Large-dimensional Central Limit Theorem with Fourth-moment Error Bounds on Convex Sets and Balls

Zhi Zhang

2021/3/3

Topics

- Some results on the large-dimensional Gaussian approximation of a sum of n independent random vectors in $\mathbb{R}^d$ together with fourth-moment error bounds on convex sets and Euclidean balls.
- Application to the bootstrap: Applied the bounds we obtained to the bootstrap approximation on balls.

Outline

Introduction and Motivation

- Let $\{\xi_i\}_{i=1}^n$ be a sequence of independent mean-zero random vectors in $\mathbb{R}^d$, $W = \sum_{i=1}^n \xi_i$ and $\Sigma = \mathrm{Var}(W)$.
- It is well known that under finite third-moment conditions and for fixed dimension d, the distribution of W can be approximated by a Gaussian distribution with error rate $O(1/\sqrt{n})$.
- Motivated by modern statistical applications, we are interested in the large-dimensional setting where d grows with n. Numerous studies have provided explicit error bounds on various distributional distances in the Gaussian approximation.

Introduction and Motivation

- However, the optimal rates, especially in terms of how rapidly d can grow with n while maintaining the validity of the Gaussian approximation, have not been fully addressed and remain a challenging open problem.
- For convex sets, Bentkus (2005) proved for the above W that if Σ is invertible and $Z \sim N(0, \Sigma)$, then

$$\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leqslant C d^{1/4} \sum_{i=1}^{n} \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^3 \quad (1)$$

where $\mathcal{A}$ is the collection of all measurable convex sets in $\mathbb{R}^d$, C is an absolute constant and $|\cdot|$ denotes the Euclidean norm when applied to a vector.

Introduction and Motivation

- The first main result is that up to a logarithmic factor,

$$\sup_{A\in\mathcal{A}} |\mathbb{P}(W \in A)-\mathbb{P}(Z \in A)| \leqslant_{\log} Cd^{1/4} \left(\sum_{i=1}^{n} \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right)^{1/2} \tag{2}$$

  where $\mathcal{A}$ is the collection of all measurable convex sets in $\mathbb{R}^d$. And they derive the similar result in the case of Euclidean balls.
- These two results share some advantages over Bentkus's works and have some applications on bootstraps approximation on balls.

Notations

- For two vectors $x, y \in \mathbb{R}^d$, $x \cdot y$ denotes their inner product. For two $d \times d$ matrices M and N, we write $\langle M, N \rangle_{H.S.}$ for their Hilbert-Schmidt inner product.

$$\langle M, N \rangle_{\mathrm{H.S.}} = \mathrm{tr}\left(M^T N \right) = \sum_i \langle Me_i, Ne_i \rangle$$

  where $\{e_i : i \in I\}$ an orthonormal basis.
- We write ∇f and $\mathrm{Hess} f$ for the gradient and Hessian matrix of f, respectively. In addition, we denote by $\nabla^r f(x)$ the r-th derivative of f at x regarded as an r-linear form: The value of $\nabla^r f(x)$ evaluated at $u_1, \dots, u_r \in \mathbb{R}^d$ is given by

$$\langle \nabla^r f(x), u_1 \otimes \cdots \otimes u_r \rangle = \sum_{j_1, \dots, j_r = 1}^{d} \partial_{j_1, \dots, j_r} f(x) u_{1,j_1} \cdots u_{r,j_r}$$

  When $u_1 = \cdots = u_r =: u$, we write $u_1 \otimes \cdots \otimes u_r = u^{\otimes r}$ for short.

Notations

- For any r -linear form T, its injective norm is defined by

$$|T|_{\vee} := \sup_{|u_1|\vee \ldots \vee |u_r| \leqslant 1} |\langle T, u_1 \otimes \cdots \otimes u_r \rangle|$$

- For an $(r-1)$ -times differentiable function $h : \mathbb{R}^d \to \mathbb{R}$, we write

$$M_r(h) := \sup_{x \neq y} \frac{\left|\nabla^{r-1} h(x) - \nabla^{r-1} h(y)\right|_{\vee}}{|x-y|}$$

- Note that $M_r(h) = \sup_{x \in \mathbb{R}^d} |\nabla^r h(x)|_{\vee}$ if h is r-times differentiable.

Main Theorem

Theorem (2.1)

Let $\xi = \{\xi_i\}_{i=1}^n$ be a sequence of centered independent random vectors in $\mathbb{R}^d$ with finite fourth moments and set $W = \sum_{i=1}^n \xi_i$. Assume $\mathrm{Var}(W) = \Sigma$ and Σ is invertible. Let $Z \sim N(0, \Sigma)$ be a centered Gaussian vector in $\mathbb{R}^d$ with covariance matrix Σ. Then,

$$\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \\ \leqslant C d^{1/4} \left(\sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right)^{1/2} \left(\left| \log \left(\sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right) \right| \vee 1 \right) \tag{3}$$

where $\mathcal{A}$ is the collection of all measurable convex sets in $\mathbb{R}^d$.

Remark of Theorem 2.1

Consider the situation where $\xi_i = X_i/\sqrt{n}$ and $\{X_1, X_2, \ldots\}$ is a sequence of i.i.d. mean-zero random vectors in $\mathbb{R}^d$ with $\operatorname{Var}(X_i) = I_d$. In this setting, $\Sigma = I_d$, and for the d-vector X_i, we have

$$\mathbb{E}\,|X_i|^3 \propto d^{3/2} \qquad \mathbb{E}\,|X_i|^4 \propto d^2$$

- RHS of (3) in Theorem 2.1 is of the order $O\left(\frac{d^{5/2}}{n}\right)^{1/2}$ up to a logarithmic factor.
- RHS of (1) in Bentkus's work is of the order $O\left(\frac{d^{7/2}}{n}\right)^{1/2}$

Therefore, subject to the requirement of the existence of the fourth moment, (3) is preferable to (1) in the large-dimensional setting where $d \to \infty$.

Theorem (2.2)

Let $\xi = \{\xi_i\}_{i=1}^n$ be a sequence of centered independent random vectors in $\mathbb{R}^d$ with finite fourth moments and set $W = \sum_{i=1}^n \xi_i$. Let $Z \sim N(0, \Sigma)$ be a centered Gaussian vector in $\mathbb{R}^d$ with covariance matrix Σ. Assume Σ is invertible. Then

$$\sup_{A \in \mathcal{B}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leqslant C\Psi(\delta(W, \Sigma)) \tag{4}$$

where $\Psi(x) = x(|\log x| \vee 1), \mathcal{B}$ is the set of all Euclidean balls in $\mathbb{R}^d$ and

$$\delta(W, \Sigma) := \left\| I_d - \mathrm{Var}\left(\Sigma^{-1/2} W\right) \right\|_{H.S.} + \left(\sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right)^{1/2}$$

Main Theorem

Theorem (2.3)

Let ξ, W and Z be as in Theorem 2.2. Assume $\operatorname{tr}\left(\Sigma^2\right) > 0$. *Then*

$$\sup_{A \in \mathcal{B}} \left|\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)\right| \leqslant \frac{C}{\operatorname{tr}\left(\Sigma^2\right)^{1/4}} \sqrt{\tilde{\delta}(W, \Sigma)} \qquad (5)$$

where

$$\tilde{\delta}(W, \Sigma) := \|\Sigma - \operatorname{Var}(W)\|_{H.S.} + \sum_{j=1}^{d} \left|\Sigma_{jj} - \operatorname{Var}\left(W_j\right)\right| + \sqrt{\sum_{i=1}^{n} \mathbb{E}\left|\xi_i\right|^4} + \sum_{j=1}^{d} \sqrt{\sum_{i=1}^{n} \mathbb{E}\left[\xi_{ij}^4\right]}$$

Remark

Under the same setting in the previous remark, note that

$$\mathbb{E}\left|\xi_i\right|^4 \leqslant d \sum_{j=1}^{d} \mathbb{E}\xi_{ij}^4$$

if $\operatorname{Var}(W) = \Sigma = I_d$, the RHS of (5) in Thm 2.3 is bounded by

$$C \max_{1 \leqslant j \leqslant d} \left(d \sum_{i=1}^{n} \mathbb{E}\xi_{ij}^4 \right)^{1/4}$$

If $\max_{1 \leqslant i \leqslant n} \max_{1 \leqslant j \leqslant d} \left(\mathbb{E}\xi_{ij}^4\right)^{1/4} = O(1/\sqrt{n})$ as $n \to \infty$, the RHS of Thm 2.3 is of order $O(\frac{d}{n})^{\frac{1}{4}}$. This converges to 0 as long as $d/n \to 0$.

Sketch of the Proof

Basic Decomposition and Sketch proof of Thm 2.1

Basic Decomposition

- The proof for Theorem 2.1 starts with approximating the indicator function 1_A for a convex set A by an appropriate smooth function h. Then, the problem amounts to establishing an appropriate bound for $\mathbb{E}h(W) - \mathbb{E}h(Z)$.
- To accomplish this, we will make use of a decomposition of $\mathbb{E}h(W) - \mathbb{E}h(Z)$ derived from the exchangeable pair approach in Stein's method for multivariate normal approximation by Chatterjee and Meckes (2008)

Stein's Equation

Lemma (cf. Götze (1991) and Meckes (2009))

Given a twice differentiable function $h : \mathbb{R}^d \to \mathbb{R}$ with bounded partial derivatives, we consider the Stein equation

$$\langle \mathrm{Hess}\, f(w), \Sigma \rangle_{H.S.} - w \cdot \nabla f(w) = h(w) - \mathbb{E}h(Z), \quad w \in \mathbb{R}^d \tag{6}$$

then

$$f(w) = \int_0^1 -\frac{1}{2(1-s)} \int_{\mathbb{R}^d} \left[h\left(\sqrt{1-s}w + \sqrt{s}\Sigma^{1/2} z \right) - \mathbb{E}h(Z) \right] \phi(z) dz ds \tag{7}$$

is a solution to (6).

Basic Decomposition

- The basic decomposition assumes that f is thrice differentiable with bounded partial derivatives. This is true if Σ is invertible or h is thrice differentiable with bounded partial derivatives.
- Let $\{\xi_1', \ldots, \xi_n'\}$ be an independent copy of $\{\xi_1, \ldots, \xi_n\}$, and let I be a random index uniformly chosen from $\{1, \ldots, n\}$ and independent of $\{\xi_1, \ldots, \xi_n, \xi_1', \ldots, \xi_n'\}$. Define

$$W' = W - \xi_I + \xi_I'$$

It is easy to verify that (W, W') has the same distribution as (W', W) (exchangeability) and

$$\mathbb{E}\left(W' - W \mid W\right) = -\frac{W}{n} \tag{8}$$

Basic Decomposition

From exchangeability and (8) we have, with $D = W' - W$

$$\begin{aligned} 0 &= \frac{n}{2}\mathbb{E}\left[D \cdot \left(\nabla f\left(W'\right) + \nabla f(W)\right)\right] \\ &= \mathbb{E}\left[\frac{n}{2} D \cdot \left(\nabla f\left(W'\right) - \nabla f(W)\right) + nD \cdot \nabla f(W)\right] \\ &= \mathbb{E}\left[\frac{n}{2}\sum_{j,k=1}^{d} D_j D_k \partial_{jk} f(W) + R_2 + nD \cdot \nabla f(W)\right] \\ &= \mathbb{E}\left[\langle \operatorname{Hess} f(W), \Sigma\rangle_{H.S.} - R_1 + R_2 - W \cdot \nabla f(W)\right] \end{aligned} \tag{9}$$

Basic Decomposition

where

$$R_1 = \sum_{j,k=1}^{d} \mathbb{E}\left\{\left(\Sigma_{jk} - \frac{n}{2} D_j D_k\right) \partial_{jk} f(W)\right\} \tag{10}$$

and

$$R_2 = \frac{n}{2} \sum_{j,k,l=1}^{d} \mathbb{E} D_j D_k D_l U \partial_{jkl} f(W + (1-U)D) \tag{11}$$

and U is a uniform random variable on $[0, 1]$ independent of everything else. From (6) and (9) we have

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = R_1 - R_2 \tag{12}$$

Basic Decomposition

We further rewrite R_1 and R_2 respectively as follows (this requires some complicated calculation). First, set

$$V = (V_{jk})_{1 \leqslant j,k \leqslant d} := \left(\mathbb{E} \left[\Sigma_{jk} - \frac{n}{2} D_j D_k \mid \xi \right] \right)_{1 \leqslant j,k \leqslant d}$$

Then we evidently have

$$R_1 = \sum_{j,k=1}^{d} \mathbb{E} V_{jk} \partial_{jk} f(W) = \mathbb{E} \langle V, \text{ Hess } f(W) \rangle_{H.S.} \tag{13}$$

Also, one can verify that (cf. Eq.(22) of Chernozhukov, Chetverikov and Kato (2014)) (we will use this result to bound R_1 later)

$$\begin{aligned} V &= \Sigma - \frac{1}{2} \sum_{i=1}^{n} \mathbb{E} \left[\xi_i \xi_i^\top \right] - \frac{1}{2} \sum_{i=1}^{n} \xi_i \xi_i^\top \\ &= (\Sigma - \mathrm{Var}(W)) - \frac{1}{2} \sum_{i=1}^{n} \left(\xi_i \xi_i^\top - \mathbb{E} \left[\xi_i \xi_i^\top \right] \right) \end{aligned} \tag{14}$$

Basic Decomposition

Next, by exchangeability we have

$$\begin{aligned}
&\mathbb{E}\left[D_j D_k D_l U \partial_{jkl} f(W + (1-U)D)\right] \\
&= -\mathbb{E}\left[D_j D_k D_l U \partial_{jkl} f\left(W' - (1-U)D\right)\right] \\
&= -\mathbb{E}\left[D_j D_k D_l U \partial_{jkl} f(W + UD)\right]
\end{aligned} \tag{15}$$

and also

$$R_2 = \frac{n}{4} \sum_{j,k,l=1}^{d} \mathbb{E}\left[D_j D_k D_l U \left\{\partial_{jkl} f(W + (1-U)D) - \partial_{jkl} f(W + UD)\right\}\right] \tag{16}$$

Basic Decomposition

If f is thrice differentiable with bounded partial derivatives, then

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = R_1 - R_2$$

where

$$R_1 = \sum_{j,k=1}^{d} \mathbb{E} V_{jk} \partial_{jk} f(W) = \mathbb{E}\langle V, \mathrm{Hess} f(W) \rangle_{H.S.}$$

$$R_2 = \frac{n}{4} \sum_{j,k,l=1}^{d} \mathbb{E}\left[D_j D_k D_l U \left\{ \partial_{jkl} f(W + (1-U)D) - \partial_{jkl} f(W + UD) \right\} \right]$$

and

$$\begin{aligned} V &= \Sigma - \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\left[\xi_i \xi_i^\top\right] - \frac{1}{2}\sum_{i=1}^{n} \xi_i \xi_i^\top \\ &= (\Sigma - \mathrm{Var}(W)) - \frac{1}{2}\sum_{i=1}^{n}\left(\xi_i\xi_i^\top - \mathbb{E}\left[\xi_i\xi_i^\top\right]\right) \end{aligned}$$

Main Theorem

Theorem (2.1)

Let $\xi = \{\xi_i\}_{i=1}^n$ be a sequence of centered independent random vectors in $\mathbb{R}^d$ with finite fourth moments and set $W = \sum_{i=1}^n \xi_i$. Assume $\mathrm{Var}(W) = \Sigma$ and Σ is invertible. Let $Z \sim N(0, \Sigma)$ be a centered Gaussian vector in $\mathbb{R}^d$ with covariance matrix Σ. Then,

$$\begin{aligned} &\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \\ \leqslant & C d^{1/4} \left(\sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right)^{1/2} \left(\left| \log \left(\sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right) \right| \vee 1 \right) \end{aligned} \tag{17}$$

where $\mathcal{A}$ is the collection of all measurable convex sets in $\mathbb{R}^d$.

Main Idea behind the Proof for Theorem 2.1

- Since $\Sigma^{-1/2}W = \sum_{i=1}^{n} \Sigma^{-1/2}\xi_i$ and $\left\{\Sigma^{-1/2}x : x \in A\right\} \in \mathcal{A}$ for all $A \in \mathcal{A}$, it suffices to consider the case $\Sigma = I_d$.
- Fix $\beta_0 > 0$. Define

$$K(\beta_0) = \sup_{W} \frac{\sup_{A\in\mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|}{\max\left\{\beta_0, \left(\sum_{i\in\mathcal{I}} \mathbb{E}|\xi_i|^4\right)^{1/2} \left(\left|\log\left(\sum_{i\in\mathcal{I}} \mathbb{E}|\xi_i|^4\right)\right| \vee 1\right)\right\}} \tag{18}$$

  where the first supremum is taken over the family of all sums $W = \sum_{i\in\mathcal{I}} \xi_i$ of finite number of independent mean-zero random vectors with $\mathbb{E}|\xi_i|^4 < \infty$ and $\mathrm{Var}(W) = I_d$.
- We will obtain a recursive inequality for $K(\beta_0)$ and prove that

$$K(\beta_0) \leqslant C d^{1/4} \tag{19}$$

  for an absolute constant C that does not depend on β_0. Equation (3) then follows by sending $\beta_0 \to 0$.

Proof of Theorem 2.1

Now we fix a $W = \sum_{i=1}^{n} \xi_i$, $n \geqslant 1$, in the aforementioned family

$$\bar{\beta} = \max\left\{\beta_0, \left(\sum_{i=1}^{n} \mathbb{E}\left|\xi_i\right|^4\right)^{1/2} \left(\left|\log\left(\sum_{i=1}^{n} \mathbb{E}\left|\xi_i\right|^4\right)\right| \vee 1\right)\right\}. \tag{20}$$

and for $A \in \mathcal{A}, \varepsilon > 0$, define

$$A^{\varepsilon} = \left\{x \in \mathbb{R}^d : \mathrm{dist}(x, A) \leqslant \varepsilon\right\} \qquad \mathrm{dist}(x, A) = \inf_{y \in A} |x - y|$$

To proceed, we need some technical lemmas.

Technical Lemmas

(Lemma 2.3 of Bentkus (2003))

Lemma (2)

For any $A \in \mathcal{A}$ and $\varepsilon > 0$, there exists a function $h_{A,\varepsilon}$ (which depends only on A and ε) such that

$$h_{A,\varepsilon}(x) = 1 \text{ for } x \in A, \quad h_{A,\varepsilon}(x) = 0 \text{ for } x \in \mathbb{R}^d \backslash A^\varepsilon, \quad 0 \leqslant h_{A,\varepsilon}(x) \leqslant 1$$

and

$$M_1\left(h_{A,\varepsilon}\right) \leqslant \frac{C}{\varepsilon}, \quad M_2\left(h_{A,\varepsilon}\right) \leqslant \frac{C}{\varepsilon^2} \tag{21}$$

where C is an absolute constant that does not depend on A and ε.

Technical Lemmas

(Theorem 4 of Ball (1993))

Lemma (3)

Let ϕ be the standard Gaussian density on $\mathbb{R}^d$, $d \geqslant 2$, and let A be a convex set in $\mathbb{R}^d$. Then

$$\int_{\partial A} \phi \leqslant 4d^{1/4} \tag{22}$$

Technical Lemmas

Using Lemma 3, one can show following lemmas of bounding the target difference between W and Z (Lemma 4.2 of Fang and Rollin (2015)).

Lemma (4)

For any d -dimensional random vector W and any $\varepsilon > 0$,

$$\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leqslant 4d^{1/4}\varepsilon + \sup_{A \in \mathcal{A}} |\mathbb{E}h_{A,\varepsilon}(W) - \mathbb{E}h_{A,\varepsilon}(Z)| \tag{23}$$

where $h_{A,\varepsilon}$ is as in Lemma 2.

Proof Continued

Before we proceed, we provide the outline of the remaining proof.

- Using equation (23) in lemma 4, we can bound $\sup_{A\in\mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|$ by bounding $\sup_{A\in\mathcal{A}} |\mathbb{E}h_{A,\varepsilon}(W) - \mathbb{E}h_{A,\varepsilon}(Z)|$.
- Then we can using basic decomposition to bound $\sup_{A\in\mathcal{A}} |\mathbb{E}h_{A,\varepsilon}(W) - \mathbb{E}h_{A,\varepsilon}(Z)|$ by considering R_1 and R_2 respectively.
- R_1 can be decomposed further into $R_{11} + R_{12}$, and each term can be bounded directly.
- To bound R_2, we divide into two cases. In the first case, $R_2 = R_{21} + R_{22}$ and we can bound two terms respectively. Besides, we will see the second case is trivial.

Proof Continued

We now fix $A \in \mathcal{A}$ (will take sup later), $0 < \varepsilon \leqslant 1$, write $h := h_{A,\varepsilon}$ and proceed to bound $|\mathbb{E}h(W) - \mathbb{E}h(Z)|$ by the basic decomposition (12). Consider the solution f to the Stein equation (6) with $\Sigma = I_d$

$$f(w) = \int_0^1 -\frac{1}{2(1-s)} \int_{\mathbb{R}^d} \left[h\left(\sqrt{1-s}w + \sqrt{s}z \right) - \mathbb{E}h(Z) \right] \phi(z) dz ds$$

Since h has bounded partial derivatives up to the second order and $\Sigma = I_d$ is invertible, f is thrice differentiable with bounded partial derivatives. Using the integration by parts formula, we have for $1 \leqslant j, k, l \leqslant d$ and any constant $0 \leqslant c_0 \leqslant 1$ that

Proof Continued

$$
\begin{aligned}
\partial_{jk} f(w) = &\int_0^{c_0} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_j h(\sqrt{1-s}w + \sqrt{s}z)\partial_k \phi(z)dzds \\
&+ \int_{c_0}^1 -\frac{1}{2s} \int_{\mathbb{R}^d} h(\sqrt{1-s}w + \sqrt{s}z)\partial_{jk}\phi(z)dzds
\end{aligned} \tag{24}
$$

and

$$
\begin{aligned}
\partial_{jkl} f(w) = &\int_0^{c_0} \frac{\sqrt{1-s}}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_{jk} h(\sqrt{1-s}w + \sqrt{s}z)\partial_l \phi(z)dzds \\
&+ \int_{c_0}^1 -\frac{\sqrt{1-s}}{2s} \int_{\mathbb{R}^d} \partial_j h(\sqrt{1-s}w + \sqrt{s}z)\partial_{kl}\phi(z)dzds
\end{aligned} \tag{25}
$$

Proof Continued

Now, using the expression of $\partial_{jk} f$ in (24) with $c_0 = \varepsilon^2$, we have

$$R_1 = R_{11} + R_{12}$$

where

$$R_{11} = \sum_{j,k=1}^{d} \mathbb{E}\left[V_{jk} \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_j h(\sqrt{1-s}W + \sqrt{s}z) \partial_k \phi(z) dz ds \right]$$

and

$$R_{12} = \sum_{j,k=1}^{d} \mathbb{E}\left[V_{jk} \int_{\varepsilon^2}^{1} -\frac{1}{2s} \int_{\mathbb{R}^d} h(\sqrt{1-s}W + \sqrt{s}z) \partial_{jk} \phi(z) dz ds \right]$$

Proof Continued

To proceed, we will utilize the following lemma (Lemma 4.3 of Fang and Röllin (2015)).

Lemma (5)

For $k \geqslant 1$ *and each map* $a : \{1, \ldots, d\}^k \to \mathbb{R}$, *we have*

$$\int_{\mathbb{R}^d} \left(\sum_{i_1,\ldots,i_k=1}^{d} a\left(i_1, \ldots, i_k\right) \frac{\partial_{i_1 \ldots i_k} \phi(z)}{\phi(z)} \right)^2 \phi(z) dz \leqslant k! \sum_{i_1,\ldots,i_k=1}^{d} \left(a\left(i_1, \ldots, i_k\right)\right)^2 \tag{26}$$

Bound for R_{11}

For R_{11}, we use the Cauchy-Schwarz inequality and the bounds in lemma 2 and lemma 5 and obtain

$$
\begin{aligned}
|R_{11}| &= \left| \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \mathbb{E} \sum_{j=1}^d \partial_j h(\sqrt{1-s}W + \sqrt{s}z) \sum_{k=1}^d V_{jk} \frac{\partial_k \phi(z)}{\phi(z)} \phi(z) dzds \right| \\
&\leqslant \frac{C}{\varepsilon} \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \mathbb{E} \left\{ \sum_{j=1}^d \left(\sum_{k=1}^d V_{jk} \frac{\partial_k \phi(z)}{\phi(z)} \right)^2 \right\}^{1/2} \phi(z) dzds \\
&\leqslant \frac{C}{\varepsilon} \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \left\{ \int_{\mathbb{R}^d} \mathbb{E} \sum_{j=1}^d \left(\sum_{k=1}^d V_{jk} \frac{\partial_k \phi(z)}{\phi(z)} \right)^2 \phi(z) dz \right\}^{1/2} ds \\
&\leqslant \frac{C}{\varepsilon} \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \left\{ \mathbb{E} \sum_{j=1}^d \sum_{k=1}^d V_{jk}^2 \right\}^{1/2} ds \leqslant C \left\{ \sum_{j,k=1}^d \mathbb{E} V_{jk}^2 \right\}^{1/2}
\end{aligned}
\tag{27}
$$

Bound for R_{11}

Recall that $\mathrm{Var}(W) = \Sigma$ and

$$V = \Sigma - \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\left[\xi_i \xi_i^\top\right] - \frac{1}{2}\sum_{i=1}^{n} \xi_i \xi_i^\top$$
$$= (\Sigma - \mathrm{Var}(W)) - \frac{1}{2}\sum_{i=1}^{n}\left(\xi_i \xi_i^\top - \mathbb{E}\left[\xi_i \xi_i^\top\right]\right)$$

we have

$$\mathbb{E} V_{jk}^2 = \frac{1}{4}\mathrm{Var}\left[\sum_{i=1}^{n} \xi_{ij}\xi_{ik}\right] = \frac{1}{4}\sum_{i=1}^{n}\mathrm{Var}\left[\xi_{ij}\xi_{ik}\right] \leqslant \frac{1}{4}\sum_{i=1}^{n}\mathbb{E}\left[\xi_{ij}^2\xi_{ik}^2\right]$$

Bound for R_{11}

and therefore,

$$\begin{aligned}
|R_{11}| &\leqslant C \left\{ \sum_{j,k=1}^{d} \sum_{i=1}^{n} \mathbb{E}\left[\xi_{ij}^2 \xi_{ik}^2\right] \right\}^{1/2} \\
&= C \left\{ \sum_{i=1}^{n} \mathbb{E}\left[\sum_{j=1}^{d} \xi_{ij}^2\right]^2 \right\}^{1/2} \\
&= C \left(\sum_{i=1}^{n} \mathbb{E}\left|\xi_i\right|^4 \right)^{1/2}
\end{aligned}$$

Bound for R_{12}

Applying similar arguments, we have, for R_{12},

$$|R_{12}| \leqslant \int_{\varepsilon^2}^{1} \frac{1}{2s} \left\{ \int_{\mathbb{R}^d} \mathbb{E} \left[\sum_{j,k=1}^{d} V_{jk} \frac{\partial_{jk}\phi(z)}{\phi(z)} \right]^2 \phi(z) dz \right\}^{1/2} ds$$

$$\leqslant C \int_{\varepsilon^2}^{1} \frac{1}{2s} \left\{ \mathbb{E} \sum_{j,k=1}^{d} V_{jk}^2 \right\}^{1/2} ds \leqslant C |\log \varepsilon| \left(\sum_{i=1}^{n} \mathbb{E} |\xi_i|^4 \right)^{1/2} \tag{28}$$

therefore,

$$|R_1| \leqslant C(|\log \varepsilon| \vee 1) \left(\sum_{i=1}^{n} \mathbb{E} |\xi_i|^4 \right)^{1/2} \tag{29}$$

Bound for R_2

Next, we bound R_2. Take $0 < \eta \leqslant 1$ arbitrarily. Using the expression of $\partial_{jkl} f$ in (25) with $c_0 = \eta^2$ and in this case, we have

$$R_2 = R_{21} + R_{22}$$

where

$$\begin{aligned} R_{21} =& \frac{1}{2} \sum_{i=1}^{n} \sum_{j,k,l=1}^{d} \mathbb{E} U \left(\xi'_{ij} - \xi_{ij}\right) \left(\xi'_{ik} - \xi_{ik}\right) \left(\xi'_{il} - \xi_{il}\right) \int_0^{\eta^2} \frac{\sqrt{1-s}}{2\sqrt{s}} \\ & \times \int_{\mathbb{R}^d} \partial_{jk} h \left(\sqrt{1-s} \left(W + (1-U)\left(\xi'_i - \xi_i\right)\right) + \sqrt{s} z\right) \partial_l \phi(z) dz ds \end{aligned}$$

Bound for R_2

and

$$R_{22} = \frac{1}{4}\sum_{i=1}^{n}\sum_{j,k,l,m=1}^{d} \mathbb{E}U(1-2U)\left(\xi'_{ij}-\xi_{ij}\right)\left(\xi'_{ik}-\xi_{ik}\right)\left(\xi'_{il}-\xi_{il}\right)\left(\xi'_{im}-\xi_{im}\right) \\ \times \int_{\eta^2}^{1} -\frac{1-s}{2s}\int_{\mathbb{R}^d} \partial_{jm}h_{22}\partial_{kl}\phi(z)dzds \tag{30}$$

where

$$h_{22} = h\left(\sqrt{1-s}\left(W + \left(U + (1-2U)U'\right)\left(\xi'_i - \xi_i\right)\right) + \sqrt{s}z\right)$$

and U' is a uniform random variable on $[0,1]$ independent of everything else.

Bound for R_2

Set $\beta_* = 0.19$ and $\sigma_* = (1 - \beta_*)^{1/2} = 0.9$. Recall that

$$\bar{\beta} = \max\left\{\beta_0, \left(\sum_{i=1}^{n} \mathbb{E}\left|\xi_i\right|^4\right)^{1/2}\left(\left|\log\left(\sum_{i=1}^{n} \mathbb{E}\left|\xi_i\right|^4\right)\right| \vee 1\right)\right\}$$

We now discuss the proof in following cases

- Case-1: $\bar{\beta} \leqslant \beta_*/d^{1/4}$
- Case-2: $\bar{\beta} > \beta_*/d^{1/4}$

The settings for β_* and σ_* will be used in bounding R_{21} and R_{22} with some specific calculation. We will skip these calculations and present the results directly.

Bound for R_2: Case-1: $\bar{\beta} \leqslant \beta_*/d^{1/4}$

In this case, we have for any $0 < \eta \leqslant 1$ and any $\varepsilon > 0$

$$|R_{21}| \leqslant \frac{C}{\varepsilon^2} \sum_{i=1}^{n} \mathbb{E}\,|\xi_i|^3 \left(d^{1/4}\varepsilon + K\left(\beta_0\right)\bar{\beta}\right)\eta \tag{31}$$

and

$$|R_{22}| \leqslant \frac{C}{\varepsilon^2} \sum_{i=1}^{n} \mathbb{E}\,|\xi_i|^4 \left(d^{1/4}\varepsilon + K\left(\beta_0\right)\bar{\beta}\right)|\log\eta| \tag{32}$$

Bound for R_2: Case-1: $\bar{\beta} \leqslant \beta_*/d^{1/4}$

By choosing appropriate η

$$\eta = \begin{cases} \frac{\sum_{i=1}^n \mathbb{E}|\xi_i|^4}{\sum_{i=1}^n \mathbb{E}|\xi_i|^3} & \text{if } \sum_{i=1}^n \mathbb{E}\left|\xi_i\right|^4 < \sum_{i=1}^n \mathbb{E}\left|\xi_i\right|^3 \\ 1 & \text{otherwise} \end{cases}$$

Hence, we have

$$\begin{aligned} |R_{21}| + |R_{22}| \leqslant & \frac{C}{\varepsilon^2} \sum_{i=1}^n \mathbb{E}\left|\xi_i\right|^4 \left(d^{1/4}\varepsilon + K\left(\beta_0\right)\bar{\beta}\right) \\ & \times \left(\left|\log\left(\sum_{i=1}^n \mathbb{E}\left|\xi_i\right|^4\right)\right| \vee 1\right) \end{aligned} \tag{33}$$

Bound for R_2: Case-1: $\bar{\beta} \leqslant \beta_* / d^{1/4}$

Therefore, in this case

$$
\begin{aligned}
&\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \\
\leqslant & 4 d^{1/4} \varepsilon + C(|\log \varepsilon| \vee 1) \left(\sum_{i=1}^{n} \mathbb{E} |\xi_i|^4 \right)^{1/2} \\
& + \frac{C}{\varepsilon^2} \sum_{i=1}^{n} \mathbb{E} |\xi_i|^4 \left(d^{1/4} \varepsilon + K(\beta_0) \bar{\beta} \right) \left(\left| \log \left(\sum_{i=1}^{n} \mathbb{E} |\xi_i|^4 \right) \right| \vee 1 \right)
\end{aligned} \tag{34}
$$

Choose

$\varepsilon = \min \left\{ \left[2C \sum_{i=1}^{n} \mathbb{E} |\xi_i|^4 \left(\left| \log \left(\sum_{i=1}^{n} \mathbb{E} |\xi_i|^4 \right) \right| \vee 1 \right) \right]^{1/2}, 1 \right\}$

with the same absolute constant C as in the third term on the right-hand side of (34)

Bound for R_2: Case-1: $\bar{\beta} \leqslant \beta_*/d^{1/4}$

If $\varepsilon < 1$, then (34) can be simplified to

$$\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leqslant \left(C d^{1/4} + \frac{K(\beta_0)}{2} \right) \bar{\beta}$$

hence

$$\frac{\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|}{\bar{\beta}} \leqslant C d^{1/4} + \frac{K(\beta_0)}{2} \tag{35}$$

If $\varepsilon = 1$, then $\sum_{i=1}^{n} \mathbb{E} |\xi_i|^4$ and $\bar{\beta}$ are bounded away from 0 by an absolute constant; hence

$$\frac{\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|}{\bar{\beta}} \leqslant \frac{2}{\bar{\beta}} \leqslant C \tag{36}$$

Bound for R_2: Case-2: $\bar{\beta} > \beta_*/d^{1/4}$

We trivially estimate

$$\frac{\sup_{A\in\mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|}{\bar{\beta}} \leqslant \frac{2}{\bar{\beta}} \leqslant \frac{2d^{1/4}}{\beta_*} \leqslant Cd^{1/4} \qquad (37)$$

Proof of Theorem 2.1

Combining both cases together, we have

$$\frac{\sup_{A\in\mathcal{A}}|\mathbb{P}(W\in A)-\mathbb{P}(Z\in A)|}{\bar{\beta}} \leqslant Cd^{1/4}+\frac{K(\beta_0)}{2}$$

Note that the right-hand side of the above bound does not depend on W. Taking supremum over W, we obtain

$$K(\beta_0) \leqslant Cd^{1/4}+\frac{K(\beta_0)}{2} \tag{38}$$

which completes the proof.

Proof of Theorem 2.2

The proof of Theorem 2.2 is quite similar to that of Theorem 2.1. It is enough to prove (4) when Σ is diagonal with positive entries. Fix $\beta_0 > 0$. Define

$$K'(\beta_0) = \sup_{W,\Sigma} \frac{\sup_{A\in\mathcal{B}} \left|\mathbb{P}(W\in A) - \mathbb{P}\left(\Sigma^{1/2}Z_0 \in A\right)\right|}{\max\{\beta_0, \Psi(\delta(W,\Sigma))\}} \qquad (39)$$

where $Z_0 \sim N(0, I_d)$ and the first supremum is taken over the family of all sums $W = \sum_{i\in\mathcal{I}} \xi_i$ of finite number of independent centered random vectors with $\mathbb{E}|\xi_i|^4 < \infty$, and diagonal matrices Σ with positive entries. We will obtain a recursive inequality for $K'(\beta_0)$ and prove that

$$K'(\beta_0) \leqslant C \qquad (40)$$

for an absolute constant C that does not depend on β_0. Equation (4) then follows by sending $\beta_0 \to 0$.

Applications

Applications on the bootstrap

Empirical bootstrap approximation for $\mathbb{P}(W \in A)$

- $X_1, \ldots, X_n$: be a sequence of centered independent vectors in $\mathbb{R}^d$ with finite fourth moments. $W := n^{-1/2} \sum_{i=1}^n X_i$, $\Sigma := Var(W)$, $Z \sim N(0, \Sigma)$. $X_1^*, \ldots, X_n^*$: be i.i.d. draws from the empirical distribution of X
- $W^* := \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i^* - \bar{X})$., where $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$
- The bootstrap analog of Theorem 2.3 is given by:

Theorem (4.1)

If $\operatorname{tr}(\Sigma^2) > 0$, *for any* $K > 0$, *we have*

$$\mathbb{P}\left(\sup_{A \in \mathcal{B}} |\mathbb{P}(W^* \in A|X) - \mathbb{P}(Z \in A)| > K\sqrt{\Delta_n} \right) \leqslant \frac{C}{K^2} \qquad (41)$$

where

$$\Delta_n := \frac{1}{n \operatorname{tr}(\Sigma^2)^{1/2}} \left(\sqrt{\sum_{i=1}^n \mathbb{E}\,|X_i|^4} + \sum_{j=1}^d \sqrt{\sum_{i=1}^n \mathbb{E}\left[X_{ij}^4\right]} \right)$$

Remark of Theorem 4.1

- Compared to the non-asymptotic bound for the quantity of $\sup_{A\in\mathcal{B}} |\mathbb{P}(W^* \in A|X) - \mathbb{P}(Z \in A)|$ under additional distribution assumption on X_i. Ours Theorem 4.1 provides better dependence on the **dimension** $d (d = o(n)$ **v.s.** $d = o(n^{1/2}))$**, at least when** $\Sigma = I_d$;
- Our result allows Σ **to be singular**;
- It's possible to give a non-asymptotic version of equation 41 but an exponential concentration if we also assume X_i are sub-Gaussian.

Wild bootstrap approximation for $\mathbb{P}(W \in A)$

Let $\{e_i\}_{i=1}^n$ be i.i.d. variables independent of $\{X_i\}_{i=1}^n$ with $\mathbb{E}e_1 = 0, \mathbb{E}e_1^2 = 1, \mathbb{E}e_1^4 < \infty$.

The $W^o := \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i X_i$. is the wild bootstrap approximation of W with multiplier variables $e_1, \ldots, e_n$.

Theorem (4.2)

If $\operatorname{tr}(\Sigma^2) > 0$, *for any* $K > 0$, *we have*

$$\mathbb{P}\left(\sup_{A \in \mathcal{B}} |\mathbb{P}(W^o \in A|X) - \mathbb{P}(Z \in A)| > K(\mathbb{E}e_1^4)^{1/4}\sqrt{\Delta_n}\right) \leqslant \frac{C}{K^2} \tag{42}$$

where Δ_n *is defined in 4.1*

Remark of Theorem 4.2

Compared to the non-asymptotic bound for the quantity of $\sup_{A\in\mathcal{B}} |\mathbb{P}(W^o \in A|X) - \mathbb{P}(Z \in A)|$ under additional distribution assumption on X_i. Our Theorem 4.2 provides better dependence on the n **and** $d(O(d/n)^{1/4}$ **v.s.** $O(d^2/n)^{1/5}))$;

Ours does not require the **unit skewness assumption** $\mathbb{E}e_1^3 = 1$ on the multiplier variables;

Thank you !