. URL: http://dx.doi.org/10.3150/23-BEJ1614.

[3] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors". In: *The Annals of Statistics* 41.6 (Dec. 2013). ISSN: 0090-5364. DOI: 10.1214/13-aos1161. URL: http://dx.doi.org/10.1214/13-AOS1161.

[4] Victor Chernozukov, Denis Chetverikov, and Kengo Kato. "Central Limit Theorems and Bootstrap in High Dimensions". In: *The Annals of Probability* 45 (Dec. 2014). DOI: 10.1214/16-AOP1113.