Bootstrapping high dimensional vector: interplay between dependence and dimensionality

Xianyang Zhang
Joint work with Guang Cheng

University of Missouri-Columbia

LDHD: Transition Workshop, 2014
Overview

- Let $x_1, x_2, \ldots, x_n$ be a sequence of mean-zero dependent random vectors in $\mathbb{R}^p$, where $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ with $1 \leq i \leq n$.

- We provide a general (non-asymptotic) theory for quantifying:
  \[
  \rho_n := \sup_{t \in \mathbb{R}} \left| P(T_X \leq t) - P(T_Y \leq t) \right|,
  \]
  where $T_X = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{ij}$ and $T_Y = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_{ij}$ with $y_i = (y_{i1}, y_{i2}, \ldots, y_{ip})'$ being a Gaussian vector.

- Key techniques: Slepian interpolation and the leave-one-block out argument (modification of Stein's leave-one-out method).

- Two examples on inference for high dimensional time series.
Outline

1. Inference for high dimensional time series
   - Uniform confidence band for the mean
   - Specification testing on the covariance structure

2. Gaussian approximation for maxima of non-Gaussian sum
   - $M$-dependent time series
   - Weakly dependent time series

3. Bootstrap
   - Blockwise multiplier bootstrap
   - Non-overlapping block bootstrap
Example I: Uniform confidence band

- Consider a $p$-dimensional *weakly dependent* time series $\{x_i\}$.

- **Goal:** construct a uniform confidence band for $\mu_0 = EX_i \in \mathbb{R}^p$ based on the observations $\{x_i\}_{i=1}^n$ with $n \ll p$.

- Consider the $(1 - \alpha)$ confidence band:

$$\left\{ \mu = (\mu_1, \ldots, \mu_p)' \in \mathbb{R}^p : \sqrt{n} \max_{1 \leq j \leq p} |\mu_j - \bar{x}_j| \leq c(\alpha) \right\},$$

where $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_p)' = \sum_{i=1}^n x_i/n$ is the sample mean.

- **Question:** how to obtain the critical value $c(\alpha)$?
Blockwise Multiplier Bootstrap

- Capture the dependence *within* and *between* the data vectors.

- Suppose \( n = b_n l_n \) with \( b_n, l_n \in \mathbb{Z} \). Define the block sum
  \[
  A_{ij} = \sum_{l=(i-1)b_n+1}^{ib_n} (x_{lj} - \bar{x}_j), \quad i = 1, 2, \ldots, l_n.
  \]

- When \( p = O(\exp(n^b)) \), \( b_n = O(n^{b'}) \) with \( 4b' + 7b < 1 \) and \( b' > 2b \).

- Define the bootstrap statistic,
  \[
  T_A = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{l_n} A_{ij} e_i \right|,
  \]
  where \( \{e_i\} \) is a sequence of i.i.d \( N(0, 1) \) random variables that are independent of \( \{x_i\} \).

- Compute \( c(\alpha) := \inf\{ t \in \mathbb{R} : P(T_A \leq t|\{x_i\}_{i=1}^n) \geq \alpha \} \).
Some numerical results

Consider a $p$-dimensional VAR(1) process,

$$x_t = \rho x_{t-1} + \sqrt{1 - \rho^2} \epsilon_t.$$

1. $\epsilon_{tj} = (\epsilon_{tj} + \epsilon_{t0})/\sqrt{2}$, where $(\epsilon_{t0}, \epsilon_{t1}, \ldots, \epsilon_{tp}) \sim \text{i.i.d } N(0, I_{p+1})$;

2. $\epsilon_{tj} = \rho_1 \zeta_{tj} + \rho_2 \zeta_{t(j+1)} + \cdots + \rho_p \zeta_{t(j+p-1)}$, where $\{\rho_j\}_{j=1}^p$ are generated independently from $U(2, 3)$, and $\{\zeta_{tj}\}$ are i.i.d $N(0, 1)$ random variables;

3. $\epsilon_{tj}$ is generated from the moving average model above with $\{\zeta_{tj}\}$ being i.i.d centralized Gamma$(4, 1)$ random variables.
Some numerical results (Con’t)

**Table:** Coverage probabilities of the uniform confidence band, where \( n = 120. \)

<table>
<thead>
<tr>
<th>( \rho ) = 0.3</th>
<th>( b_n = 4 )</th>
<th>( 95% )</th>
<th>( 99% )</th>
<th>( b_n = 6 )</th>
<th>( 95% )</th>
<th>( 99% )</th>
<th>( b_n = 8 )</th>
<th>( 95% )</th>
<th>( 99% )</th>
<th>( b_n = 10 )</th>
<th>( 95% )</th>
<th>( 99% )</th>
<th>( b_n = 12 )</th>
<th>( 95% )</th>
<th>( 99% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>89.7</td>
<td>97.2</td>
<td>90.5</td>
<td>97.5</td>
<td>90.1</td>
<td>97.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>92.5</td>
<td>98.3</td>
<td>91.6</td>
<td>97.8</td>
<td>91.6</td>
<td>97.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94.6</td>
<td>99.0</td>
<td>91.5</td>
<td>97.6</td>
<td>92.4</td>
<td>97.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95.0</td>
<td>99.2</td>
<td>91.8</td>
<td>97.8</td>
<td>91.6</td>
<td>97.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94.8</td>
<td>99.3</td>
<td>91.3</td>
<td>97.9</td>
<td>92.0</td>
<td>97.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \rho ) = 0.5</td>
<td>( b_n = 4 )</td>
<td>( 76.9 )</td>
<td>( 92.9 )</td>
<td>( 83.5 )</td>
<td>( 94.0 )</td>
<td>( 83.3 )</td>
<td>( 93.7 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>87.1</td>
<td>96.3</td>
<td>87.3</td>
<td>96.2</td>
<td>87.4</td>
<td>95.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91.6</td>
<td>98.3</td>
<td>88.8</td>
<td>96.6</td>
<td>89.4</td>
<td>96.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>92.5</td>
<td>98.6</td>
<td>89.8</td>
<td>97.1</td>
<td>89.3</td>
<td>97.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>93.0</td>
<td>99.0</td>
<td>90.0</td>
<td>97.2</td>
<td>90.5</td>
<td>97.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example II: Specification testing on the covariance structure

For a mean-zero $p$-dimensional time series $\{x_i\}$, define
$$\Gamma(h) = Ex_{i+h}x_i' \in \mathbb{R}^{p \times p}.$$ 

Consider
$$H_0 : \Gamma(h) = \tilde{\Gamma}(h) \text{ versus } H_a : \Gamma(h) \neq \tilde{\Gamma}(h),$$
for some $h \in \Lambda \subseteq \{0, 1, 2, \ldots \}$.

Special cases:

1. $\Lambda = \{0\}$: testing the covariance structure. See Cai and Jiang (2011), Chen et al. (2010), Li and Chen (2012) and Qiu and Chen (2012) for some developments when $\{x_i\}$ are i.i.d.

2. $\Lambda = \{1, 2, \ldots, H\}$ and $\tilde{\Gamma}(h) = 0$ for $h \in \Lambda$: white noise testing.
Testing for white noise

- Consider the white noise testing problem. Our test is given by
  \[ T = \sqrt{n} \max_{1 \leq h \leq H} \max_{1 \leq j, k \leq p} |\hat{\gamma}_{jk}(h)|, \]
  where \( \hat{\Gamma}(h) = \sum_{i=1}^{n-h} x_{i+h}x_i' / n = (\hat{\gamma}_{jk}(h))_{j,k=1}^p. \)

- Let \( z_i = (z_{i,1}, \ldots, z_{i,p^2H}) = (\text{vec}(x_{i+1}x_i'), \ldots, \text{vec}(x_{i+H}x_i'))' \in \mathbb{R}^{p^2H} \) for \( i = 1, \ldots, N := n - H. \)

- Suppose \( N = b_n l_n \) for \( b_n, l_n \in \mathbb{Z}. \) Define
  \[ T_A = \max_{1 \leq j \leq p^2H} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{l_n} A_{ij} e_i \right|, \quad A_{ij} = \sum_{l=(i-1)b_n+1}^{ib_n} (z_{l,j} - \bar{z}_j), \]
  where \( \{e_i\} \) is a sequence of i.i.d \( N(0, 1) \) random variables that are independent of \( \{x_i\}, \) and \( \bar{z}_j = \sum_{i=1}^{N} z_{i,j} / n. \)

- Compute \( c(\alpha) := \inf \{ t \in \mathbb{R} : P(T_A \leq t|\{x_i\}_{i=1}^n) \geq \alpha \}, \) and reject the white noise null hypothesis if \( T > c(\alpha). \)
Some numerical results

We are interested in testing,

$$H_0 : \Gamma(h) = 0, \quad \text{for } 1 \leq h \leq L,$$

versus

$$H_a : \Gamma(h) \neq 0, \quad \text{for some } 1 \leq h \leq L.$$

Consider the following data generating processes:

1. multivariate normal: $x_{tj} = \rho_1 \zeta_{tj} + \rho_2 \zeta_{t(j+1)} + \cdots + \rho_p \zeta_{t(j+p-1)}$, where $\{\rho_j\}_{j=1}^p$ are generated independently from $U(2, 3)$, and $\{\zeta_{tj}\}$ are i.i.d $N(0, 1)$ random variables;

2. multivariate ARCH model: $x_t = \Sigma_t^{1/2} \epsilon_t$ with $\epsilon_t \sim N(0, I_p)$ and $\Sigma_t = 0.1 I_p + 0.9 x_{t-1} x'_{t-1}$, where $\Sigma_t^{1/2}$ is a lower triangular matrix based on the Cholesky decomposition of $\Sigma_t$;

3. VAR(1) model: $x_t = \rho x_{t-1} + \sqrt{1 - \rho^2} \epsilon_t$, where $\rho = 0.2$ and the errors $\{\epsilon_t\}$ are generated according to $\overline{1}$. 
Some numerical results (Con’t)

**Table:** Rejection percentages for testing the uncorrelatedness, where \( n = 240 \) and the actual number of parameters is \( p^2 \times L \).

<table>
<thead>
<tr>
<th></th>
<th>( p = 20, 1 )</th>
<th>( p = 20, 2 )</th>
<th>( p = 20, 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 5% )</td>
<td>( 1% )</td>
<td>( 5% )</td>
<td>( 1% )</td>
</tr>
<tr>
<td>( L = 1 )</td>
<td>( b_n = 1 )</td>
<td>4.3 0.8 2.8 0.3</td>
<td>90.3 71.9</td>
</tr>
<tr>
<td></td>
<td>( b_n = 4 )</td>
<td>5.0 1.0 1.0 0.3</td>
<td>86.3 63.3</td>
</tr>
<tr>
<td></td>
<td>( b_n = 8 )</td>
<td>5.3 1.2 1.6 0.9</td>
<td>86.0 59.2</td>
</tr>
<tr>
<td></td>
<td>( b_n = 12 )</td>
<td>5.1 1.0 2.3 1.4</td>
<td>86.5 59.2</td>
</tr>
<tr>
<td>( L = 3 )</td>
<td>( b_n = 1 )</td>
<td>4.7 1.0 2.3 0.3</td>
<td>79.4 57.7</td>
</tr>
<tr>
<td></td>
<td>( b_n = 4 )</td>
<td>3.6 0.7 0.6 0.3</td>
<td>74.0 46.2</td>
</tr>
<tr>
<td></td>
<td>( b_n = 8 )</td>
<td>3.7 0.4 1.3 0.8</td>
<td>71.4 41.0</td>
</tr>
<tr>
<td></td>
<td>( b_n = 12 )</td>
<td>4.0 0.6 2.2 1.3</td>
<td>72.1 40.6</td>
</tr>
</tbody>
</table>
Maxima of non-Gaussian sum

- The above applications hinge on a general theoretical result.

- Let \( x_1, x_2, \ldots, x_n \) be a sequence of mean-zero dependent random vectors in \( \mathbb{R}^p \), where \( x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})' \) with \( 1 \leq i \leq n \).

- **Target:** approximate the distribution of

\[
T_X = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{ij}.
\]
Gaussian approximation

Let \( y_1, y_2, \ldots, y_n \) be a sequence of mean-zero Gaussian random vectors in \( \mathbb{R}^p \), where \( y_i = (y_{i1}, y_{i2}, \ldots, y_{ip})' \) with \( 1 \leq i \leq n \).

Suppose that \( \{y_i\} \) preserves the autocovariance structure of \( \{x_i\} \), i.e.,

\[
\text{cov}(y_i, y_j) = \text{cov}(x_i, x_j).
\]

Goal: quantify the Kolmogrov distance

\[
\rho_n := \sup_{t \in \mathbb{R}} |P(T_X \leq t) - P(T_Y \leq t)|,
\]

where \( T_Y = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_{ij} \).
Existing results in the independent case

**Question:** how large $p$ can be in relation with $n$ so that $\rho_n \to 0$?

- Bentkus (2003): $\rho_n \to 0$ provided that $p^{7/2} = o(n)$.

- Chernozhukov et al. (2013): $\rho_n \to 0$ if $p = O(\exp(n^b))$ with $b < 1/7$ (an astounding improvement).

**Motivation:** study the interplay between the dependence structure and the growth rate of $p$ so that $\rho_n \to 0$. 

Dependence Structure I: $M$-dependent time series

- A time series $\{x_i\}$ is called $M$-dependent if for $|i - j| > M$, $x_i$ and $x_j$ are independent.

- Under suitable restrictions on the tail of $x_i$ and weak dependence assumptions uniformly across the components of $x_i$, we show that

$$\rho_n \lesssim \frac{M^{1/2} (\log(pn/\gamma) \vee 1)^{7/8}}{n^{1/8}} + \gamma,$$

for some $\gamma \in (0, 1)$.

- When $p = O(\exp(n^b))$ for $b < 1/11$, and $M = O(n^{b'})$ with $4b' + 7b < 1$, we have

$$\rho_n \leq Cn^{-c}, \quad c, C > 0.$$

- If $b' = 0$ (i.e., $M = O(1)$), our result allows $b < 1/7$ [Chernozhukov et al. (2013)].
Dependence Structure II: Physical dependence measure [Wu (2005)]

- The sequence \( \{x_i\} \) has the following causal representation,

\[
x_i = G(\ldots, \epsilon_{i-1}, \epsilon_i),
\]

where \( G \) is a measurable function and \( \{\epsilon_i\} \) is a sequence of i.i.d random variables.

- Let \( \{\epsilon_i'\} \) be an i.i.d copy of \( \{\epsilon_i\} \) and define

\[
x_i^* = G(\ldots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_i).
\]

- The strength of the dependence can be quantified via

\[
\theta_{i,j,q}(x) = (E|x_{ij} - x_{ij}^*|^q)^{1/q}, \quad \Theta_{i,j,q}(x) = \sum_{l=i}^{+\infty} \theta_{l,j,q}(x).
\]
Bound on the Kolmogrov distance

Theorem

Under suitable conditions on the tail of \( \{x_i\} \) and certain weak dependence assumptions, we have

\[
\rho_n \lesssim n^{-1/8} M^{1/2} l_n^{7/8} + (n^{1/8} M^{-1/2} l_n^{-3/8}) \frac{q}{1+q} \left( \sum_{j=1}^{p} \Theta_{M,j,q} \right)^{\frac{1}{1+q}} + \gamma,
\]

where \( \Theta_{i,j,q} = \Theta_{i,j,q}(x) \lor \Theta_{i,j,q}(y) \).

- The tradeoff between the first two terms reflects the interaction between the dimensionality and dependence;

- Key step in the proof: \( M \)-dependent approximation.
The bound on the Kolmogorov distance (Con’t)

**Corollary**

Suppose that

1. \( \max_{1 \leq j \leq p} \Theta_{M,j,q} = O(\rho^M) \) for \( \rho < 1 \) and \( q \geq 2 \);
2. \( \rho = O(\exp(n^b)) \) for \( 0 < b < 1/11 \).

Then we have

\[
\rho_n \leq C n^{-c}, \quad c, C > 0.
\]
Dimension free dependence structure

Question: is there any so-called “dimension free dependence structure”? What kind of dependence assumption will not affect the increase rate of \( p \)?

- For a permutation \( \pi(\cdot) \), \( (x_{i\pi(1)}, \ldots, x_{i\pi(p)}) = (z_1, z_2) \).

- Suppose \( \{z_i\} \) is a \( s \)-dimensional time series and \( \{z_i\} \) is a \( p-s \) dimensional sequence of independent variables.

- Assume that \( \{z_i\} \) and \( \{z_i\} \) are independent, and \( s/p \to 0 \).

- Under suitable assumptions, it can be shown that for \( p = O(\exp(n^b)) \) with \( b < 1/7 \),

\[
\rho_n \leq Cn^{-c}, \quad c, C > 0.
\]
Resampling

Summary: for $M$-dependent or more generally weakly dependent time series, we have shown that

$$\rho_n := \sup_{t \in \mathbb{R}} \left| P(T_X \leq t) - P(T_Y \leq t) \right| \leq Cn^{-c}, \quad c, C > 0.$$  

**Question:** in practice the autocovariance structure of $\{x_i\}$ is typically unknown. How can we approximate the distribution of $T_X$ or $T_Y$?

**Solution:** Resampling method.
Blockwise multiplier bootstrap

1. Suppose \( n = b_n l_n \). Compute the block sum,

\[
A_{ij} = \sum_{l=(i-1)b_n+1}^{ib_n} x_{lj}, \quad i = 1, 2, \ldots, l_n.
\]

2. Generate a sequence of i.i.d \( N(0, 1) \) random variables \( \{e_i\} \) and compute

\[
T_A = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{i=1}^{l_n} A_{ij} e_i.
\]

3. Repeat step 2 several times and compute the \( \alpha \)-quantile of \( T_A \)

\[
c_{T_A}(\alpha) = \inf\{t \in \mathbb{R} : P(T_A \leq t | \{x_i\}_{i=1}^n) \geq \alpha\}.
\]
Validity of the blockwise multiplier bootstrap

**Theorem**

Under suitable assumptions, we have for \( p = O(\exp(n^b)) \) with \( 0 < b < 1/15 \),

\[
\sup_{\alpha \in (0,1)} \left| P\left( T_X \leq c_{T_A}(\alpha) \right) - \alpha \right| \lesssim n^{-c}, \quad c > 0.
\]
Non-overlapping block bootstrap

1. Let $A_{1j}^*, \ldots, A_{lnj}^*$ be an i.i.d draw from the empirical distribution of
   $\{A_{ij}\}_{i=1}^{ln}$ and compute

   \[
   T_{A^*} = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{i=1}^{ln} (A_{ij}^* - \bar{A}_j), \quad \bar{A}_j = \sum_{i=1}^{ln} A_{ij} / l_n.
   \]

2. Repeat the above step several times to obtain the $\alpha$-quantile of $T_{A^*}$,

   \[
   c_{T_{A^*}}(\alpha) = \inf\{t \in \mathbb{R} : P(T_{A^*} \leq t|\{x_i\}_{i=1}^{n}) \geq \alpha\}.
   \]

**Theorem**

Under suitable assumptions, we have with probability $1 - o(1)$,

\[
\sup_{\alpha \in (0,1)} \left| P(T_X \leq c_{T_{A^*}}(\alpha)|c_{T_{A^*}}(\alpha)) - \alpha \right| = o(1).
\]
Future works

1. Choice of the block size in the blockwise multiplier bootstrap and non-overlapping block bootstrap;

Thank you!