Bootstrapping Spectral Statistics in High Dimensions

Miles Lopes

UC Davis

joint work with Andrew Blandino and Alex Aue

Bootstrap for sample covariance matrices

- Suppose we have i.i.d. observations $X_1, \ldots, X_n \in \mathbb{R}^p$, and let $\widehat{\Sigma}$ be the sample covariance matrix.
- Let $T = \varphi(\widehat{\Sigma})$ denote a statistic of interest.
- We would like to estimate $\sqrt{\mathrm{var}(T)}$, or more generally, approximate the sampling distribution of T.
- The non-parametric bootstrap offers a general way to solve these problems.

Bootstrap for sample covariance matrices

- Suppose we have i.i.d. observations $X_1, \dots, X_n \in \mathbb{R}^p$, and let $\widehat{\Sigma}$ be the sample covariance matrix.
- Let $T = \varphi(\widehat{\Sigma})$ denote a statistic of interest.
- We would like to estimate $\sqrt{\text{var}(T)}$, or more generally, approximate the sampling distribution of T.
- The non-parametric bootstrap offers a general way to solve these problems.

Non-parametric bootstrap.

For: $b = 1, \dots, B$:

- Sample n points $X_1^*, \dots, X_n^*$ with replacement from $\{X_1, \dots, X_n\}$.
- Form the sample covariance matrix $\widehat{\Sigma}^*$ associated with $X_1^*, \dots, X_n^*$.
- Compute $T_b^* := \varphi(\widehat{\Sigma}^*)$.

Return: the empirical distribution of $T_1^*, \dots, T_B^*$.

Some past work and recent developments

In 1985, Beran and Srivastava showed that the standard bootstrap generally works for smooth functionals of $\widehat{\Sigma}$ when $p \ll n$. (Exceptions arise for non-smooth functionals, or tied population eigenvalues.)

The paper Hall, Lee, Park, Paul (2009) develops a remedy for tied eigenvalues, as well as a generalization to functional data. (A good literature survey is also provided for many other papers in the $p \ll n$ setting.)

When $p \asymp n$, relatively little is known about bootstrap consistency.

Recently, El Karoui and Purdom (2016), have studied the non-parametric bootstrap, and have demonstrated some negative empirical results for $\lambda_1(\widehat{\Sigma})$. They also prove bootstrap consistency for a fixed number of the largest sample eigenvalues when Σ is effectively low-dimensional (e.g. when the true spectrum decays).

Difficulties in high dimensions

- Why do difficulties arise when p is large?

Difficulties in high dimensions

- Why do difficulties arise when p is large?

- If it were possible, we would prefer to draw an i.i.d. sample from the (unknown) distribution $\mathbb{P}$ underlying $\mathcal{D} = \{X_1, \ldots, X_n\}$.

Difficulties in high dimensions

- Why do difficulties arise when p is large?
- If it were possible, we would prefer to draw an i.i.d. sample from the (unknown) distribution $\mathbb{P}$ underlying $\mathcal{D} = \{X_1, \dots, X_n\}$.
- Instead, the bootstrap uses an i.i.d. sample from the empirical distribution $\widehat{\mathbb{P}}$, which places mass $1/n$ at each point in $\mathcal{D}$.

Difficulties in high dimensions

- Why do difficulties arise when p is large?
- If it were possible, we would prefer to draw an i.i.d. sample from the (unknown) distribution $\mathbb{P}$ underlying $\mathcal{D} = \{X_1, \ldots, X_n\}$.
- Instead, the bootstrap uses an i.i.d. sample from the empirical distribution $\widehat{\mathbb{P}}$, which places mass $1/n$ at each point in $\mathcal{D}$.
- Key issue: If p is large, then $\widehat{\mathbb{P}}$ is often a poor substitute for $\mathbb{P}$.

Alternatives to the non-parametric bootstrap?

- The empirical distribution $\widehat{\mathbb{P}}$ can fail because it “tries too hard” to approximate the “entire” distribution $\mathbb{P}$.

Alternatives to the non-parametric bootstrap?

- The empirical distribution $\widehat{\mathbb{P}}$ can fail because it "tries too hard" to approximate the "entire" distribution $\mathbb{P}$.

- We only need to approximate $\mathbb{P}$ with respect to the parameters that are "relevant" for our statistic.

Alternatives to the non-parametric bootstrap?

- The empirical distribution $\widehat{\mathbb{P}}$ can fail because it "tries too hard" to approximate the "entire" distribution $\mathbb{P}$.

- We only need to approximate $\mathbb{P}$ with respect to the parameters that are "relevant" for our statistic.

- For certain classes of statistics T, we can rely on probability theory to determine the relevant parameters (i.e. invariance principles or universality results).

Alternatives to the non-parametric bootstrap?

- The empirical distribution $\widehat{\mathbb{P}}$ can fail because it "tries too hard" to approximate the "entire" distribution $\mathbb{P}$.
- We only need to approximate $\mathbb{P}$ with respect to the parameters that are "relevant" for our statistic.
- For certain classes of statistics T, we can rely on probability theory to determine the relevant parameters (i.e. invariance principles or universality results).
- Then, we can sample from a suitable parametric distribution $\mathbb{P}_{\widehat{\boldsymbol{\theta}}}$.

Alternatives to the non-parametric bootstrap?

- The empirical distribution $\widehat{\mathbb{P}}$ can fail because it "tries too hard" to approximate the "entire" distribution $\mathbb{P}$.
- We only need to approximate $\mathbb{P}$ with respect to the parameters that are "relevant" for our statistic.
- For certain classes of statistics T, we can rely on probability theory to determine the relevant parameters (i.e. invariance principles or universality results).
- Then, we can sample from a suitable parametric distribution $\mathbb{P}_{\widehat{\boldsymbol{\theta}}}$.

  (This is similar to the "parametric bootstrap", but not quite the same, since we are not trying to model all of $\mathbb{P}$.)

Alternatives to the non-parametric bootstrap?

- The empirical distribution $\widehat{\mathbb{P}}$ can fail because it "tries too hard" to approximate the "entire" distribution $\mathbb{P}$.
- We only need to approximate $\mathbb{P}$ with respect to the parameters that are "relevant" for our statistic.
- For certain classes of statistics T, we can rely on probability theory to determine the relevant parameters (i.e. invariance principles or universality results).
- Then, we can sample from a suitable parametric distribution $\mathbb{P}_{\widehat{\boldsymbol{\theta}}}$.

  (This is similar to the "parametric bootstrap", but not quite the same, since we are not trying to model all of $\mathbb{P}$.)

Question: This viewpoint can be used to understand "spectral statistics" and "max statistics" in high dimensions. Perhaps there are more examples?

A basic model for studying covariance matrices

Let $\Sigma \in \mathbb{R}^{p \times p}$ be a population covariance matrix.

Suppose $X \in \mathbb{R}^{n \times p}$ is a data matrix with i.i.d. rows generated as

$$X_i = \Sigma^{1/2} Z_i \tag{1}$$

where the vectors $Z_1, \dots, Z_n \in \mathbb{R}^p$ have i.i.d. entries with $\mathbb{E}[Z_{ij}] = 0$, $\mathbb{E}[Z_{ij}^2] = 1$ and $\mathbb{E}[Z_{ij}^4] = \kappa$.

Define the sample covariance matrix

$$\widehat{\Sigma} = \frac{1}{n} X^\top X. \tag{2}$$

Linear Spectral Statistics (LSS)

A natural class of prototype statistics for investigating bootstrap consistency are *linear spectral statstics*, which have the form

$$T = \tfrac{1}{p} \sum_{j=1}^{p} f(\lambda_j(\widehat{\Sigma})), \tag{3}$$

where f is a smooth function on $[0, \infty)$.

Linear Spectral Statistics (LSS)

A natural class of prototype statistics for investigating bootstrap consistency are *linear spectral statstics*, which have the form

$$T = \tfrac{1}{p} \textstyle\sum_{j=1}^{p} f(\lambda_j(\widehat{\Sigma})), \tag{3}$$

where f is a smooth function on $[0, \infty)$.

Examples:

- The choice $f(x) = \log(x)$ leads to $\log(\det(\widehat{\Sigma}))$.
- The choice $f(x) = x^k$, leads to $\mathrm{tr}(\widehat{\Sigma}^k)$ (cf. Schatten norms)
- The normal log-likelihood ratio statistic for testing sphericity is

$$p \log(\mathrm{tr}(\widehat{\Sigma})) - \log(\det(\widehat{\Sigma})).$$

- Raj Rao et al (2008) developed testing procedures based on $\mathrm{tr}(\widehat{\Sigma}^k)$.
- Various tests of sphericity are "asymptotically equivalent" to transformations of LSS (Dobriban 2016)

Background ideas for developing a new bootstrap

Bai and Silverstein established an important CLT for LSS in 2004. (See also Jonsson 1982.)

If $\mathbb{E}[Z_{ij}^4] = 3$, and $p/n \to c \in (0, \infty)$, then under "standard assumptions" we have

$$p(T - \mathbb{E}[T]) \xrightarrow{w} N(0, \sigma^2).$$

Background ideas for developing a new bootstrap

Bai and Silverstein established an important CLT for LSS in 2004. (See also Jonsson 1982.)

If $\mathbb{E}[Z_{ij}^4] = 3$, and $p/n \to c \in (0, \infty)$, then under "standard assumptions" we have

$$p(T - \mathbb{E}[T]) \xrightarrow{w} N(0, \sigma^2).$$

Note: The normalization is $\mathcal{O}(p)$ rather than $\mathcal{O}(\sqrt{n})$.

Background ideas for developing a new bootstrap

Bai and Silverstein established an important CLT for LSS in 2004. (See also Jonsson 1982.)

If $\mathbb{E}[Z_{ij}^4] = 3$, and $p/n \to c \in (0, \infty)$, then under "standard assumptions" we have

$$p(T - \mathbb{E}[T]) \xrightarrow{w} N(0, \sigma^2).$$

Note: The normalization is $\mathcal{O}(p)$ rather than $\mathcal{O}(\sqrt{n})$.

$$\sigma^2 = \frac{-1}{2\pi^2} \oint\oint \frac{f(z_1)f(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} \frac{d}{dz_1}\underline{m}(z_1) \frac{d}{dz_2}\underline{m}(z_2) dz_1 dz_2$$

Background ideas for developing a new bootstrap

$$p(T - \mathbb{E}[T]) \xrightarrow{w} N(0, \sigma^2).$$

Important property: Under the setup of Bai and Silverstein (2004), the parameter σ only depends on the limiting spectral distribution of Σ.

So, roughly speaking, this result says that under certain conditions, the laws of LSS are asymptotically governed by just the eigenvalues $\Lambda = \text{diag}(\lambda_1(\Sigma), \ldots, \lambda_p(\Sigma))$, rather than the entire matrix Σ.

This is a major reduction in complexity.

A "parametric bootstrap" approach

More good news: The eigenvalues Λ can be estimated well in high dimensions — under modest assumptions.

(Sparsity and/or low-rank conditions are not needed; cf. Ledoit and Wolf (2015), Kong and Valiant (2017), Chaudhuri, Jun, and Paul (Friday), among others.)

A "parametric bootstrap" approach

More good news: The eigenvalues Λ can be estimated well in high dimensions — under modest assumptions.

(Sparsity and/or low-rank conditions are not needed; cf. Ledoit and Wolf (2015), Kong and Valiant (2017), Chaudhuri, Jun, and Paul (Friday), among others.)

Intuitive procedure: Generate a "new dataset" X^* that nearly matches the observed data X with respect to Λ.

Then, we just compute the statistic T^* arising from the "new data" X^*.

(This is akin to a parametric bootstrap.)

A “parametric bootstrap” approach

More good news: The eigenvalues Λ can be estimated well in high dimensions — under modest assumptions.

(Sparsity and/or low-rank conditions are not needed; cf. Ledoit and Wolf (2015), Kong and Valiant (2017), Chaudhuri, Jun, and Paul (Friday), among others.)

Intuitive procedure: Generate a “new dataset” X^* that nearly matches the observed data X with respect to Λ.

Then, we just compute the statistic T^* arising from the “new data” X^*.

(This is akin to a parametric bootstrap.)

One extra detail: The kurtosis $\kappa = \mathbb{E}[Z_{ij}^4]$ matters too, but that's ok.

Proposed method: Spectral Bootstrap

Goal. Approximate the distribution of $T = \frac{1}{p}\sum_{j=1}^{p} f(\lambda_j(\widehat{\Sigma}))$.

Proposed method: Spectral Bootstrap

Goal. Approximate the distribution of $T = \frac{1}{p}\sum_{j=1}^{p} f(\lambda_j(\widehat{\Sigma}))$.

Before resampling, first compute estimates $\widehat{\kappa}$ and $\widehat{\Lambda}$.

Proposed method: Spectral Bootstrap

Goal. Approximate the distribution of $T = \frac{1}{p}\sum_{j=1}^{p} f(\lambda_j(\widehat{\Sigma}))$.

Before resampling, first compute estimates $\widehat{\kappa}$ and $\widehat{\Lambda}$.

Algorithm. (Spectral Bootstrap)

For $b = 1, \ldots, B$:

Proposed method: Spectral Bootstrap

Goal. Approximate the distribution of $T = \frac{1}{p}\sum_{j=1}^{p} f(\lambda_j(\widehat{\Sigma}))$.

Before resampling, first compute estimates $\widehat{\kappa}$ and $\widehat{\Lambda}$.

Algorithm. (Spectral Bootstrap)

For $b = 1, \ldots, B$:

- Generate a random matrix $Z^* \in \mathbb{R}^{n \times p}$ whose entries Z^*_{ij} are drawn i.i.d. from Pearson$(0, 1, 0, \widehat{\kappa})$. (Recall $X_i = \Sigma^{1/2} Z_i$)

Proposed method: Spectral Bootstrap

Goal. Approximate the distribution of $T = \frac{1}{p}\sum_{j=1}^{p} f(\lambda_j(\widehat{\Sigma}))$.

Before resampling, first compute estimates $\widehat{\kappa}$ and $\widehat{\Lambda}$.

Algorithm. (Spectral Bootstrap)

For $b = 1, \ldots, B$:

- Generate a random matrix $Z^* \in \mathbb{R}^{n\times p}$ whose entries Z^*_{ij} are drawn i.i.d. from Pearson$(0, 1, 0, \widehat{\kappa})$. (Recall $X_i = \Sigma^{1/2} Z_i$)
- Compute $\widehat{\Sigma}^* := \frac{1}{n}\widehat{\Lambda}^{1/2}(Z^{*\top} Z^*)\widehat{\Lambda}^{1/2}$. (Note $\widehat{\Sigma} = \frac{1}{n}\Sigma^{1/2} Z^\top Z \Sigma^{1/2}$)

Proposed method: Spectral Bootstrap

Goal. Approximate the distribution of $T = \frac{1}{p}\sum_{j=1}^{p} f(\lambda_j(\widehat{\Sigma}))$.

Before resampling, first compute estimates $\widehat{\kappa}$ and $\widehat{\Lambda}$.

Algorithm. (Spectral Bootstrap)

For $b = 1, \ldots, B$:

- Generate a random matrix $Z^* \in \mathbb{R}^{n\times p}$ whose entries Z^*_{ij} are drawn i.i.d. from Pearson$(0, 1, 0, \widehat{\kappa})$. (Recall $X_i = \Sigma^{1/2} Z_i$)
- Compute $\widehat{\Sigma}^* := \frac{1}{n}\widehat{\Lambda}^{1/2}(Z^{*\top} Z^*)\widehat{\Lambda}^{1/2}$. (Note $\widehat{\Sigma} = \frac{1}{n}\Sigma^{1/2} Z^\top Z \Sigma^{1/2}$)
- Compute the eigenvalues of $\widehat{\Sigma}^*$, and denote them by $(\lambda_1^*, \ldots, \lambda_p^*)$.

Proposed method: Spectral Bootstrap

Goal. Approximate the distribution of $T = \frac{1}{p}\sum_{j=1}^{p} f(\lambda_j(\widehat{\Sigma}))$.

Before resampling, first compute estimates $\widehat{\kappa}$ and $\widehat{\Lambda}$.

Algorithm. (Spectral Bootstrap)

For $b = 1, \dots, B$:

- Generate a random matrix $Z^* \in \mathbb{R}^{n\times p}$ whose entries Z^*_{ij} are drawn i.i.d. from Pearson$(0, 1, 0, \widehat{\kappa})$. (Recall $X_i = \Sigma^{1/2} Z_i$)
- Compute $\widehat{\Sigma}^* := \frac{1}{n}\widehat{\Lambda}^{1/2}(Z^{*\top} Z^*)\widehat{\Lambda}^{1/2}$. (Note $\widehat{\Sigma} = \frac{1}{n}\Sigma^{1/2} Z^\top Z \Sigma^{1/2}$)
- Compute the eigenvalues of $\widehat{\Sigma}^*$, and denote them by $(\lambda_1^*, \dots, \lambda_p^*)$.
- Compute the statistic, $T_b^* := \frac{1}{p}\sum_{j=1}^{p} f(\lambda_j^*)$

Return: the empirical distribution of the values $T_1^*, \dots, T_B^*$.

Generalizing to other spectral statistics

Let $\psi : \mathbb{R}^p \to \mathbb{R}$ be a generic (non-linear) function, and consider the statistic

$$T = \psi(\lambda_1(\widehat{\Sigma}), \ldots, \lambda_p(\widehat{\Sigma})).$$

Key point: To bootstrap T, we only need change the last step.

(This is a distinct benefit of the bootstrap in relation to formulas.)

Generalizing to other spectral statistics

Let $\psi : \mathbb{R}^p \to \mathbb{R}$ be a generic (non-linear) function, and consider the statistic

$$T = \psi(\lambda_1(\widehat{\Sigma}), \dots, \lambda_p(\widehat{\Sigma})).$$

Key point: To bootstrap T, we only need change the last step.

(This is a distinct benefit of the bootstrap in relation to formulas.)

For $b = 1, \dots, B$:

- $\dots$
- Compute the eigenvalues of $\widehat{\Sigma}^*$, and denote them by $(\lambda_1^*, \dots, \lambda_p^*)$.
- Compute the statistic, $T_b^* := \psi(\lambda_1^*, \dots, \lambda_p^*)$

Return: the empirical distribution of the values $T_1^*, \dots, T_B^*$.

Estimating kurtosis

Recall $\kappa = \mathbb{E}[Z_{ij}^4]$, and all row vectors satisfy $X_i = \Sigma^{1/2} Z_i$.

Our estimate of κ is based on a general formula for the variance of a quadratic form

$$\kappa = 3 + \frac{\mathsf{Var}(\|X_1\|_2^2) - 2\|\Sigma\|_F^2}{\sum_{j=1}^p \sigma_j^4}.$$

It turns out that all the quantities on the right side have ratio-consistent estimators when $p \asymp n$ under standard conditions.

Estimating kurtosis (cont.)

Recall the kurtosis formula

$$\kappa = 3 + \frac{\mathrm{Var}(\|X_1\|_2^2) - 2\|\Sigma\|_F^2}{\sum_{j=1}^p \sigma_j^4}.$$

Remarks.

- Estimating $\|\Sigma\|_F^2$ is somewhat tricky because the naive estimate is biased, and so the bias must be corrected. It is known from Bai and Saranadasa (1996) that as $(n, p) \to \infty$

$$\frac{\|\widehat{\Sigma}\|_F^2 - \frac{1}{n}\mathrm{tr}(\widehat{\Sigma})^2}{\|\Sigma\|_F^2} = 1 + o_{\mathbb{P}}(1).$$

Estimating kurtosis (cont.)

Recall the kurtosis formula

$$\kappa = 3 + \frac{\mathrm{Var}(\|X_1\|_2^2) - 2\|\Sigma\|_F^2}{\sum_{j=1}^p \sigma_j^4}.$$

Remarks.

- Estimating $\|\Sigma\|_F^2$ is somewhat tricky because the naive estimate is biased, and so the bias must be corrected. It is known from Bai and Saranadasa (1996) that as $(n, p) \to \infty$

$$\frac{\|\widehat{\Sigma}\|_F^2 - \frac{1}{n}\mathrm{tr}(\widehat{\Sigma})^2}{\|\Sigma\|_F^2} = 1 + o_{\mathbb{P}}(1).$$

- Also, it can be shown that

$$\frac{\sum_{j=1}^p \left(\frac{1}{n}\sum_{i=1}^n X_{ij}^2\right)^2}{\sum_{j=1}^p \sigma_j^4} = 1 + o_{\mathbb{P}}(1).$$

Estimating eigenvalues

We use the QUEST method (Ledoit and Wolf, 2015).

Let H_p denote the spectral the distribution function associated with $\lambda_1(\Sigma), \ldots, \lambda_p(\Sigma)$,

$$H_p(t) := \frac{1}{p} \sum_{j=1}^{p} 1\{\lambda_j(\Sigma) \leq t\}.$$

The QUEST method provides an estimate $\widehat{H}_p$ of this distribution.

In turn, the quantiles of $\widehat{H}_p$ can be used to estimate $\lambda_1(\Sigma), \ldots, \lambda_p(\Sigma)$.

Estimating eigenvalues

We use the QUEST method (Ledoit and Wolf, 2015).

Let H_p denote the spectral the distribution function associated with $\lambda_1(\Sigma), \dots, \lambda_p(\Sigma)$,

$$H_p(t) := \frac{1}{p} \sum_{j=1}^{p} 1\{\lambda_j(\Sigma) \leq t\}.$$

The QUEST method provides an estimate $\widehat{H}_p$ of this distribution.

In turn, the quantiles of $\widehat{H}_p$ can be used to estimate $\lambda_1(\Sigma), \dots, \lambda_p(\Sigma)$.

Consistency. If there is a limiting population distribution H such that

$$H_p \xrightarrow{w} H,$$

then the QUEST estimator is consistent in the sense that (under standard assumptions)

$$\widehat{H}_p \xrightarrow{w} H \quad \text{almost surely.}$$

Estimating eigenvalues

We use the QUEST method (Ledoit and Wolf, 2015).

Let H_p denote the spectral the distribution function associated with $\lambda_1(\Sigma), \dots, \lambda_p(\Sigma)$,

$$H_p(t) := \frac{1}{p} \sum_{j=1}^{p} 1\{\lambda_j(\Sigma) \leq t\}.$$

The QUEST method provides an estimate $\widehat{H}_p$ of this distribution.

In turn, the quantiles of $\widehat{H}_p$ can be used to estimate $\lambda_1(\Sigma), \dots, \lambda_p(\Sigma)$.

Consistency. If there is a limiting population distribution H such that

$$H_p \xrightarrow{w} H,$$

then the QUEST estimator is consistent in the sense that (under standard assumptions)

$$\widehat{H}_p \xrightarrow{w} H \quad \text{almost surely.}$$

The QUEST method also performs well in practice.

Main result: bootstrap consistency

Main assumptions for bootstrap consistency.

- $p/n \to c \in (0, \infty)$
- $\lambda_p(\Sigma)$ and $\lambda_1(\Sigma)$ bounded away from 0 and ∞
- Finite 8th moment: $\mathbb{E}[Z_{11}^8] < \infty$.
- $H_p \xrightarrow{w} H$
- Asymptotic "regularity" of population eigenvectors (more later)

Main result: consistency of spectral bootstrap

Let d_{LP} be the Lévy-Prohorov metric on distributions.

Note: Convergence in d_{LP} is equivalent to weak convergence.

Theorem 1 (LBA 2019, consistency of spectral bootstrap)

Under the stated assumptions, as $(n, p) \to \infty$,

$$d_{LP}\Big(\mathcal{L}\big(p(T^* - \mathbb{E}^*[T^*])|X\big)\,,\,\mathcal{L}\big(p(T - \mathbb{E}[T])\big)\Big) \to 0 \quad \textit{in} \quad \mathbb{P}_X\textit{-probability.}$$

Main result: consistency of spectral bootstrap

Let d_{LP} be the Lévy-Prohorov metric on distributions.

Note: Convergence in d_{LP} is equivalent to weak convergence.

Theorem 1 (LBA 2019, consistency of spectral bootstrap)

Under the stated assumptions, as $(n, p) \to \infty$,

$$d_{LP}\Big(\mathcal{L}\big(p(T^* - \mathbb{E}^*[T^*])|X\big)\,,\, \mathcal{L}\big(p(T - \mathbb{E}[T])\big)\Big) \to 0 \quad \textit{in} \quad \mathbb{P}_X\textit{-probability.}$$

Regularity of eigenvectors

Let U be the matrix of eigenvectors of Σ, and consider the non-random quantity

$$\Delta_p(z_1, z_2) := \frac{1}{p} \sum_{j=1}^{p} \left[U D_n(z_1) U^\top \right]_{jj} \left[U D_n(z_2) U^\top \right]_{jj}$$

where $z_1, z_2 \in \mathbb{C} \setminus \mathbb{R}$, and $D_n(\cdot) \in \mathbb{C}^{p \times p}$ is a diagonal matrix that only depends on the spectrum of Σ.

Regularity of eigenvectors

Let U be the matrix of eigenvectors of Σ, and consider the non-random quantity

$$\Delta_p(z_1, z_2) := \frac{1}{p} \sum_{j=1}^{p} \left[U D_n(z_1) U^\top \right]_{jj} \left[U D_n(z_2) U^\top \right]_{jj}$$

where $z_1, z_2 \in \mathbb{C} \setminus \mathbb{R}$, and $D_n(\cdot) \in \mathbb{C}^{p \times p}$ is a diagonal matrix that only depends on the spectrum of Σ.

Also define

$$\Delta_p'(z_1, z_2) := \frac{1}{p} \sum_{j=1}^{p} \left[D_n(z_1) \right]_{jj} \left[D_n(z_2) \right]_{jj}.$$

Regularity of eigenvectors

Let U be the matrix of eigenvectors of Σ, and consider the non-random quantity

$$\Delta_p(z_1, z_2) := \frac{1}{p} \sum_{j=1}^{p} \left[U D_n(z_1) U^\top \right]_{jj} \left[U D_n(z_2) U^\top \right]_{jj}$$

where $z_1, z_2 \in \mathbb{C} \setminus \mathbb{R}$, and $D_n(\cdot) \in \mathbb{C}^{p \times p}$ is a diagonal matrix that only depends on the spectrum of Σ.

Also define

$$\Delta'_p(z_1, z_2) := \frac{1}{p} \sum_{j=1}^{p} \left[D_n(z_1) \right]_{jj} \left[D_n(z_2) \right]_{jj}.$$

Regularity of eigenvectors. We say that the eigenvectors of Σ are regular if for any $z_1, z_2 \in \mathbb{C} \setminus \mathbb{R}$, as $(n, p) \to \infty$

$$\Delta_p(z_1, z_2) = \Delta'_p(z_1, z_2) + o(1).$$

Regularity of eigenvectors

Let U be the matrix of eigenvectors of Σ, and consider the non-random quantity

$$\Delta_p(z_1, z_2) := \frac{1}{p}\sum_{j=1}^{p}\left[UD_n(z_1)U^\top\right]_{jj}\left[UD_n(z_2)U^\top\right]_{jj}$$

where $z_1, z_2 \in \mathbb{C}\setminus\mathbb{R}$, and $D_n(\cdot) \in \mathbb{C}^{p\times p}$ is a diagonal matrix that only depends on the spectrum of Σ.

Also define

$$\Delta_p'(z_1, z_2) := \frac{1}{p}\sum_{j=1}^{p}\left[D_n(z_1)\right]_{jj}\left[D_n(z_2)\right]_{jj}.$$

Regularity of eigenvectors. We say that the eigenvectors of Σ are regular if for any $z_1, z_2 \in \mathbb{C}\setminus\mathbb{R}$, as $(n, p) \to \infty$

$$\Delta_p(z_1, z_2) = \Delta_p'(z_1, z_2) + o(1).$$

Note: This assumption is not needed when $\kappa = 3$. Empirical results suggest that this assumption is not too much of a concern.

Regularity of eigenvectors (cont.)

Example 1. (Rank k perturbations, $k \to \infty$).

Suppose $\lambda_1(\Sigma)$ is bounded away from ∞, and let Λ be otherwise unrestricted.

If U is of the form

$$U = I_{p \times p} - 2\Pi,$$

where Π is any orthogonal projection matrix of rank k, and $k = o(p)$, then the eigenvectors are regular.

This is a fairly substantial perturbation from the diagonal case.

Regularity of eigenvectors (cont.)

Example 2. (Spiked covariance models).

Suppose Λ is of the form

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_k, 1, \ldots, 1)$$

where $k = o(p)$, and $\lambda_1 = \lambda_1(\Sigma)$ is bounded away from infinity.

Then, any choice of U will be regular.

Comments on proof ideas

The proof builds on recent work by Najim and Yao (2016), which develops a CLT for LSS of the form

$$d_{\mathrm{LP}}\Big(\mathcal{L}(p(T-\mathbb{E}[T]))\ ,\ \mathcal{L}(V_n)\Big) \to 0,$$

where V_n is a Gaussian rv with the appropriate mean and variance.
The formulation of the CLT in terms of a metric (as opposed to a weak limit) is helpful in analyzing the bootstrap.

Comments on proof ideas

The proof builds on recent work by Najim and Yao (2016), which develops a CLT for LSS of the form

$$d_{\mathrm{LP}}\Big(\mathcal{L}(p(T-\mathbb{E}[T]))\ ,\ \mathcal{L}(V_n)\Big) \to 0,$$

where V_n is a Gaussian rv with the appropriate mean and variance.
The formulation of the CLT in terms of a metric (as opposed to a weak limit) is helpful in analyzing the bootstrap.

Another ingredient worth mentioning is the Helffer-Sjöstrand formula, which shows that LSS can be represented as a linear functional of the empirical Stieltjes transform. Namely, if

$$\widehat{m}_p(z) = \tfrac{1}{p}\mathrm{tr}\big((\widehat{\Sigma} - zI_{p\times p})^{-1}\big),$$

then

$$T = \phi_f(\widehat{m}_p),$$

where ϕ_f is a linear functional depending on f.

Simulations for LSS ($\kappa > 3$)

Recall $X_i = \Sigma^{1/2} Z_i$.

- Z_i generated with standardized i.i.d. t-dist (df=20)
- kurtosis $\kappa \approx 3.4$
- decaying population spectrum is $\lambda_j = j^{-1/2}$
- population eigenvectors uniformly drawn from Haar measure

We tabulate the std. dev., 95th percentile, and 99th percentile of $p(T - \mathbb{E}[T])$.

Simulations for LSS ($\kappa > 3$)

Recall $X_i = \Sigma^{1/2} Z_i$.

- Z_i generated with standardized i.i.d. t-dist (df=20)
- kurtosis $\kappa \approx 3.4$
- decaying population spectrum is $\lambda_j = j^{-1/2}$
- population eigenvectors uniformly drawn from Haar measure

We tabulate the std. dev., 95th percentile, and 99th percentile of $p(T - \mathbb{E}[T])$.

	$f(x) = x$						$f(x) = \log(x)$					
(n,p)	std. dev.		95th		99th		std. dev.		95th		99th	
(500,200)	0.16		0.27		0.36		1.07		1.82		2.41	
	0.17	**(0.01)**	**0.28**	**(0.03)**	**0.39**	**(0.06)**	**1.08**	**(0.08)**	**1.76**	**(0.20)**	**2.51**	**(0.35)**
(500,400)	0.18		0.29		0.41		4.41		7.03		9.77	
	0.18	**(0.02)**	**0.30**	**(0.04)**	**0.42**	**(0.06)**	**4.27**	**(0.33)**	**6.72**	**(0.70)**	**9.29**	**(1.18)**
(500,600)	0.17		0.29		0.40		-		-		-	
	0.18	**(0.02)**	**0.30**	**(0.04)**	**0.43**	**(0.07)**						

Simulations for LSS ($\kappa < 3$)

Recall $X_i = \Sigma^{1/2} Z_i$.

- Z_i generated with standardized i.i.d. Beta(6,6)
- kurtosis $\kappa = 2.6$
- decaying population spectrum is $\lambda_j = j^{-1/2}$
- population eigenvectors uniformly drawn from Haar measure

	$f(x) = x$						$f(x) = \log(x)$					
(n,p)	std. dev.		95th		99th		std. dev.		95th		99th	
(500,200)	0.14		0.23		0.33		0.93		1.51		1.92	
	0.14	(**0.01**)	**0.23**	(**0.03**)	**0.32**	(**0.05**)	**0.93**	(**0.08**)	**1.52**	(**0.17**)	**2.15**	(**0.31**)
(500,400)	0.15		0.25		0.34		1.65		2.64		3.64	
	0.14	(**0.01**)	**0.24**	(**0.03**)	**0.34**	(**0.05**)	**1.70**	(**0.13**)	**2.81**	(**0.31**)	**3.97**	(**0.56**)
(500,600)	0.16		0.26		0.34		-		-		-	
	0.15	(**0.01**)	**0.25**	(**0.03**)	**0.35**	(**0.05**)						

What about other spectral statistics?

In principle, the proposed method can be applied to any other spectral statistic.

Below, we present some simulation results for some *non-linear* statistics:

- $T_{\text{max}} := \lambda_1(\widehat{\Sigma})$.
- $T_{\text{gap}} := \lambda_1(\widehat{\Sigma}) - \lambda_2(\widehat{\Sigma})$

Simulations for non-linear statistics ($\kappa < 3$)

Recall $X_i = \Sigma^{1/2} Z_i$.

- Z_i generated with standardized i.i.d. Beta(6,6)
- kurtosis $\kappa = 2.6$
- decaying population spectrum is $\lambda_j = j^{-1/2}$
- population eigenvectors uniformly drawn from Haar measure

Simulations for non-linear statistics ($\kappa < 3$)

Recall $X_i = \Sigma^{1/2} Z_i$.

- Z_i generated with standardized i.i.d. Beta(6,6)
- kurtosis $\kappa = 2.6$
- decaying population spectrum is $\lambda_j = j^{-1/2}$
- population eigenvectors uniformly drawn from Haar measure

	$T_{\max} - \mathbb{E}[T_{\max}]$						$T_{\text{gap}} - \mathbb{E}[T_{\text{gap}}]$					
(n,p)	std. dev.		95th		99th		std. dev.		95th		99th	
(500,200)	0.06		0.11		0.15		0.08		0.13		0.17	
	0.06	**(0.01)**	**0.09**	**(0.01)**	**0.13**	**(0.02)**	**0.07**	**(0.01)**	**0.11**	**(0.01)**	**0.16**	**(0.03)**
(500,400)	0.06		0.10		0.15		0.08		0.13		0.18	
	0.06	**(0.01)**	**0.09**	**(0.01)**	**0.13**	**(0.02)**	**0.07**	**(0.01)**	**0.11**	**(0.01)**	**0.16**	**(0.03)**
(500,600)	0.06		0.11		0.14		0.07		0.13		0.17	
	0.06	**(0.01)**	**0.09**	**(0.01)**	**0.13**	**(0.02)**	**0.07**	**(0.01)**	**0.11**	**(0.02)**	**0.16**	**(0.03)**

Simulations for non-linear statistics ($\kappa > 3$)

Recall $X_i = \Sigma^{1/2} Z_i$.

- Z_i generated with standardized i.i.d. t-dist (df=20)
- kurtosis $\kappa \approx 3.4$
- decaying population spectrum is $\lambda_j = j^{-1/2}$
- population eigenvectors uniformly drawn from Haar measure

	$T_{\max} - \mathbb{E}[T_{\max}]$						$T_{\text{gap}} - \mathbb{E}[T_{\text{gap}}]$					
(n,p)	std. dev.		95th		99th		std. dev.		95th		99th	
(500,200)	0.06		0.10		0.15		0.07		0.12		0.17	
	0.07	**(0.01)**	**0.11**	**(0.02)**	**0.17**	**(0.03)**	**0.08**	**(0.01)**	**0.13**	**(0.02)**	**0.19**	**(0.03)**
(500,400)	0.06		0.10		0.14		0.07		0.13		0.17	
	0.07	**(0.01)**	**0.11**	**(0.02)**	**0.17**	**(0.03)**	**0.08**	**(0.01)**	**0.13**	**(0.02)**	**0.19**	**(0.03)**
(500,600)	0.06		0.11		0.16		0.08		0.13		0.18	
	0.07	**(0.01)**	**0.11**	**(0.02)**	**0.16**	**(0.03)**	**0.08**	**(0.01)**	**0.13**	**(0.02)**	**0.19**	**(0.03)**

Summary of bootstrap for spectral statistics

- We have identified LSS as a general class of statistics for which bootstrapping can succeed in high dimensions.
- This offers general-purpose way to approximate the laws of LSS without relying on asymptotic formulas.
- The method is akin to the parametric bootstrap — using the fact that spectral statistics may depend on relatively few parameters of the full data-generating distribution.
- Numerical results are encouraging.
- The method appears to extend to some non-linear spectral statistics — for which asymptotic formulas are often unavailable.
- Further work on non-linear statistics is underway . . .

paper available at

- https://arxiv.org/abs/1709.08251
- to appear in Biometrika

Thanks to

- you for your attention
- NSF grant DMS-1613218