High-dimensional bootstrap and asymptotic expansion

Yuta Koike [*][†]

November 11, 2024

Abstract

The recent seminal work of Chernozhukov, Chetverikov and Kato has shown that bootstrap approximation for the maximum of a sum of independent random vectors is justified even when the dimension is much larger than the sample size. In this context, numerical experiments suggest that third-moment match bootstrap approximations would outperform normal approximation even without studentization, but the existing theoretical results cannot explain this phenomenon. In this paper, we first show that Edgeworth expansion, if justified, can give an explanation for this phenomenon. In particular, we derive an asymptotic expansion formula of the bootstrap coverage probability and show that the third-moment match wild bootstrap is second-order accurate in high-dimensions even without studentization when the covariance matrix has identical diagonal entries and bounded eigenvalues. In addition, we show the validity of the asymptotic expansion when appropriate random vectors have Stein kernels.

1 Introduction

Let $X_1, \dots, X_n$ be independent centered random vectors in $\mathbb{R}^d$ with finite variance. Set

$$S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i.$$

The aim of this paper is to investigate the accuracy of bootstrap approximation for the maximum type statistics

$$T_n := \max_{1 \le j \le d} S_{n,j} \quad \text{and} \quad \|S_n\|_\infty := \max_{1 \le j \le d} |S_{n,j}|,$$

when both n and d tend to infinity. The seminal work of Chernozhukov, Chetverikov & Kato [6] has established Gaussian type approximations for these statistics under very mild assumptions when the dimension d is possibly much larger than the sample size n. To be precise, let Z be a centered Gaussian vector in $\mathbb{R}^d$ with the same covariance matrix as S_n, say Σ. Gaussian analogs of T_n and $\|S_n\|_\infty$ are respectively given by

$$Z^\vee := \max_{1 \le j \le d} Z_j \quad \text{and} \quad \|Z\|_\infty := \max_{1 \le j \le d} |Z_j|.$$

Under mild moment assumptions, Chernozhukov, Chetverikov & Kato [6] have shown that

$$\sup_{t \in \mathbb{R}} |P(T_n \le t) - P(Z^\vee \le t)| = O\left(\left(\frac{\log^a(dn)}{n}\right)^b\right) \tag{1.1}$$

[*]Graduate School of Mathematical Sciences, University of Tokyo
[†]CREST, Japan Science and Technology Agency

holds with $a = 7$ and $b = 1/8$. An analogous result also holds for $\|S_n\|_\infty$. This result implies that given a significance level $\alpha \in (0,1)$, the probability $P(T_n \geq c^G_{1-\alpha})$ is approximately equal to α as long as $\log d = o(n^{1/7})$, where $c^G_{1-\alpha}$ is the $(1-\alpha)$-quantile of $Z^\vee$. Therefore, we can use $c^G_{1-\alpha}$ as a critical value to construct asymptotically $(1-\alpha)$-level simultaneous confidence intervals or α-level tests for a high-dimensional vector of parameters; see [1, 9] for details. In practice, $c^G_{1-\alpha}$ is not computable because Σ is generally unknown, so we need to replace it by an estimate. In [6], this is implemented by the Gaussian wild (or multiplier) bootstrap: Let $w_1, \dots, w_n$ be i.i.d. standard normal variables independent of the data $X_1, \dots, X_n$. Define the Gaussian wild bootstrap version of S_n as follows:

$$S^*_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i (X_i - \bar{X}), \quad \text{where } \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i. \tag{1.2}$$

We may naturally expect that $c^G_{1-\alpha}$ would be well-approximated by the $(1-\alpha)$-quantile of the conditional law of $T^*_n := \max_{1\leq j\leq d} S^*_{n,j}$ given the data, say $\hat{c}_{1-\alpha}$. This is formally justified by [6]: They essentially prove

$$P(T_n \geq \hat{c}_{1-\alpha}) = \alpha + O\left(\left(\frac{\log^a(dn)}{n}\right)^b\right) \tag{1.3}$$

with $a = 7$ and $b = 1/8$. The successive work [7] have improved the convergence rates of (1.1) and (1.3) to $b = 1/6$. They also proved the left hand side of (1.1) can be replaced by $\sup_{A\in\mathcal{R}} |P(S_n \in A) - P(Z \in A)|$, where $\mathcal{R} := \{\prod_{j=1}^d [a_j, b_j] : a_j \leq b_j,\ j = 1, \dots, d\}$ is the class of rectangles in $\mathbb{R}^d$.

It is easy to see that the conditional law of S^*_n given the data is $N(0, \widehat{\Sigma}_n)$, where $\widehat{\Sigma}_n$ is the sample covariance matrix: $\widehat{\Sigma}_n := n^{-1}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^\top$. Hence, the Gaussian wild bootstrap is essentially a feasible version of normal approximation for T_n. Then, it is natural to ask whether the approximation accuracy can be improved by more sophisticated bootstrap methods such as the empirical and non-Gaussian wild bootstraps. In the fixed-dimensional setting, it is well-known that standard bootstrap methods improve the approximation accuracy in the coverage probabilities upon normal approximation only when the statistic of interest is asymptotically pivotal (cf. [15, Chapter 3] and [20, Section 3]). However, despite that T_n and $\|S_n\|_\infty$ are not asymptotically pivotal in general, numerical experiments suggest that third-moment match bootstrap methods would outperform normal approximation (cf. [8, 11]). To appreciate this, we depict in Fig. 1 the P-P plot for the rejection rate $P(T_n \geq \hat{c}_{1-\alpha})$ against the nominal significance level α when $n = 200$ and $d = 400$, where $\hat{c}_{1-\alpha}$ is computed either the Gaussian wild bootstrap or a wild bootstrap with third-moment match. We can clearly see that the latter performance is much better than the former.

Deng & Zhang [11] tried to explain this phenomenon by showing that convergence rates of third-moment match bootstrap approximations have a better dimension dependence, i.e. they achieve $a = 5$ and $b = 1/6$ in (1.3). Later, however, it was shown in [16] that the same convergence rate is achieved by normal approximation, i.e. (1.1) holds with $a = 5$ and $b = 1/6$. Chernozhukov *et al.* [8] have further improved the convergence rate to $a = 5$ and $b = 1/4$ for both normal and bootstrap approximations. Meanwhile, if we require Σ to be invertible, it is possible to achieve the Berry–Esseen rate $n^{-1/2}$ up to a log factor even in the high-dimensional setting. Results in this direction first appeared in Fang & Koike [12], where the following

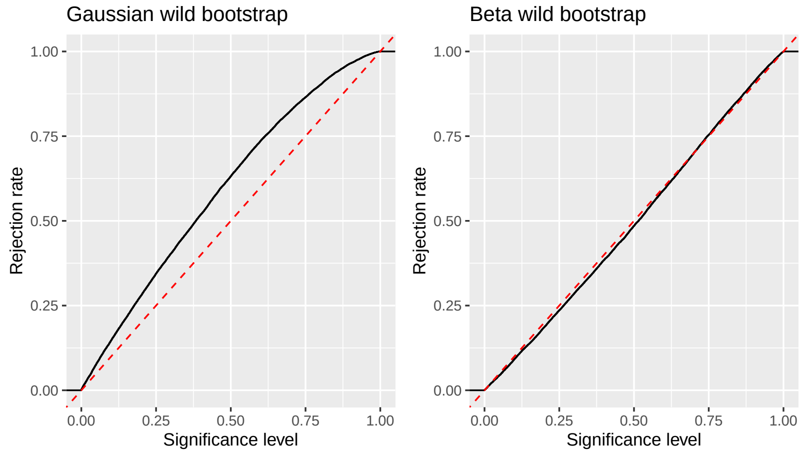

Figure 1: PP-plots for the rejection rate $P(T_n \geq \hat{c}_{1-\alpha})$ against the nominal significance level α when $n = 200$ and $d = 400$. The rejection rate is evaluated based on 20,000 Monte Carlo iterations. The critical value $\hat{c}_{1-\alpha}$ is computed by the Gaussian wild bootstrap for the left panel and the wild bootstrap with w_1 generated from the standardized beta distribution with parameters α, β given by (2.6) with $\nu = 0.1$ for the right panel, respectively. The number of bootstrap replications is 499. $X_1, \dots, X_n$ are generated from a Gaussian copula model with gamma marginals as in the simulation study. The parameter matrix is $R = (0.2^{|j-k|})_{1 \leq j,k \leq d}$.

result is obtained when $X_1, \dots, X_n$ are log-concave:

$$\sup_{A \in \mathcal{R}} |P(S_n \in A) - P(Z \in A)| = O\left(\sqrt{\frac{\log^3 d}{n}} \log n\right). \tag{1.4}$$

This rate is known to be optimal up to the $\log n$ factor in terms of both n and d; see Proposition 1.1 in [12]. This type of results has been further investigated in [10, 17, 21]. In particular, Chernozhukov *et al.* [10] have obtained the above nearly optimal rate when $\max_{i,j} |X_{ij}|$ is bounded. Further, in some situations, the rate $n^{-1/2}$ is (nearly) attainable even when Σ is (asymptotically) degenerate; see [13, 14, 22]. Nevertheless, all of these improvements are valid for normal approximation and thus do not explain the superior performances of third-moment match bootstrap approximations.

In this paper, we aim to explain the superior performance of bootstrap approximation in high-dimensions using Edgeworth expansion and related techniques. We first prove the validity of Edgeworth expansion for S_n in the high-dimensional setting when X_i have Stein kernels (cf. Definition 2.1). This also allows us to derive a valid Edgeworth expansion for the wild bootstrap statistic S_n^* when the weights w_i have Stein kernels. In particular, our results cover the simulation setting for Fig. 1 (cf. Example 2.1). Next, we develop an asymptotic expansion formula of $P(T_n \geq \hat{c}_{1-\alpha})$ in Theorem 3.3. As a conseuqnece, we find that the wild bootstrap with third moment match is second-order accurate *without studentization* when $d \geq n$ and Σ has identical diagonal entries and bounded eigenvalues, revealing the blessing of dimensionality in this context; see Corollary 3.1.

The full version of the paper is available at arXiv: `https://arxiv.org/abs/2404.05006`.

Notation Throughout the paper, we assume that S_n has an invertible covariance matrix Σ and denote by σ_* the square root of the minimum eigenvalue of Σ. We also set $\overline{\sigma} = \max_{j=1,\dots,d} \sqrt{\Sigma_{jj}}$ and $\underline{\sigma} = \min_{j=1,\dots,d} \sqrt{\Sigma_{jj}}$. Further, $w_1, \dots, w_n$ denote i.i.d. random variables independent of $X_1, \dots, X_n$. They are used to define the wild bootstrap statistic S_n^* in (1.2). We always assume $\mathrm{E}[w_1] = 0$ and $\mathrm{E}[w_1^2] = 1$. Also, P^* and E^* denote the conditional probability and expectation given the data $X_1, \dots, X_n$, respectively. For $p \in (0, 1)$, $\hat{c}_p$ denotes the conditional p-quantile of T_n^* given the data, i.e. $\hat{c}_p := \inf\{t \in \mathbb{R} : P^*(T_n^* \leq t) \geq p\}$.

For a vector $x \in \mathbb{R}^d$, we set $|x| := \sqrt{\sum_{j=1}^d x_j^2}$ and $x^\vee := \max_{1 \leq j \leq d} x_j$. We denote by $\mathbf{1}_d = (1, \dots, 1)^\top \in \mathbb{R}^d$ the all-ones vector in $\mathbb{R}^d$. For $r \in \mathbb{N}$, $(\mathbb{R}^d)^{\otimes r}$ denotes the set of real-valued d-dimensional r-arrays $V = (V_{j_1,\dots,j_r})_{1 \leq j_1,\dots,j_r \leq d}$. In particular, $(\mathbb{R}^d)^{\otimes 1} = \mathbb{R}^d$ and $(\mathbb{R}^d)^{\otimes 2}$ is the set of $d \times d$ matrices. For $U \in (\mathbb{R}^d)^{\otimes q}$ and $V \in (\mathbb{R}^d)^{\otimes r}$, we set $U \otimes V := (U_{i_1,\dots,i_q} V_{j_1,\dots,j_r})_{1 \leq i_1,\dots,i_q,j_1,\dots,j_r \leq d} \in (\mathbb{R}^d)^{\otimes(q+r)}$. We write $U^{\otimes 2} = U \otimes U$ for short. When $q = r$, we also set $\langle U, V \rangle := \sum_{j_1,\dots,j_r=1}^d U_{j_1,\dots,j_r} V_{j_1,\dots,j_r}$. In particular, when $q = r = 1$, $\langle U, V \rangle$ is the Euclidean inner product of U and V which we also write $U \cdot V$. In addition, we set $\|V\|_1 := \sum_{j_1,\dots,j_r=1}^d |V_{j_1,\dots,j_r}|$ and $\|V\|_\infty := \max_{1 \leq j_1,\dots,j_r \leq d} |V_{j_1,\dots,j_r}|$. Further, for $x \in \mathbb{R}^d$, we define $x^{\otimes r} := (x_{j_1} \cdots x_{j_r})_{1 \leq j_1,\dots,j_r \leq d} \in (\mathbb{R}^d)^{\otimes r}$. Finally, we set

$$\overline{X^r} := \frac{1}{n} \sum_{i=1}^n X_i^{\otimes r}.$$

Given an r-times differentiable function $h : \mathbb{R}^d \to \mathbb{R}$, we set $\nabla^r h(x) := (\partial_{j_1,\dots,j_r} h(x))_{1 \leq j_1,\dots,j_r \leq d} \in (\mathbb{R}^d)^{\otimes r}$ for $x \in \mathbb{R}^d$, where $\partial_{j_1,\dots,j_r} = \frac{\partial^r}{\partial x_{j_1} \cdots \partial x_{j_r}}$. For $m \in \mathbb{N} \cup \{\infty\}$, $C_b^m(\mathbb{R}^d)$ denotes the set of bounded C^m functions with bounded derivatives.

For an invertible matrix V, ϕ_V denotes the density of $N(0, V)$. We write $\phi_d = \phi_{I_d}$ for short, where I_d is the $d \times d$ identity matrix. Further, we write $\phi = \phi_1$ for short. Φ denotes the standard normal distribution function. Also, for a distribution function $F : \mathbb{R} \to [0, 1]$, its (generalized) inverse is defined as $F^{-1}(p) = \inf\{t \in \mathbb{R} : F(t) \geq p\}$, $p \in (0, 1)$. We refer to Appendix A.1 in [4] for useful properties of inverse distribution functions.

For a random vector ξ and $p \in (1, \infty)$, we set $\|\xi\|_p := (\mathrm{E}[|\xi|^p])^{1/p}$ (recall that $|\cdot|$ is the Euclidean norm). Further, for $\alpha > 0$, we set $\|\xi\|_{\psi_\alpha} := \inf\{t > 0 : \mathrm{E}[\exp\{(|\xi|/t)^\alpha\}] \leq 2\}$. For two random vectors ξ and η, we write $\xi \stackrel{d}{=} \eta$ if ξ has the same law as η.

We assume $d \geq 3$ whenever we consider an expression containing $\log d$. A similar convention is applied to n.

2 Valid Edgeworth expansion in high-dimensions

Let us formally define the notion of Stein kernel.

Definition 2.1 (Stein kernel)**.** Let ξ be a random vector in $\mathbb{R}^d$ with $\mathrm{E}[\|\xi\|_\infty] < \infty$. A measurable function $\tau : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ is called a *Stein kernel* for (the law of) ξ if $\mathrm{E}[\|\tau(\xi)\|_\infty] < \infty$ and

$$\mathrm{E}[(\xi - \mathrm{E}[\xi]) \cdot \nabla h(\xi)] = \mathrm{E}[\langle \tau(\xi), \nabla^2 h(\xi) \rangle] \tag{2.1}$$

for any $h \in C_b^2(\mathbb{R}^d)$.

The concept of Stein kernel was originally introduced in Stein [27, Lecture VI] for the univariate case. Although its partial multivariate extension dates back to [5], general treatments have started in more recent studies of [18, 26], stemming from the discovery of connection to Malliavin calculus due to Nourdin & Peccati [25] (the so-called *Malliavin–Stein method*). We refer to [23] for the recent development.

Remark 2.1 (Alternative definition)**.** Our definition of Stein kernel is taken from [18]. In the literature, the definition of Stein kernel often requires (2.1) to hold with ∇h on the both sides replaced by any bounded C^1 function $h : \mathbb{R}^d \to \mathbb{R}^d$ with bounded derivatives. Except for the case $d = 1$, this requirement is slightly stronger than ours. Nevertheless, as far as the author knows, this stronger requirement has so far been met by all known constructions of Stein kernels, including all the examples of this paper.

The validity of Edgeworth expansion for S_n is ensured if the summands have Stein kernels:

Theorem 2.1 (Edgeworth expansion for S_n)**.** *Suppose that X_i has a Stein kernel τ_i^X for every $i = 1, \dots, n$. Suppose also that there exists a constant $b > 0$ such that*

$$\|X_{ij}\|_{\psi_1} \leq b, \qquad \|\tau_{i,jk}^X(X_i)\|_{\psi_{1/2}} \leq b^2 \tag{2.2}$$

for all $i = 1, \dots, n$ and $j, k = 1, \dots, d$. Further, assume $\log^3 d \leq n$. Then there exists a universal constant $C > 0$ such that

$$\sup_{A \in \mathcal{R}} \left| P(S_n \in A) - \int_A p_n(z) dz \right| \leq C \frac{b^5}{\sigma_*^5} \frac{\log^3 d}{n} \log n. \tag{2.3}$$

Remark 2.2. Here and below, we do not intend to optimize the dependence of bounds on b and σ_*.

Example 2.1 (Gaussian copula model)**.** Let R be a $d \times d$ positive semidefinite symmetric matrix with unit diagonals. Also, for every $j = 1, \dots, d$, let μ_j be a non-degenerate probability distribution on $\mathbb{R}$ (i.e. μ_j is not the unit mass at a point), and denote by F_j its distribution function. The Gaussian copula model $U = (U_1, \dots, U_d)^\top$ with parameter matrix R and marginal distributions $\mu_1, \dots, \mu_d$ is defined as $U_j = F_j^{-1}(\Phi(Z_j))$ for $j = 1, \dots, d$, where $Z \sim N(0, R)$.

Proposition 2.1 (Stein kernel of Gaussian copula model)**.** *Suppose that there exists a constant $\kappa > 0$ such that for every $j = 1, \dots, d$ and any Borel set $B \subset \mathbb{R}$,*

$$\liminf_{h \downarrow 0} \frac{\mu_j(B^h) - \mu_j(B)}{h} \geq \kappa \min\{\mu_j(B), 1 - \mu_j(B)\}, \tag{2.4}$$

where $B^h := \{t \in \mathbb{R} : |t - s| < h$ for some $s \in B\}$. Then $X := U - \mathrm{E}[U]$ has a Stein kernel τ and

$$\max_{1 \leq j \leq d} \|X_j\|_{\psi_1} \leq C\kappa^{-1}, \qquad \max_{1 \leq j,k \leq d} \|\tau_{jk}(X)\|_{\psi_1} \leq C\kappa^{-2}$$

for some universal constant $C > 0$.

The maximal constant κ satisfying (2.4) is called the *Cheeger (isoperimetric) constant* of μ_j. We refer to [3, Theorem 1.3] for a useful equivalent formulation in the univariate case. When μ_j is log-concave, then (2.4) is satisfied with $\kappa = 1/\sqrt{3\operatorname{Var}[X_j]}$ by Proposition 4.1 in [2]. Since the gamma distribution with shape parameter ≥ 1 is log-concave, Proposition 2.1 shows that the simulated model in the introduction satisfies the assumptions of Theorem 2.1.

Other constructions of multivariate Stein kernels are found in [23, Section 4], although it does not seem straightforward to verify the second condition of (2.2) for them.

We turn to Edgeworth expansion for S_n^*. Its validity is ensured if the weight variables have Stein kernels:

Theorem 2.2 (Edgeworth expansion for S_n^*). *Suppose that* $\max_{1\le i\le n}\max_{1\le j\le d}\|X_{ij}\|_{\psi_1} \le b$ *for some constant* $b>0$. *Suppose also that* w_1 *satisfies either of the following conditions:*

(i) w_1 *has a Stein kernel* τ^* *and there exists a constant* $b_w\ge 1$ *such that* $|w_1|\le b_w$ *and* $|\tau^*(w_1)|\le b_w^2$.

(ii) $w_1\sim N(0,1)$. *We set* $b_w=1$ *in this case.*

Further, assume $\log^3 d\le n$. *Set* $\gamma:=\mathrm{E}[w_1^3]$. *Then there exists a universal constant* $C>0$ *such that*

$$\sup_{A\in\mathcal{R}}\left|P^*(S_n^*\in A)-\int_A\hat{p}_{n,\gamma}(z)dz\right|\le C\frac{b_w^5b^5}{\sigma_*^5}\frac{\log^3(dn)}{n}\log n \tag{2.5}$$

with probability at least $1-1/n$.

We can construct a random variable w_1 satisfying Condition (i) and $\mathrm{E}[w_1^3]=1$ as follows: Let η be a random variable following the beta distribution with parameters $\alpha,\beta>0$. Then $w:=(\eta-\mathrm{E}[\eta])/\sqrt{\mathrm{Var}[\eta]}$ satisfies (i) by [19, Example 4.9]. Also, we have

$$\mathrm{E}[w_1^3]=\frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}=\frac{2(1-2\mu)\sqrt{1+\nu}}{(2+\nu)\sqrt{\mu(1-\mu)}},$$

where $\mu=\alpha/(\alpha+\beta)$ and $\nu=\alpha+\beta$. From this expression, given a positive constant $\nu>0$, we have $\mathrm{E}[w_1^3]=1$ if we set

$$\alpha=\nu\frac{c-(2+\nu)\sqrt{c}}{2c},\quad\beta=\nu\frac{c+(2+\nu)\sqrt{c}}{2c}\quad\text{with } c=\nu^2+20\nu+20. \tag{2.6}$$

A drawback of Theorem 2.2 is that two-point distributions do not admit Stein kernels. In particular, it does not cover Mammen's wild bootstrap examined in the simulation study of [11]. However, the above standardized beta distribution becomes closer to Mammen's two-point distribution as ν is closer to 0, and their numerical difference virtually vanishes. Our simulation study shows that the beta wild bootstrap with $\nu=0.1$ performs very similarly to Mammen's one.

3 Second-order accurate approximation

Our next aim is to develop an asymptotic expansion of the bootstrap coverage probability. Such an expansion is conventionally derived with the help of Cornish–Fisher expansion (cf. Section 3.5.2 in [15]), so we first develop such expansions for T_n and T_n^* in our setting.

Before starting discussions, we introduce some notation used throughout this section. For $t\in\mathbb{R}$, we set $A(t):=(-\infty,t]^d$. We denote by f_Σ the density of $Z^\vee$, where $Z\sim N(0,\Sigma)$. Note that f_Σ is a C^∞ function since Σ is invertible. Finally, we set $\varsigma_d:=\sqrt{\mathrm{Var}[Z^\vee]\log d}$.

3.1 Cornish–Fisher expansion

This section develops Cornish–Fisher type expansions for T_n and T_n^*.

Theorem 3.1 (Cornish–Fisher expansion for T_n). *Under the assumptions of Theorem 2.1, let $\lambda > 0$ be a constant such that $b/\sigma_* \leq \lambda$. Then, for any $\varepsilon \in (0, 1/2)$, there exist positive constants c and C depending only on λ and ε such that if*

$$\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3 d}{n}\log n \leq c, \tag{3.1}$$

then

$$\sup_{\varepsilon<p<1-\varepsilon}\left|c_p - \left(c_p^G - \frac{Q_n(c_p^G)}{f_\Sigma(c_p^G)}\right)\right| \leq \frac{C}{\sqrt{\log d}}\frac{\varsigma_d^3}{\sigma_*^2}\frac{\log^3 d}{n}\log n, \tag{3.2}$$

where c_p is the p-quantile of T_n and

$$Q_n(t) := \int_{A(t)}\{p_n(z) - \phi_\Sigma(z)\}dz = -\frac{1}{6\sqrt{n}}\langle \mathrm{E}[\overline{X^3}], \int_{A(t)}\nabla^3\phi_\Sigma(z)dz\rangle, \qquad t \in \mathbb{R}.$$

Theorem 3.2 (Cornish–Fisher expansion for T_n^*). *Under the assumptions of Theorem 2.2, let $\lambda > 0$ be a constant such that $b/\sigma_* \leq \lambda$. Then, for any $\varepsilon \in (0, 1/2)$, there exist positive constants c and C depending only on λ, ε and b_w such that if*

$$\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}\log n \leq c, \tag{3.3}$$

then

$$\sup_{\varepsilon<p<1-\varepsilon}\left|\hat{c}_p - \left(c_p^G - \frac{\hat{Q}_{n,\gamma}(c_p^G)}{f_\Sigma(c_p^G)}\right)\right| \leq \frac{C}{\sqrt{\log d}}\frac{\varsigma_d^3}{\sigma_*^2}\frac{\log^3(dn)}{n}\log n \tag{3.4}$$

with probability at least $1 - 1/n$, where

$$\hat{Q}_{n,\gamma}(t) := \int_{A(t)}\{\hat{p}_{n,\gamma}(z) - \phi_\Sigma(z)\}dz.$$

3.2 Asymptotic expansion of coverage probability

For a $d \times d$ matrix V, $\mathrm{vec}(V)$ denotes the d^2-dimensional vector obtained by stacking the columns of V. For two random vectors ξ and η, the random vector $(\xi^\top, \eta^\top)^\top$ will be denoted by (ξ, η) for simplicity.

Theorem 3.3 (Asymptotic expansion of bootstrap coverage probability). *Suppose that the assumptions of Theorem 3.2 are satisfied. For every $i = 1, \dots, n$, set $Y_i := \mathrm{vec}(X_i^{\otimes 2} - \mathrm{E}[X_i^{\otimes 2}])$ and suppose that the $(d + d^2)$-dimensional random vector (X_i, Y_i) has a Stein kernel $\bar{\tau}_i$ of the form*

$$\bar{\tau}_i = \begin{pmatrix} \tau_i^X & \tau_i^{XY} \\ \tau_i^{YX} & \tau_i^Y \end{pmatrix} \tag{3.5}$$

with τ_i^X an $(\mathbb{R}^d)^{\otimes 2}$-valued function and such that

$$\begin{aligned} &\max_{1\leq j,k\leq d}\|\tau_{i,jk}^X(X_i, Y_i)\|_{\psi_{1/2}} \leq b^2, \qquad \max_{1\leq j,k\leq d^2}\|\tau_{i,jk}^Y(X_i, Y_i)\|_{\psi_{1/4}} \leq b^4, \\ &\max_{1\leq j\leq d, 1\leq k\leq d^2}\left(\|\tau_{i,jk}^{XY}(X_i, Y_i)\|_{\psi_{1/3}} + \|\tau_{i,kj}^{YX}(X_i, Y_i)\|_{\psi_{1/3}}\right) \leq b^3. \end{aligned} \tag{3.6}$$

Then, for any $\varepsilon \in (0, 1/2)$, there exist positive constants c and C depending only on λ, ε and b_w such that if (3.3) *holds, then*

$$\sup_{\varepsilon<\alpha<1-\varepsilon}\left|P(T_n \geq \hat{c}_{1-\alpha}) - \left(\alpha - (1-\gamma)Q_n(c_{1-\alpha}^G) - \mathrm{E}[R_n(\alpha)]\right)\right| \leq C\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}\log n,$$

where

$$R_n(\alpha) := \frac{1}{\sqrt{n}} \frac{\left\langle \overline{X^3} \otimes \mathbf{1}_d, \Psi_\alpha^{\otimes 2} \right\rangle}{2 f_\Sigma(c_{1-\alpha}^G)}, \qquad \Psi_\alpha := \int_{A(c_{1-\alpha}^G)} \nabla^2 \phi_\Sigma(z) dz.$$

Remark 3.1 (Univariate case)**.** When $d = 1$ and $\Sigma = 1$, the above asymptotic expansion formula reduces to

$$\begin{cases} \alpha - \frac{\mathrm{E}[\overline{X^3}]}{6\sqrt{n}} \{2(c_{1-\alpha}^G)^2 + 1\} \phi(c_{1-\alpha}^G) & \text{if } \gamma = 0, \\ \alpha - \frac{\mathrm{E}[\overline{X^3}]}{2\sqrt{n}} (c_{1-\alpha}^G)^2 \phi(c_{1-\alpha}^G) & \text{if } \gamma = 1. \end{cases}$$

These recover the asymptotic expansion formulae for normal and empirical bootstrap coverage probabilities, respectively; see e.g. [20, Eqs.(2)–(3)] (note that $c_{1-\alpha}^G = \Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$ when $d = 1$).

The new assumption in Theorem 3.3 is the existence of a (nice) Stein kernel for (X_i, Y_i). This assumption can be viewed as a counterpart of joint Cramér's condition for X_i and Y_i that is typically imposed to derive a univariate counterpart of Theorem 3.3; see e.g. Eq.(2.54) in [15]. It is natural in this sense, but the verification is not easy in practice. Here, we give one sufficient condition following Mikulincer [24]'s idea of using the Malliavin–Stein method.

Lemma 3.1. *Let G be a standard Gaussian vector in $\mathbb{R}^{d'}$. Let $\psi : \mathbb{R}^{d'} \to \mathbb{R}^d$ be a locally Lipschitz function such that $\mathrm{E}[|\psi(G)|^2] < \infty$ and $\max_{1 \le j \le d} \mathrm{E}[|\nabla \psi_j(G)|^2] < \infty$. Then $X := \psi(G) - \mathrm{E}[\psi(G)]$ has a Stein kernel τ such that*

$$\|\tau_{jk}(X)\|_p \le \|\nabla \psi_j(G)\|_{2p} \|\nabla \psi_k(G)\|_{2p} \tag{3.7}$$

for all $p \ge 1$ and $j, k = 1, \dots, d$. In addition,

$$\|X_j\|_p \le \sqrt{p-1} \|\nabla \psi_j(G)\|_p \tag{3.8}$$

for any even integer $p \ge 2$ and $j = 1, \dots, d$.

Moreover, if we further assume $\mathrm{E}[|\psi(G)|^4] < \infty$ and $\max_{1 \le j \le d} \mathrm{E}[|\psi(G)|^2 |\nabla \psi_j(G)|^2] < \infty$, then for $Y = \mathrm{vec}(X^{\otimes 2} - \mathrm{E}[X^{\otimes 2}])$, (X, Y) has a Stein kernel of the form (3.5) *and satisfies*

$$\max_{1 \le j,k \le d} \|\tau_{i,jk}^X(X,Y)\|_p \le \max_{1 \le j \le d} \|\nabla \psi_j(G)\|_{2p}^2,$$
$$\max_{1 \le j \le d, 1 \le l \le d^2} \left(\|\tau_{i,jl}^{XY}(X,Y)\|_p \vee \|\tau_{i,lj}^{YX}(X,Y)\|_p \right) \le 2 \max_{1 \le j,k,l \le d} \|\nabla \psi_j(G)\|_{2p} \|X_l \nabla \psi_k(G)\|_{2p},$$
$$\max_{1 \le l,m \le d^2} \|\tau_{i,lm}^Y(X,Y)\|_p \le 4 \max_{1 \le j,k \le d} \|X_j \nabla \psi_k(G)\|_{2p}^2$$

for all $p \ge 1$.

Example 3.1 (Gaussian copula model)**.** Consider the same setting as Example 2.1. Proposition 2.1 can be extended as follows.

Proposition 3.1. *Set $Y = \mathrm{vec}(X^{\otimes 2} - \mathrm{E}[X^{\otimes 2}])$. Under the assumptions of Proposition 2.1, (X, Y) has a Stein kernel of the form* (3.5) *and satisfies* (3.6) *with $b = C\kappa^{-1}$ for some universal constant $C > 0$.*

Now we discuss implications of Theorem 3.3 to the second-order accuracy of standard bootstrap approximations. An easy consequence is that any wild bootstrap approximation is second-order accurate when $\mathrm{E}[\overline{X^3}] = 0$ as long as w_1 satisfies the assumptions in Theorem 2.2. However, simulation results suggest that the choice of w_1 would affect the performance even when $\mathrm{E}[\overline{X^3}] = 0$, so there is still room to investigate.

The following corollary gives a more interesting implication:

Corollary 3.1. *Under the assumptions of Theorem 3.3, suppose additionally that* $\mathrm{E}[w_1^3] = 1$, $\varepsilon \geq 2e^{-d/2}$, $\underline{\sigma} = \overline{\sigma} =: \sigma$ *and the maximum eigenvalue of* Σ *is bounded by* $K\sigma^2$ *with some constant* $K > 0$. *Then there exist a constant* $C > 0$ *depending only on* λ, ε, K *and* b_w *such that*

$$\sup_{\varepsilon<\alpha<1-\varepsilon} |P(T_n \geq \hat{c}_{1-\alpha}) - \alpha| \leq C \left(\frac{\varsigma_d^3}{\sigma_*^3} \frac{\log^3(dn)}{n} \log n + \frac{\varsigma_d}{\sigma} \sqrt{\frac{\log^3 d}{dn}} \right). \tag{3.9}$$

Observe that the second term on the right hand side of (3.9) is divided by $\sqrt{d}$. Hence, Corollary 3.1 implies that the third-moment match wild bootstrap is second-order accurate if $d \geq n$ and Σ has identical diagonal entries and bounded eigenvalues with respect to d. This seems to be a new result on the blessing of dimensionality, although too high-dimensionality is harmful due to the first term of the bound.

Acknowledgments The author thanks Xiao Fang and Ryo Imai for valuable discussions about the subject of this paper. This work was partly supported by JST CREST Grant Number JPMJCR2115 and JSPS KAKENHI Grant Numbers JP22H00834, JP22H00889, JP22H01139, JP24K14848.

References

[1] Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C. & Kato, K. (2018). High-dimensional econometrics and regularized GMM. Working paper. arXiv: 1806.01888.

[2] Bobkov, S. G. (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.* **27**, 1903–1921.

[3] Bobkov, S. G. & Houdré, C. (1997). Isoperimetric constants for product probability measures. *Ann. Probab.* **25**, 184–205.

[4] Bobkov, S. G. & Ledoux, M. (2019). *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*, vol. 261 of *Memoirs of the American Mathematical Society*. American Mathematical Society, Providence.

[5] Cacoullos, T. & Papathanasiou, V. (1992). Lower variance bounds and a new proof of the central limit theorem. *J. Multivariate Anal.* **43**, 173–184.

[6] Chernozhukov, V., Chetverikov, D. & Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41**, 2786–2819.

[7] Chernozhukov, V., Chetverikov, D. & Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45**, 2309–2353.

[8] Chernozhukov, V., Chetverikov, D., Kato, K. & Koike, Y. (2022). Improved central limit theorem and bootstrap approximation in high dimensions. *Ann. Statist.* **50**, 2562–2586.

[9] Chernozhukov, V., Chetverikov, D., Kato, K. & Koike, Y. (2023). High-dimensional data bootstrap. *Annu. Rev. Stat. Appl.* **10**, 427–449.

[10] Chernozhukov, V., Chetverikov, D. & Koike, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* **33**, 2374–2425.

[11] Deng, H. & Zhang, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48**, 3643–3671.

[12] Fang, X. & Koike, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31**, 1660–1686.

[13] Fang, X. & Koike, Y. (2022). Sharp high-dimensional central limit theorems for log-concave distributions. *Ann. Inst. Henri Poincaré Probab. Stat. (forthcoming).* Working paper is available at arXiv: 2207.14536.

[14] Fang, X., Koike, Y., Liu, S.-H. & Zhao, Y.-K. (2023). High-dimensional central limit theorems by Stein's method in the degenerate case. Working paper. arXiv: 2305.17365v1.

[15] Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer, New York.

[16] Koike, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4**, 643–696.

[17] Kuchibhotla, A. K. & Rinaldo, A. (2020). High-dimensional CLT for sums of non-degenerate random vectors: $n^{-1/2}$-rate. Preprint, arXiv: 2009.13673.

[18] Ledoux, M., Nourdin, I. & Peccati, G. (2015). Stein's method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.* **25**, 256–306.

[19] Ley, C., Reinert, G. & Swan, Y. (2017). Stein's method for comparison of univariate distributions. *Probab. Surv.* **14**, 1–52.

[20] Liu, R. Y. & Singh, K. (1987). On a partial correction by the bootstrap. *Ann. Statist.* **15**, 1713–1718.

[21] Lopes, M. E. (2022). Central limit theorem and bootstrap approximation in high dimensions with near $1/\sqrt{n}$ rates. *Ann. Statist.* **50**, 2492–2513.

[22] Lopes, M. E., Lin, Z. & Müller, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Ann. Statist.* **48**, 1214–1229.

[23] Mijoule, G., Raič, M., Reinert, G. & Swan, Y. (2023). Stein's density method for multivariate continuous distributions. *Electron. J. Probab.* **28**, 1–40.

[24] Mikulincer, D. (2022). A CLT in Stein's distance for generalized Wishart matrices and higher-order tensors. *Int. Math. Res. Not. IMRN* **2022**, 7839–7872.

[25] Nourdin, I. & Peccati, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145**, 75–118.

[26] Nourdin, I., Peccati, G. & Swan, Y. (2014). Entropy and the fourth moment phenomenon. *J. Funct. Anal.* **266**, 3170–3207.

[27] Stein, C. (1986). *Approximate computation of expectations*. Institute of Mathematical Statistics, Hayward, California.