The Annals of Statistics
2022, Vol. 50, No. 5, 2545–2561
https://doi.org/10.1214/22-AOS2192
© Institute of Mathematical Statistics, 2022

NEW EDGEWORTH-TYPE EXPANSIONS WITH FINITE SAMPLE GUARANTEES

BY MAYYA ZHILOVA[a]

School of Mathematics, Georgia Institute of Technology, [a]*mzhilova@math.gatech.edu*

We establish higher-order nonasymptotic expansions for a difference between probability distributions of sums of i.i.d. random vectors in a Euclidean space. The derived bounds are uniform over two classes of sets: the set of all Euclidean balls and the set of all half-spaces. These results allow to account for an impact of higher-order moments or cumulants of the considered distributions; the obtained error terms depend on a sample size and a dimension explicitly. The new inequalities outperform accuracy of the normal approximation in existing Berry–Esseen inequalities under very general conditions. Under some symmetry assumptions on the probability distribution of random summands, the obtained results are optimal in terms of the ratio between the dimension and the sample size. The new technique which we developed for establishing nonasymptotic higher-order expansions can be interesting by itself. Using the new higher-order inequalities, we study accuracy of the nonparametric bootstrap approximation and propose a bootstrap score test under possible model misspecification. The results of the paper also include explicit error bounds for general elliptic confidence regions for an expected value of the random summands, and optimality of the Gaussian anticoncentration inequality over the set of all Euclidean balls.

1. Introduction. The Edgeworth series had been introduced by Edgeworth [16, 17] and Chebyshev [41], and developed by Cramér [15] (see Section 2.9 by Hall [24] for a detailed overview of early works about the Edgeworth series). Since that time, the Edgeworth expansion has become one of the major asymptotic techniques for approximation of a c.d.f. or a p.d.f. In particular, the Edgeworth expansion is a powerful instrument for establishing rates of convergence in the CLT and for studying accuracy of the bootstrap.

In this paragraph, we recall a basic form of the Edgeworth series and their properties that are useful for comparison with the proposed results; this statement can be found in Chapter 5 by Hall [24] (see also Bhattacharya and Rao [7], Kolassa [27], Skovgaard [40]). Let $S_n := n^{-1/2}\sum_{i=1}^n X_i$ for i.i.d. $\mathbb{R}^d$-valued random vectors $\{X_i\}_{i=1}^n$ with $\mathbb{E}X_i = 0$, $\Sigma := \operatorname{Var}(X_i)$, and $\mathbb{E}|X_i^{\otimes(k+2)}| < \infty$. Let $\mathscr{A}$ denote a class of sets $A \subseteq \mathbb{R}^d$ satisfying

$$\sup_{A\in\mathscr{A}} \int_{(\partial A)^\varepsilon} \varphi(x)\,dx = O(\varepsilon), \quad \varepsilon \downarrow 0, \tag{1.1}$$

where $\varphi(x)$ is the p.d.f. of $\mathcal{N}(0, I_d)$, and $(\partial A)^\varepsilon$ denotes the set of points distant no more than ε from the boundary ∂A of A. This condition holds for any measurable convex set in $\mathbb{R}^d$. Let also $\psi(t) := \mathbb{E}e^{it^T X_1}$. If the Cramér condition

$$\limsup_{\|t\|\to\infty} |\psi(t)| < 1 \tag{1.2}$$

Received December 2020; revised September 2021.
MSC2020 subject classifications. Primary 62E17, 62F40; secondary 62F25.
Key words and phrases. Edgeworth series, dependence on dimension, higher-order accuracy, multivariate Berry–Esseen inequality, chi-square approximation, finite sample inference, anticoncentration inequality, bootstrap, elliptic confidence sets, linear contrasts, bootstrap score test, model misspecification.

is fulfilled, then

$$\mathbb{P}(S_n \in A) = \int_A \left\{ \varphi_\Sigma(x) + \sum_{j=1}^{k} n^{-j/2} P_j\big(-\varphi_\Sigma : \{\kappa_j\}\big)(x) \right\} dx + o\big(n^{-k/2}\big) \tag{1.3}$$

for $n \to \infty$. The remainder term equals $o(n^{-k/2})$ uniformly in $A \in \mathscr{A}$, $\varphi_\Sigma(x)$ denotes the p.d.f. of $\mathcal{N}(0, \Sigma)$; κ_j are cumulants of X_1, and $P_j(-\varphi_\Sigma : \{\kappa_j\})(x)$ is a density of a signed measure, recovered from the series expansion of the characteristic function of X_1 using the inverse Fourier transform. In the multivariate case, a calculation of an expression for P_j for large j is rather involved since the number of terms included in it grows with j (see McCullagh [33]).

Expansion (1.3) does not hold for arbitrary random variables, in particular, Cramér's condition (1.2) holds if a probability distribution of X_1 has a nondegenerate absolutely continuous component. Condition (1.1) does not take into account dependence on dimension d. Indeed, if d is not reduced to a generic constant, then the right-hand side of (1.1) depends on d in different ways for major classes of sets. Let us refer to the works of Ball [2], Bentkus [5], Klivans, O'Donnell and Servedio [26], Chernozhukov, Chetverikov and Kato [13], Belloni, Bugni and Chernozhukov [4], where the authors established anticoncentration inequalities for important classes of sets.

Due to the asymptotic form of the Edgeworth series (1.3) for probability distributions, this kind of expansions is typically used in the asymptotic framework (for $n \to \infty$) without taking into account dependence of the remainder term $o(n^{-k/2})$ on the dimension. To the best of our knowledge, there have been no studies on accuracy of the Edgeworth expansions in finite sample multivariate setting so far. In this paper, we consider this framework and establish approximating bounds of type (1.3) with explicit dependence on dimension d and sample size n; this is useful for numerous contemporary applications, where it is important to track dependence of error terms on d and n. Furthermore, these results allow to account for an impact of higher-order moments of the considered distributions, which is important for deriving approximation bounds with higher-order accuracy. In order to derive the explicit multivariate higher-order expansions, we propose a novel proof technique that can be interesting and useful by itself.

One of the major applications of the proposed approximation bounds is the study of a performance of bootstrapping procedures in the nonasymptotic multivariate setting. In statistical inference, the bootstrap is one of the basic methods for estimation of probability distributions and quantiles of various statistics. Bootstrapping is well known for its *good finite sample performance* (see, e.g., Horowitz [25]), for this reason it is widely used in applications. However, a majority of theoretical results about the bootstrap are asymptotic (for $n \to \infty$), and most of the existing works about bootstrapping in the nonasymptotic high-dimensional/multivariate setting are quite recent. Arlot, Blanchard and Roquain [1] studied generalized weighted bootstrap for construction of nonasymptotic confidence bounds in ℓ_r-norm for $r \in [1, +\infty)$ for the mean value of high dimensional random vectors with a symmetric and bounded (or with Gaussian) distribution. Chernozhukov, Chetverikov and Kato [12] established results about accuracy of Gaussian approximation and bootstrapping for maxima of sums of high-dimensional vectors in a very general set-up. In [14] the authors extended and improved the results from maxima to general hyperractangles and sparsely convex sets. Bootstrap approximations can provide *faster rates of convergence* than the normal approximation (see Præstgaard and Wellner [36], Barbe and Bertail [3], Liu [30], Mammen [32], Lahiri [28], and references therein), however most of the existing results on this topic had been established in an asymptotic framework. In 2016, in Zhilova [46], we derived higher-order properties of the nonparametric and multiplier bootstrap. Those results were the first progress made on the higher-order accuracy of the bootstrap in the nonasymptotic (and multivariate) framework. In the present paper we derive new and much more general results. In particular, one of the implications of the proposed approximation bounds is an improvement of the Berry–Esseen inequality by Bentkus [5]. In Section 1.1 below, we summarize the contribution and the structure of the paper.

1.1. *Contribution and structure of the paper.* In Section 2, we establish expansions for the difference between probability distributions of $S_n := n^{-1/2}\sum_{i=1}^n X_i$ for i.i.d. random vectors $\{X_i\}_{i=1}^n$ and $\mathcal{N}(0,\Sigma)$, $\Sigma := \operatorname{Var}(S_n)$. The bounds are uniform over two classes of subsets of $\mathbb{R}^d$: the set $\mathscr{B}$ of all ℓ_2-balls, and the set $\mathscr{H}$ of all half-spaces. These classes of sets are useful when one works with linear or quadratic approximations of a smooth function of S_n; they are also useful for construction of confidence sets based on linear contrasts, for elliptic confidence regions, and for χ^2-type approximations in various parametric models where a multivariate statistic is asymptotically normal. In Sections 6 and 7 we consider examples of elliptic confidence regions, Rao's score test for a simple null hypothesis, and its bootstrap version that remains valid even in case of a misspecified parametric model.

In Theorem 2.1, where we study higher-order accuracy of the normal approximation of S_n over the class $\mathscr{B}$, the approximation error is $\le Cn^{-1/2}R_3 + C\sqrt{d^2/n} + Cd^2/n$. R_3 is a sublinear function of the third moment $\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}$, and $|R_3| \le \|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}\|_{\mathrm{F}}$ for the Frobenius norm $\|\cdot\|_{\mathrm{F}}$. The derived expressions for the error terms as well as the numerical constants are explicit. One of the implications of this result is an improvement of the Berry–Esseen inequality by Bentkus [5] that has the best known error rate for the class $\mathscr{B}$ (Remark 2.1 provides a detailed comparison between these results).

The proposed approximation bounds are not restricted to the normal approximation. In Theorems 2.2, 2.4, we consider the uniform bounds between the distributions of S_n and $S_{T,n} := n^{-1/2}\sum_{i=1}^n T_i$ for i.i.d. random vectors $\{T_i\}_{i=1}^n$ with the same expected value as X_i but possibly different covariance matrices. Here the error terms include a sublinear function of the differences $\mathbb{E}(X_1^{\otimes j}) - \mathbb{E}(T_1^{\otimes j})$ for $j = 2, 3$.

Let us also emphasize that the derived expansions impose considerably weaker conditions on probability distributions of X_i and T_i than the Edgeworth expansions (1.3) since our results do not require the Cramér condition (1.2) to be fulfilled, and they assume a smaller number of finite moments. Furthermore, the constants in our results do not depend on d and n, which allows to track dependence of the error terms on them. To the best of our knowledge, there have been no such results obtained so far.

In Section 3, we describe key steps of the proofs and the new technique which we developed for establishing the nonasymptotic higher-order expansions.

In Section 4, we consider the case of symmetrically distributed X_i. The error term in the normal approximation bound is $\le C(d^{3/2}/n)^{1/2}$, which is smaller than the error term $\le C(d^2/n)^{1/2}$ provided by Theorem 2.1 for the general case. Furthermore, we construct a lower bound, based on the example by Portnoy [35], showing that in this case the relation $d^{3/2}/n \to 0$ is required for consistency of the normal approximation.

In Section 5, we study accuracy of the nonparametric bootstrap approximation over set $\mathscr{B}$, using the higher-order methodology from Section 2. The resulting error terms depend on the quantities that characterize the sub-Gaussian tail behavior of X_i (proportional to their Orlicz ψ_2-norms) explicitly.

In Section 8, in the Supplementary Material [47], we collect statements from the earlier paper [46] which are used in the proofs of main results; we also provide improved bounds for constants in these statements and show optimality of the Gaussian anticoncentration bound over set $\mathscr{B}$. Proofs of the main results are presented in Sections 9 and 10 in [47].

1.2. *Notation.* For a vector $X = (x_1, \dots, x_d)^T \in \mathbb{R}^d$, $\|X\|$ denotes the Euclidean norm, $\mathbb{E}|X^{\otimes k}| < \infty$ denotes that $\mathbb{E}|x_{i_1} \cdots x_{i_k}| < \infty$ for all integer $i_1, \dots, i_k \in \{1, \dots, d\}$. For tensors $A, B \in \mathbb{R}^{d^{\otimes k}}$, their inner product equals $\langle A, B\rangle := \sum_{1 \le i_j \le d} a_{i_1,\dots,i_k} b_{i_1,\dots,i_k}$, where $a_{i_1,\dots,i_k}$ and $b_{i_1,\dots,i_k}$ are elements of A and B. The operator norm of A (for $k \ge 2$) induced by the Euclidean norm is denoted by $\|A\| := \sup\{\langle A, \gamma_1 \otimes \cdots \otimes \gamma_k\rangle : \|\gamma_j\| = 1, \gamma_j \in \mathbb{R}^d, j = 1, \dots, k\}$ (see Wang et al. [43]). The Frobenius norm is $\|A\|_{\mathrm{F}} = \sqrt{\langle A, A\rangle}$. The maximum norm is $\|A\|_{\max} := \max\{|a_{i_1,\dots,i_k}| : i_1, \dots, i_k \in \{1, \dots, d\}\}$. For a function $f : \mathbb{R}^d \mapsto \mathbb{R}$ and $h \in \mathbb{R}^d$, $f^{(s)}(x)h^s$ denotes the higher-order directional derivative $(h^T \nabla)^s f(x)$. $\varphi(x)$ denotes the p.d.f. of the standard normal distribution in $\mathbb{R}^d$. C, c denote positive generic constants. The abbreviations p.d. and p.s.d. denote positive definite and positive semi-definite matrices correspondingly.

2. Higher-order approximation bounds. Denote for random vectors X, Y in $\mathbb{R}^d$

$$\Delta_{\mathscr{B}}(X, Y) := \sup_{r \ge 0, t \in \mathbb{R}^d} \big|\mathbb{P}(\|X - t\| \le r) - \mathbb{P}(\|Y - t\| \le r)\big|. \tag{2.1}$$

Introduce the following functions:

$$h_1(\beta) := h_2(\beta) + (1 - \beta^2)^{-1}\beta^{-4}, \qquad h_2(\beta) := (1 - \beta^2)^2\beta^{-4} \tag{2.2}$$

for $\beta \in (0, 1)$. Let $\{X_i\}_{i=1}^n$ be i.i.d. $\mathbb{R}^d$-valued random vectors with $\mathbb{E}|X_i^{\otimes 4}| < \infty$ and p.d. covariance matrix $\Sigma := \operatorname{Var}(X_i)$. Without loss of generality, assume that $\mathbb{E}X_i = 0$. The following theorem provides the higher-order approximation bounds between $S_n := n^{-1/2}\sum_{i=1}^n X_i$ and the multivariate normal random vector $Z_\Sigma \sim \mathcal{N}(0, \Sigma)$ in terms of the distance $\Delta_{\mathscr{B}}(S_n, Z_\Sigma)$.

THEOREM 2.1. *Suppose that the conditions above are fulfilled, then it holds for any* $\beta \in (0, 1)$

$$\begin{aligned}\Delta_{\mathscr{B}}(S_n, Z_\Sigma) \le{}& (\sqrt{6}\beta^3)^{-1} R_3 n^{-1/2} \\ &+ 2C_{B,4}\|\Sigma^{-1}\|\|\Sigma\|\{(h_1(\beta) + (4\beta^4)^{-1})\mathbb{E}\|\Sigma^{-1/2}X_1\|^4 + d^2 + 2d\}^{1/2} n^{-1/2} \\ &+ (2\sqrt{6})^{-1}\{h_1(\beta)\mathbb{E}\|\Sigma^{-1/2}X_1\|^4 + h_2(\beta)(d^2 + 2d)\}n^{-1},\end{aligned}$$

where $C_{B,4} \ge 9.5$ *is a constant independent from* d, n, *and from a probability distribution of* X_i (*see the definition of* $C_{B,4}$ *in the proof after formula* (9.18) *in the Supplementary Material* [47]); R_3 *is a sublinear function of* $\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}$ *such that, in general,*

$$|R_3| \le \|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}\|_{\mathrm{F}} \le \|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}\|d.$$

Furthermore, if N *is the number of nonzero elements in* $\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}$ *and* $N \le d^2$, $m_3 := \|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}\|_{\max}$, *then*

$$|R_3| \le m_3\sqrt{N} \le m_3 d$$

(*a detailed definition of* R_3 *is given in* (9.12) *in the Supplementary Material* [47]).

COROLLARY 1. *Let* $\beta = 0.829$, *close to the local minimum of* $h_1(\beta)$, *then*

$$\begin{aligned}\Delta_{\mathscr{B}}(S_n, Z_\Sigma) \le{}& 0.717\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}\|dn^{-1/2} \\ &+ 2C_{B,4}\|\Sigma^{-1}\|\|\Sigma\|\{7.51\mathbb{E}\|\Sigma^{-1/2}X_1\|^4 + d^2 + 2d\}^{1/2} n^{-1/2} \\ &+ \{1.43\mathbb{E}\|\Sigma^{-1/2}X_1\|^4 + 0.043(d^2 + 2d)\}n^{-1}.\end{aligned}$$

Let also $m_4 := \|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 4}\|_{\max}$, *then*

$$\begin{aligned}\Delta_{\mathscr{B}}(S_n, Z_\Sigma) \le 0.717\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}\|dn^{-1/2} \\ + 2C_{B,4}\|\Sigma^{-1}\|\|\Sigma\|\{(7.51m_4+1)d^2+2d\}^{1/2}n^{-1/2} \\ + \{(1.425m_4+0.043)d^2+0.086d\}n^{-1}.\end{aligned} \tag{2.3}$$

Hence, if all the terms in (2.3) *except d and n are bounded by a generic constant* $C > 0$, *then*

$$\Delta_{\mathscr{B}}(S_n, Z_\Sigma) \le C\{\sqrt{d^2/n} + d^2/n\}. \tag{2.4}$$

REMARK 2.1. The Berry–Esseen inequality by Bentkus [5] shows that for $\Sigma = I_d$ and $\mathbb{E}\|X_1\|^3 < \infty$, $\Delta_{\mathscr{B}}(S_n, Z_\Sigma) \le c\mathbb{E}\|X_1\|^3 n^{-1/2}$. The term $(\sqrt{6}\beta^3)^{-1}n^{-1/2}R_3$ in Theorem 2.1 has an explicit constant and, since this is a sublinear function of the third moment of $\Sigma^{-1/2}X_1$, it can be considerably smaller than the third moment of the ℓ_2-norm $\|\Sigma^{-1/2}X_1\|$. Corollary 1 shows that the error term in Theorem 2.1 depends on d and n as $C(\sqrt{d^2/n} + d^2/n)$, which *improves the Berry–Esseen approximation error* $C\sqrt{d^3/n}$ in terms of the ratio between d and n. Theorem 2.1 imposes a stronger moment assumption than the Berry–Esseen bound by Bentkus [5], since the latter inequality assumes only 3 finite moments of $\|X_i\|$. However, the theorems considered here require much weaker conditions than the Edgeworth expansions (1.3) that would assume in general at least 5 finite moments of $\|X_i\|$ and the Cramér condition (1.2).

REMARK 2.2. Since functions $h_1(\beta)$, $h_2(\beta)$ are known explicitly (2.2), the expression of the approximation error term in Theorem 2.1 contains explicit constants and it even allows to optimize the error term (w.r.t. parameter $\beta \in (0, 1)$), depending on R_3, $\mathbb{E}\|\Sigma^{-1/2}X_1\|^4$, and other terms as well. Therefore, the results in this paper allow to address the problem of finding an optimal constant in Berry–Esseen inequalities (see, e.g., Shevtsova [39]).

The following statement is an extension of Theorem 2.1 to a general (not necessarily normal) approximating distribution. Let $\{T_i\}_{i=1}^n$ be i.i.d random vectors in $\mathbb{R}^d$, with $\mathbb{E}T_i = 0$, p.d. covariance matrix $\operatorname{Var}(T_i) = \Sigma_T$, and $\mathbb{E}|T_i^{\otimes 4}| < \infty$. Let also $S_{T,n} := n^{-1/2}\sum_{i=1}^n T_i$.

THEOREM 2.2. *Let* $\{X_i\}_{i=1}^n$ *satisfy conditions of Theorem* 2.1. *First, consider the case* $\operatorname{Var}(T_i) = \operatorname{Var}(X_i) = \Sigma$. *Denote*

$$\bar{V}_4 := \mathbb{E}\|\Sigma^{-1/2}X_1\|^4 + \mathbb{E}\|\Sigma^{-1/2}T_1\|^4.$$

It holds for any $\beta \in (0, 1)$

$$\begin{aligned}\Delta_{\mathscr{B}}(S_n, S_{T,n}) \le (\sqrt{6}\beta^3)^{-1}\bar{R}_{3,T}n^{-1/2} + (2\sqrt{6})^{-1}h_1(\beta)\bar{V}_4 n^{-1}, \\ + \sqrt{8}C_{B,4}\|\Sigma^{-1}\|\|\Sigma\|\{(h_1(\beta) + (4\beta^4)^{-1})\bar{V}_4 + 2d^2 + 4d\}^{1/2}n^{-1/2},\end{aligned}$$

where $\bar{R}_{3,T}$ *is a sublinear function of* $\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3} - \mathbb{E}(\Sigma^{-1/2}T_1)^{\otimes 3}$ *such that, in general,*

$$\begin{aligned}|\bar{R}_{3,T}| &\le \|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3} - \mathbb{E}(\Sigma^{-1/2}T_1)^{\otimes 3}\|_{\mathrm{F}} \\ &\le \|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3} - \mathbb{E}(\Sigma^{-1/2}T_1)^{\otimes 3}\|d.\end{aligned}$$

Furthermore, if N_T is the number of nonzero elements in $\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3} - \mathbb{E}(\Sigma^{-1/2}T_1)^{\otimes 3}$, *and* $N_T \le d^2$, $m_{3,T} = \|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3} - \mathbb{E}(\Sigma^{-1/2}T_1)^{\otimes 3}\|_{\max}$, *then*

$$|\bar{R}_{3,T}| \le m_{3,T}\sqrt{N_T} \le m_{3,T}d.$$

Now consider the case when $\operatorname{Var} X_i = \Sigma$ *and* $\operatorname{Var} T_i = \Sigma_T$ *are not necessarily equal to each other. Let* $\lambda_0^2 > 0$ *denote the minimum of the smallest eigenvalues of* Σ *and* Σ_T. *Denote*

$$V_4 := \mathbb{E}\|X_1\|^4 + \mathbb{E}\|T_1\|^4, \qquad v_4 := \|\Sigma\|^2 + \|\Sigma_T\|^2.$$

It holds for any $\beta \in (0,1)$

$$\begin{aligned}\Delta_{\mathscr{B}}(S_n, S_{T,n}) \le{}& (\sqrt{2}\beta^2\lambda_0^2)^{-1}\|\Sigma - \Sigma_T\|_{\mathrm{F}} + (\sqrt{6}\beta^3)^{-1} R_{3,T} n^{-1/2} \\ &+ 4\sqrt{2} C_{B,4}\lambda_0^{-2}\{h_1(\beta)V_4 + (d^2+2d)(v_4+1/2)\}^{1/2} n^{-1/2} \\ &+ 2(\sqrt{6}\lambda_0^4)^{-1}\{h_1(\beta)V_4 + (d^2+2d)v_4\} n^{-1},\end{aligned}$$

where $R_{3,T}$ *is a sublinear function of* $\mathbb{E}(X_1^{\otimes 3}) - \mathbb{E}(T_1^{\otimes 3})$ *such that, in general,*

$$|R_{3,T}| \le \lambda_0^{-3}\|\mathbb{E}(X_1^{\otimes 3}) - \mathbb{E}(T_1^{\otimes 3})\|_{\mathrm{F}} \le \lambda_0^{-3}\|\mathbb{E}(X_1^{\otimes 3}) - \mathbb{E}(T_1^{\otimes 3})\| d$$

(*a detailed definition of* $R_{3,T}$ *is given in* (9.24) *in the Supplementary Material* [47]).

Below we consider the uniform distance between the probability distributions of S_n and Z_Σ over the set of all half-spaces in $\mathbb{R}^d$:

$$\Delta_{\mathscr{H}}(S_n, Z_\Sigma) := \sup_{x\in\mathbb{R},\gamma\in\mathbb{R}^d} |\mathbb{P}(\gamma^T S_n \le x) - \mathbb{P}(\gamma^T Z_\Sigma \le x)|. \tag{2.5}$$

Denote $h_3(\beta) := 3\beta^{-4}\{1-(1-\beta^2)^2\}$ for $\beta \in (0,1)$ (similarly to h_1, h_2 introduced in (2.2)).

THEOREM 2.3. *Given the conditions of Theorem* 2.1, *it holds* $\forall \beta \in (0,1)$

$$\begin{aligned}\Delta_{\mathscr{H}}(S_n, Z_\Sigma) \le{}& (\sqrt{6}\beta^3)^{-1}\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}\| n^{-1/2} \\ &+ C_{H,4}\{(h_1(\beta)+\beta^{-4})\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 4}\| + h_3(\beta)\}^{1/2} n^{-1/2} \\ &+ (2\sqrt{6})^{-1}\{h_1(\beta)\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 4}\| + 3h_2(\beta)\} n^{-1},\end{aligned}$$

where $C_{H,4} \ge 9.5$ *is a constant independent from* d, n, *and a probability distribution of* X_i (*see the definition of* $C_{H,4}$ *in the proof after formula* (9.28) *in* [47]).

COROLLARY 2. *Let* $\beta = 0.829$, *close to the local minimum of* $h_1(\beta)$, *then*

$$\begin{aligned}\Delta_{\mathscr{H}}(S_n, Z_\Sigma) \le{}& 0.717\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}\| n^{-1/2} \\ &+ C_{H,4}\{9.10\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 4}\| + 5.731\}^{1/2} n^{-1/2} \\ &+ \{1.425\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 4}\| + 0.127\} n^{-1}.\end{aligned}$$

REMARK 2.3. The inequalities that we establish for the class $\mathscr{H}$ are *dimension-free*. Indeed, the approximation errors in Theorems 2.3, 2.4 and Corollary 2 depend only on numerical constants, on $n^{-1/2}$, n^{-1}, and on the operator norms of the third and the fourth moments of $\Sigma^{-1/2}X_1$:

$$\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes j}\| = \sup_{\gamma\in\mathbb{R}^d, \|\gamma\|=1} \mathbb{E}(\gamma^T\Sigma^{-1/2}X_1)^j \quad \text{for } j = 3, 4.$$

REMARK 2.4. Recalling the arguments in Remark 2.1, $\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3}\|$ in the latter statement depends on the third moment of X_1 sublinearly. Furthermore, the classical Berry–Esseen theorem by Berry [6] and Esseen [20] (that requires $\mathbb{E}|X_i^{\otimes 3}| < \infty$) gives an error term $\le c\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 4}\|^{3/4}n^{-1/2}$ which is $\ge \sqrt{\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 4}\|/n}$ because $\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 4}\| \ge 1$. This justifies that Theorem 2.3 can have a better accuracy than the result for $\Delta_{\mathscr{H}}$ implied by the classical Berry–Esseen inequality when, for example, $\mathbb{E}X_1^{\otimes 3} = 0$ and $\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 4}\|$ is rather big (e.g., for the logistic or von Mises distributions).

The following statement extends Theorem 2.3 to the case of a general (not necessarily normal) approximating distribution, similarly to Theorem 2.2. Recall that $v_4 = \|\Sigma\|^2 + \|\Sigma_T\|^2$ and denote

$$\bar{V}_{T,4} := \|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 4}\| + \|\mathbb{E}(\Sigma^{-1/2}T_1)^{\otimes 4}\|, \qquad V_{T,4} := \|\mathbb{E}(X_1^{\otimes 4})\| + \|\mathbb{E}(T_1^{\otimes 4})\|.$$

THEOREM 2.4. *Given the conditions of Theorem 2.2, it holds* $\forall \beta \in (0,1)$

$$\begin{aligned}\Delta_{\mathscr{H}}(S_n, S_{T,n}) \le{}& (\sqrt{6}\beta^3)^{-1}\|\mathbb{E}(\Sigma^{-1/2}X_1)^{\otimes 3} - \mathbb{E}(\Sigma^{-1/2}T_1)^{\otimes 3}\|n^{-1/2}\\ &+ C_{H,4}\{(h_1(\beta) + \beta^{-4})\bar{V}_{T,4} + 2h_3(\beta)\}^{1/2}n^{-1/2}\\ &+ (2\sqrt{6})^{-1}h_1(\beta)\bar{V}_{T,4}n^{-1}.\end{aligned}$$

Consider the case when $\operatorname{Var} X_i = \Sigma$ *and* $\operatorname{Var} T_i = \Sigma_T$ *are not necessarily equal to each other. Let* $\lambda_0^2 > 0$ *denote the minimum of the smallest eigenvalues of* Σ *and* Σ_T. *It holds* $\forall \beta \in (0,1)$

$$\begin{aligned}\Delta_{\mathscr{H}}(S_n, S_{T,n}) \le{}& (\sqrt{2}\beta^2\lambda_0^2)^{-1}\|\Sigma - \Sigma_T\|\\ &+ (\sqrt{6}\beta^3\lambda_0^3)^{-1}\|\mathbb{E}(X_1^{\otimes 3}) - \mathbb{E}(T_1^{\otimes 3})\|n^{-1/2}\\ &+ 4\sqrt{2}C_{H,4}\lambda_0^{-2}\{h_1(\beta)V_{T,4} + 3(v_4 + 1/2)\}^{1/2}n^{-1/2}\\ &+ 2(\sqrt{6}\lambda_0^4)^{-1}\{h_1(\beta)V_{T,4} + 3v_4\}n^{-1}.\end{aligned}$$

3. New proof technique. In this section, we describe the key steps and ideas that we develop in the proofs of Theorems 2.1 and 2.2 for the class $\mathscr{B}$. Theorems 2.3 and 2.4 about half-spaces are derived in an analogous way.

First, we use the triangle inequality

$$\Delta_{\mathscr{B}}(S_n, Z_\Sigma) \le \Delta_{\mathscr{B}}(S_n, \tilde{S}_n) + \Delta_{\mathscr{B}}(\tilde{S}_n, Z_\Sigma), \tag{3.1}$$

where $\tilde{S}_n = n^{-1/2}\sum_{i=1}^n Y_i$ for i.i.d. random summands Y_i constructed in a special way (see definitions (9.3), (9.4) in [47]). We define Y_i such that they have matching moments of orders 1, 2, 3 with the original random vectors X_i, and at the same time they have a normal component which plays crucial role in the smoothing technique that we describe below.

To the term $\Delta_{\mathscr{B}}(S_n, \tilde{S}_n)$ in (3.1) we apply the Berry–Esseen type inequality from [46] (this result is discussed in more detail in Section 8 in the Supplementary Material [47]), which yields

$$\begin{aligned}&\Delta_{\mathscr{B}}(S_n, \tilde{S}_n)\\ &\quad\le 2C_{B,4}\|\Sigma^{-1}\|\|\Sigma\|\sqrt{\{(h_1(\beta) + 0.25/\beta^4)\mathbb{E}\|\Sigma^{-1/2}X_1\|^4 + d^2 + 2d\}/n}.\end{aligned} \tag{3.2}$$

Here the error rate $C\sqrt{d^2/n}$ comes from the higher-order moment matching property between the random summands of S_n and $\tilde{S}_n$, which improves the ratio $C\sqrt{d^3/n}$ between dimension d and sample size n in the classical Berry–Esseen result by Bentkus [5] (in the classical Berry–Esseen theorem one uses, in general, only first two matching moments which is smaller than first three moments). Also the square root in this expression naturally comes from the smoothing argument used for derivation of the Berry–Esseen inequality with the best-known rate w.r.t. d and n, and it is unavoidable for the distance $\Delta_{\mathscr{B}}(S_n, \tilde{S}_n)$ under the mild conditions on X_i imposed here. The proof of the result in [46] is based on an extension of the proof of the Berry–Esseen inequaliy by Bentkus [5].

For the term $\Delta_{\mathscr{B}}(\tilde{S}_n, Z_\Sigma)$ in (3.1), we exploit the structure of $\tilde{S}_n$ in order to construct the higher-order expansion that allows to compare moments of X_i and Z_Σ. It holds

$$\tilde{S}_n \overset{d}{=} \tilde{Z} + n^{-1/2}\sum_{i=1}^{n} U_i, \tag{3.3}$$

where $\tilde{Z} \sim \mathcal{N}(0, \beta^2\Sigma)$ for $\beta \in (0, 1)$ that enters the resulting bounds as a free parameter and can be used for optimizing the approximation error terms w.r.t. it. Random vectors $\{U_i\}_{i=1}^n$ are i.i.d. and independent from $\tilde{Z}$, and $\{X_i\}_{i=1}^n$, hence the expression in (3.3) has multivariate normal distribution, conditionally on $\{U_i\}$. Also

$$\mathbb{E}U_i = \mathbb{E}X_i = 0, \qquad \operatorname{Var} U_i = (1-\beta^2)\Sigma, \qquad \mathbb{E}U_i^{\otimes 3} = \mathbb{E}X_i^{\otimes 3}. \tag{3.4}$$

We introduce the following representation of the probability distribution functions of $\tilde{S}_n$ and Z_Σ. Let $B \in \mathscr{B}$ and $B' := \{x \in \mathbb{R}^d : \beta\Sigma^{1/2}x \in B\}$, and $Z_0 := \beta^{-1}\Sigma^{-1/2}\tilde{Z} \sim \mathcal{N}(0, I_d)$, then it holds

$$\begin{aligned}
\mathbb{P}(\tilde{S}_n \in B) &= \mathbb{E}\biggl\{\mathbb{P}\biggl(\tilde{Z} + n^{-1/2}\sum_{i=1}^{n} U_i \in B | \{U_i\}_{i=1}^n\biggr)\biggr\} \\
&= \mathbb{E}\biggl\{\mathbb{P}\biggl(Z_0 + n^{-1/2}\beta^{-1}\sum_{i=1}^{n} \Sigma^{-1/2}U_i \in B' | \{U_i\}_{i=1}^n\biggr)\biggr\} \\
&= \mathbb{E}\int_{B'} \varphi\biggl(t - n^{-1/2}\beta^{-1}\sum_{i=1}^{n} \Sigma^{-1/2}U_i\biggr)\,dt,
\end{aligned} \tag{3.5}$$

for $\varphi(t)$ denoting the p.d.f. of Z_0. In this way, *we use the normal component* of $\tilde{S}_n$ to represent $\mathbb{P}(\tilde{S}_n \in B)$ *as an expectation of a smooth function* of the sum if i.i.d. random vectors $n^{-1/2}\beta^{-1}\sum_{i=1}^n \Sigma^{-1/2}U_i$ that have *matching moments* with the original samples X_i. The same representation holds for the approximating distribution Z_Σ. Let $Z_i \sim \mathcal{N}(0, (1-\beta^2)\Sigma)$ be i.i.d., independent from all other random vectors with the same first two moments as U_i, then $Z_\Sigma \overset{d}{=} \tilde{Z} + n^{-1/2}\sum_{i=1}^n Z_i$,

$$\begin{aligned}
\mathbb{P}(Z_\Sigma \in B) &= \mathbb{E}\biggl\{\mathbb{P}\biggl(\tilde{Z} + n^{-1/2}\sum_{i=1}^{n} Z_i \in B | \{Z_i\}_{i=1}^n\biggr)\biggr\} \\
&= \mathbb{E}\biggl\{\mathbb{P}\biggl(Z_0 + n^{-1/2}\beta^{-1}\sum_{i=1}^{n} \Sigma^{-1/2}Z_i \in B' | \{Z_i\}_{i=1}^n\biggr)\biggr\} \\
&= \mathbb{E}\int_{B'} \varphi\biggl(t - n^{-1/2}\beta^{-1}\sum_{i=1}^{n} \Sigma^{-1/2}Z_i\biggr)\,dt.
\end{aligned} \tag{3.6}$$

Now we represent the difference $\mathbb{P}(\tilde{S}_n \in B) - \mathbb{P}(Z_\Sigma \in B)$ as the following telescoping sum (the general telescopic sum principle or the swapping method is due to Lindeberg [29]; see also Trotter [42] and Chatterjee [11]):

$$\begin{aligned}
&\mathbb{P}(\tilde{S}_n \in B) - \mathbb{P}(Z_\Sigma \in B) \\
&\quad = \sum_{i=1}^{n} \mathbb{E}\int_{B'} \{\varphi(t - (n^{1/2}\beta\Sigma^{1/2})^{-1}U_i - s_i)
\end{aligned} \tag{3.7}$$

$$- \varphi\big(t - (n^{1/2}\beta\Sigma^{1/2})^{-1} Z_i - s_i\big)\big\} \, dt,$$

where $s_i = n^{-1/2}\beta^{-1}\sum_{k=1}^{i-1}\Sigma^{-1/2}Z_k + n^{-1/2}\beta^{-1}\sum_{k=i+1}^{n}\Sigma^{-1/2}U_k$ for $i = 1, \dots, n$, the sums are taken equal zero if index k runs beyond the specified range. Random vectors s_i are independent from U_i and Z_i which is used in the proof together with the Taylor expansion of $\varphi(t)$ and the matching moments property (3.4). Further details of the calculations are available in Section 9 in the Supplementary Material [47]. The resulting error bound

$$\Delta_{\mathscr{B}}(\tilde{S}_n, Z_\Sigma) \le (\sqrt{6}\beta^3)^{-1} R_3 n^{-1/2} + (2\sqrt{6})^{-1}\{h_1(\beta)\mathbb{E}\|\Sigma^{-1/2}X_1\|^4 + h_2(\beta)(d^2 + 2d)\} n^{-1},$$

is fully explicit, nonasymptotic, and is analogous to the terms in the classical Edgeworth series.

The proof of Theorem 2.2 uses an analogous approach. First, we write the triangle inequality

$$\Delta_{\mathscr{B}}(S_n, S_{T,n}) \le \Delta_{\mathscr{B}}(S_n, \tilde{S}_n) + \Delta_{\mathscr{B}}(S_{T,n}, \tilde{S}_{T,n}) + \Delta_{\mathscr{B}}(\tilde{S}_n, \tilde{S}_{T,n}), \tag{3.8}$$

where $\tilde{S}_{T,n} = n^{-1/2}\sum_{i=1}^{n} Y_{T,i}$ is constructed similarly to the approximating sum $\tilde{S}_n$ (see (9.19)–(9.21) in [47] for their explicit definitions). The terms $\Delta_{\mathscr{B}}(S_n, \tilde{S}_n)$, $\Delta_{\mathscr{B}}(S_{T,n}, \tilde{S}_{T,n})$ are bounded similarly to $\Delta_{\mathscr{B}}(S_n, \tilde{S}_n)$ in (3.2), and the term $\Delta_{\mathscr{B}}(\tilde{S}_n, \tilde{S}_{T,n})$ is expanded in the same way as $\Delta_{\mathscr{B}}(\tilde{S}_n, Z_\Sigma)$ using the smooth normal components, the telescoping sum representations, and the Taylor series expansions.

Let us emphasize that the proposed proof technique is much more simple than many existing methods of deriving rates of convergence in the normal approximation. Furthermore, it is not restricted to the case when an approximation distribution is normal and it allows to obtain explicit error terms and constants under very mild conditions. To the best of our knowledge, this is a novel technique and it has not been used in earlier literature.

REMARK 3.1. I submitted the first version of this paper containing the new proof technique to *The Annals of Statistics* on June 5, 2020 (submission number AOS2006-011). In about 3 months after that Lopes [31] (see https://arxiv.org/abs/2009.06004v1) used a very similar approach and labeled it as "implicit Gaussian smoothing." Lopes [31] considered the problem of bounding the distance

$$\Delta_{\mathscr{R}}(S_n, Z_\Sigma) = \sup_{R\in\mathscr{R}} \big|\mathbb{P}(S_n \in R) - \mathbb{P}(Z_\Sigma \in R)\big|, \tag{3.9}$$

where $\mathscr{R}$ is a set of all hyperrectagles in $\mathbb{R}^d$. The approach used by Lopes [31] has major similarities with the approach that we present here. These ideas play crucial role for my solution in the present paper as well as for the proofs in Lopes [31]. Below we describe these similarities:

(i) Lopes [31] uses the normal part of a random sum similarly to (3.3) in order to represent its probability distribution of a sum of independent random vectors as an expected value of a smooth function (via the Gaussian distribution) similarly to (3.5), (3.6).

(ii) The smooth function obtained via the Gaussian distribution in part (i) is expanded using the Taylor series as in (3.7) (see also (9.7)–(9.9) in Section 9 in the Supplementary Material [47]); Lopes [31] uses the second-order Taylor expansion with the remainder in the same form as here (see (9.15), (9.6)). The Taylor expansion is applied to the differences in the telescoping sum similarly to (3.7) and (9.7)–(9.9).

(iii) The expansion in (ii) allows to apply the moment matching strategy as in (9.11), (9.13).

(iv) Lopes [31] uses the induction technique based on the proof by Bentkus [5], similarly to our approach for (3.2).

I would like to emphasize that the approach presented here as well as the ideas from our 2016 paper [46] are new and first appeared in my work, previous to Lopes [31].

4. Approximation bounds for symmetric distributions and optimality of the error rate. In this section, we consider the case when the probability distribution of $X_i - \mathbb{E}X_i$ has some symmetry properties this assumption can be formulated in terms of the condition (4.1) on moments of X_i. Suppose for i.i.d. $\mathbb{R}^d$-valued random vectors $\{X_i\}_{i=1}^n$ that $\mathbb{E}|X_i^{\otimes 6}| < \infty$ and their covariance matrix $\Sigma := \operatorname{Var}(X_i)$ is p.d. Without loss of generality, assume that $\mathbb{E}X_i = 0$. Let $X = (x_1, \dots, x_d)$ be an i.i.d. copy of X_i, we assume that

$$\mathbb{E}p(x_1, \dots, x_d) = 0 \tag{4.1}$$

for any monomial $p : \mathbb{R}^d \mapsto \mathbb{R}$ that has degree ≤ 5 and contains an odd power of x_j for at least one $j \in \{1, \dots, d\}$. In addition, we assume that there exist a random vector U_L in $\mathbb{R}^d$ with $\mathbb{E}|U_L^{\otimes 6}| < \infty$ and a p.d. covariance matrix $\Sigma_L \in \mathbb{R}^{d\times d}$ such that the following moment matching property holds for $Z_{\Sigma,L} \sim \mathcal{N}(0, \Sigma_L)$ independent from U_L and $L := Z_{\Sigma_L} + U_L$:

$$\mathbb{E}(L^{\otimes j}) = \mathbb{E}(X_i^{\otimes j}) \quad \forall j \in \{1, \dots, 5\}. \tag{4.2}$$

We introduced this condition in an earlier paper [46]; Lemmas 3.1, 3.2 in that paper show that under certain conditions on the support of X_i there exists a probability distribution $\mathcal{L}(L)$ which complies with these conditions (see Lemma 8.2 in Section 8 in [47] for further details). Also, because of property (4.1), it is sufficient to assume that there exist only 6 finite absolute moments (instead of 7 finite absolute moments as stated in Lemma 3.1 in [46] for the general case).

THEOREM 4.1. *Let $\{X_i\}_{i=1}^n$ follow the conditions above, take $\lambda_z^2 > 0$ equal to the smallest eigenvalue of Σ_L, and $Z_\Sigma \sim \mathcal{N}(0, \Sigma)$ in $\mathbb{R}^d$, then it holds*

$$\begin{aligned}\Delta_{\mathscr{B}}(S_n, Z_\Sigma) \le C_{B,6}\{\lambda_z^{-6}\mathbb{E}(\|X_1\|^6 + \|L_1\|^6)\}^{1/4} n^{-1/2} \\ + (4!)^{-1/2}\lambda_z^{-4}\|\mathbb{E}(X_1^{\otimes 4}) - \mathbb{E}(Z_\Sigma^{\otimes 4})\|_{\mathrm{F}} n^{-1} \\ + (6!)^{-1/2}\lambda_z^{-6}\{\mathbb{E}\|U_L\|^6 + \mathbb{E}\|Z_\Sigma\|^6\} n^{-2},\end{aligned}$$

where $C_{B,6} = 2.9 C_{\ell_2} C_{\phi,6} \ge 2.9$ is a constant independent from d, n, and probability distribution of X_i (it is discussed in detail in Section 8 in [47]). Let $m_{6,\mathrm{sym}}$ denote the maximum of the 6-th moments of the coordinates of X_1, L_1, Z_Σ, then the above inequality implies

$$\begin{aligned}\Delta_{\mathscr{B}}(S_n, Z_\Sigma) &\le (4!)^{-1/2}\lambda_z^{-4}\|\mathbb{E}(X_1^{\otimes 4}) - \mathbb{E}(Z_\Sigma^{\otimes 4})\|_{\mathrm{F}} n^{-1} \\ &\quad + C_{B,6}(\lambda_z^{-6} m_{6,\mathrm{sym}})^{1/4} d^{3/4} n^{-1/2} + (6!)^{-1/2}(\lambda_z^{-6} m_{6,\mathrm{sym}}) d^3 n^{-2} \\ &\le 8^{-1/2}\lambda_z^{-4}\|\mathbb{E}(X_1^{\otimes 4}) - \mathbb{E}(Z_\Sigma^{\otimes 4})\|_{\max} d n^{-1} \\ &\quad + C_{B,6}(\lambda_z^{-6} m_{6,\mathrm{sym}})^{1/4} d^{3/4} n^{-1/2} + (6!)^{-1/2}(\lambda_z^{-6} m_{6,\mathrm{sym}}) d^3 n^{-2}.\end{aligned} \tag{4.3}$$

Below we consider the example by Portnoy [35] (Theorem 2.2 in [35]), using the notation in the present paper, and we derive a lower bound for $\Delta_{\mathscr{B}}(S_n, Z_\Sigma)$ with the ratio between d and n similar to the error term in Theorem 4.1. Proposition 4.1 and Lemma 4.1 imply that for $\{X_i\}_{i=1}^n$, Z as in Theorem 4.2, and for sufficiently large d, n

$$C d^{3/2}/n \le \Delta_{\mathscr{B}}(S_n, Z) \le C(d^{3/2}/n)^{1/2}.$$

THEOREM 4.2 (Portnoy [35]). *Let i.i.d. random vectors X_i have the following mixed normal distribution*:

$$X_i|\{Z_i\} \sim \mathcal{N}(0, Z_i Z_i^T) \quad \text{for i.i.d. } Z_i \sim \mathcal{N}(0, I_d).$$

Let also $S_n = n^{-1/2}\sum_{i=1}^n X_i$, $Z \sim \mathcal{N}(0, I_d)$. If $d \to \infty$ such that $d/n \to 0$ as $n \to \infty$, then

$$\|S_n\|^2 = \|Z\|^2 + D_n, \quad D_n = O_p(d^2/n).$$

PROPOSITION 4.1. *Let $\{X_i\}_{i=1}^n$ and Z be as in Theorem 4.2. If $d \to \infty$ such that $d/n \to 0$ as $n \to \infty$, then*

$$\Delta_{\mathscr{B}}(S_n, Z) \ge \Delta_L(\|S_n\|^2, \|Z\|^2) = O(d^{3/2}/n),$$

where $\Delta_L(X, Y) := \inf\{\varepsilon > 0 : G(x - \varepsilon) - \varepsilon \le F(x) \le G(x + \varepsilon) + \varepsilon \text{ for all } x \in \mathbb{R}\}$ denotes the Lévy distance between the c.d.f.-s of X and Y, equal $F(x)$ and $G(x)$, respectively.

LEMMA 4.1. *$\{X_i\}_{i=1}^n$ in Theorem 4.2 satisfy conditions of Theorem 4.1 for $\lambda_z = (1 - \sqrt{2/5})^{1/2}$.*

5. Bootstrap approximation. Here we consider the nonparametric or Efron's bootstrapping scheme for S_n (by Efron [18], Efron and Tibshirani [19]) and study its accuracy in the framework of Theorems 2.1 and 2.2. Let $\{X_i\}_{i=1}^n$ be i.i.d. $\mathbb{R}^d$-valued random vectors with $\mathbb{E}|X_i^{\otimes 4}| < \infty$, p.d. $\Sigma := \operatorname{Var} X_i$ and $\mu := \mathbb{E}X_i$. Introduce resampled variables $X_1^*, \dots, X_n^*$ with zero mean, according to the nonparametric bootstrapping scheme

$$\mathbb{P}^*(X_i^* = X_j - \bar{X}) = 1/n \quad \forall i, j = 1, \dots, n, \tag{5.1}$$

where $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ and $\mathbb{P}^*(\cdot) := \mathbb{P}(\cdot|\{X_i\}_{i=1}^n)$, $\mathbb{E}^*(\cdot) := \mathbb{E}(\cdot|\{X_i\}_{i=1}^n)$. Hence, $\{X_j^*\}_{j=1}^n$ are i.i.d. and

$$\mathbb{E}^*(X_j^*) = \mathbb{E}(X_i - \mu) = 0, \qquad \mathbb{E}^*(X_j^{*\otimes k}) = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^{\otimes k} \approx_{\mathbb{P}} \mathbb{E}(X_i - \mu)^{\otimes k}$$

for $k \ge 1$; the sign $\approx_{\mathbb{P}}$ denotes "closeness" with high probability. Denote the centered sum S_n and its bootstrap approximation as follows:

$$S_{0,n} := n^{-1/2}\sum_{i=1}^n (X_i - \mu), \qquad S_n^* := n^{-1/2}\sum_{i=1}^n X_i^*.$$

In order to quantify the accuracy of the bootstrap approximation of the probability distribution of $S_{0,n}$, we compare the empirical moments $n^{-1}\sum_{i=1}^n (X_i - \bar{X})^{\otimes k}$ and the population moments $\mathbb{E}(X_i - \mu)^{\otimes k}$ for $k = 2, 3$, using exponential concentration bounds. For this purpose, we introduce condition (5.2) on tail behavior of coordinates of $X_i - \mu$. Here we follow the notation from Section 2.3 by Boucheron, Lugosi and Massart [9]. A random real-valued variable x belongs to class $\mathcal{G}(\sigma^2)$ of sub-Gaussian random variables with variance factor $\sigma^2 > 0$ if

$$\mathbb{E}\{\exp(\gamma x)\} \le \exp(\gamma^2\sigma^2/2) \quad \forall \gamma \in \mathbb{R}. \tag{5.2}$$

We assume that every coordinate of random vectors $X_i - \mu$, $i = 1, \dots, n$ belongs to the class $\mathcal{G}(\sigma^2)$ for some $\sigma^2 > 0$. Let also

$$\begin{aligned} &C_1(t) := 2\{4\sqrt{2t} + 3tn^{-1/2}\}, \qquad C_2(t) := 4\sqrt{2}(\sqrt{8}t + t^{3/2}n^{-1/2}), \\ &t_* := \log n + \log(2dn + d^2 + 3d), \qquad C_{j*} := C_j(t_*) \text{ for } j = 1, 2. \end{aligned} \tag{5.3}$$

$\lambda_{\min}(\Sigma) > 0$ denotes the smallest eigenvalue of the covariance matrix Σ. We consider d and n such that

$$\sigma^2 C_{1*} d/\sqrt{n} < \lambda_{\min}(\Sigma). \tag{5.4}$$

This condition allows to ensure that the approximation bound in Theorem 5.1 holds with high probability. Recall that $h_1(\beta) = (1-\beta^2)^2\beta^{-4} + (1-\beta^2)^{-1}\beta^{-4}$ for $\beta \in (0,1)$.

THEOREM 5.1. *If the conditions introduced above are fulfilled, then it holds with probability* $\geq 1 - n^{-1}$

$$\Delta_{\mathscr{B}}(S_{0,n}, S_n^*) \leq \delta_{\mathscr{B}},$$

where $\delta_{\mathscr{B}} = \delta_{\mathscr{B}}(d, n, \mathcal{L}(X_i))$ *is defined as follows*:

$$\delta_{\mathscr{B}} := (\sqrt{2}\beta^2\lambda_0^2)^{-1}\{\sigma^2 C_{1*} d/\sqrt{n}\} \tag{5.5}$$

$$+ (\sqrt{6}\beta^3\lambda_0^3)^{-1}[4\sigma\sqrt{2dn^{-2}t_*}\{\|\Sigma\|_{\mathrm{F}} + \sigma^2 t_* d/n\} \tag{5.6}$$

$$+ \sigma^2 d^{3/2} n^{-1} C_{2*}\{1 + 3n^{-1/2}\} + \|\mathbb{E}(X_1 - \mu)^{\otimes 3}\|_{\mathrm{F}} n^{-1/2}] \tag{5.7}$$

$$+ 4\sqrt{2}C_{B,4}\lambda_0^{-2}\{h_1(\beta)[\mathbb{E}\|X_1 - \mu\|^4 + 8(1 + n^{-2})\{2\sigma^2 t_* d/n\}^2]$$

$$+ (d^2 + 2d)(3\|\Sigma\|^2 + 2\{\sigma^2 C_{1*} d/\sqrt{n}\}^2 + 1/2)\}^{1/2} n^{-1/2}$$

$$+ 2(\sqrt{6}\lambda_0^4)^{-1}\{h_1(\beta)[\mathbb{E}\|X_1 - \mu\|^4 + 8(1 + n^{-2})\{2\sigma^2 t_* d/n\}^2]$$

$$+ (d^2 + 2d)[3\|\Sigma\|^2 + 2\{\sigma^2 C_{1*} d/\sqrt{n}\}^2]\} n^{-1}$$

for arbitrary $\beta \in (0,1)$ *and for* $\lambda_0^2 := \lambda_{\min}(\Sigma) - \sigma^2 C_{1*} d/\sqrt{n}$.

REMARK 5.1. The explicit approximation error $\delta_{\mathscr{B}}$ in Theorem 5.1 allows to evaluate accuracy of the bootstrap in terms of d, n, σ^2, and moments of X_i. In general, $\delta_{\mathscr{B}} \leq C_*\{\sqrt{d^2/n} + d^2/n\}$ (for C_* depending on the log-term t_* and moments of X_i), however the expression for $\delta_{\mathscr{B}}$ provides a much more detailed and accurate characterization of the approximation error. The proof of this result is based on the second statement in Theorem 2.2 (for Σ and Σ_T not necessarily equal to each other). The first term on the right-hand side of (5.5) and the summands in (5.6), (5.7) characterize the distances between the population moments $\mathbb{E}(X_i - \mu)^{\otimes k}$ and their consistent estimators $\mathbb{E}^*(X_j^{*\otimes k})$ (for $k = 2$ and 3, respectively). The rest of the summands in the expression for $\delta_{\mathscr{B}}$ correspond to the higher-order remainder terms which leads to smaller error terms for a sufficiently large n.

6. Elliptic confidence sets. An elliptic confidence set is one of the major types of confidence regions in statistical theory and applications. They are commonly constructed for parameters of (generalized) linear regression models, in ANOVA methods, and in various parametric models where a multivariate statistic is asymptotically normal. As for example, in the case of the score function considered in Section 7. See, for instance, Friendly, Monette and Fox [22] for an overview of statistical problems and applications involving elliptic confidence regions.

In this section, we construct confidence regions for an expected value of i.i.d. random vectors $\{X_i\}_{i=1}^n$, using the bootstrap-based quantiles. Since the considered set-up is rather general, the provided results can be used for various applications, where one is interested in estimating a mean of an observed sample in a nonasymptotic and multivariate setting. See, for example, Arlot, Blanchard and Roquain [1], where the authors constructed nonasymptotic confidence bounds in ℓ_r-norm for the mean value of high dimensional random vectors and considered a number of important practical applications.

Let $W \in \mathbb{R}^{d\times d}$ be a p.d. symmetric matrix. W is supposed to be known, it defines the quadratic form of an elliptic confidence set: $B_W(x_0, r) := \{x \in \mathbb{R}^d : (x-x_0)^T W(x-x_0) \le r\}$, for $x_0 \in \mathbb{R}^d$, $r \ge 0$. There are various ways of how one can interpret and use W in statistical models. For example, W can serve for weighting an impact of residuals in linear regression models in the presence of errors' heteroscedasticity (cf. weighted least squares estimation); for regularized least squares estimators in the linear regression model (e.g., ridge regression) W denotes a regularized covariance matrix of the LSE; see [22] for further examples.

In Proposition 6.1 below, we construct an elliptic confidence region for $\mathbb{E}X_1$ based on the bootstrap approximation established in Section 5. Let $\bar{X}^* := n^{-1}\sum_{j=1}^n X_j^*$ for the i.i.d. bootstrap sample $\{X_j^*\}_{j=1}^n$ generated from the empirical distribution of $\{X_i - \bar{X}\}_{i=1}^n$ for $\bar{X} = n^{-1}\sum_{i=1}^n X_i$. Let also

$$q_\alpha^* := \inf\{t > 0 : (1-\alpha) \le \mathbb{P}^*(n^{1/2}\|W^{1/2}\bar{X}^*\| \le t)\}$$

denote $(1-\alpha)$-quantile of the bootstrap statistic $n^{1/2}\|W^{1/2}\bar{X}^*\|$ for arbitrary $\alpha \in (0, 1)$. We assume that coordinates of vectors $\{W^{1/2}(X_i - \mathbb{E}X_i)\}_{i=1}^n$ are sub-Gaussian with variance factor $\sigma_W^2 > 0$ (i.e., condition (5.2) is fulfilled). Let also d, n be such that $\sigma_W^2 C_{1*} d/\sqrt{n} < \lambda_{\min}(W^{1/2}\Sigma W^{1/2})$ (for C_{1*} defined in (5.3)). Theorem 5.1 implies the following statement.

PROPOSITION 6.1. *If the conditions above are fulfilled, it holds*

$$\left|\mathbb{P}(n^{1/2}\|W^{1/2}(\bar{X} - \mathbb{E}X_1)\| \le q_\alpha^*) - (1-\alpha)\right| \le \delta_W.$$

δ_W is analogous to $\delta_{\mathscr{B}}$, where we take $\Sigma := W^{1/2}\Sigma W^{1/2}$, $\sigma^2 := \sigma_W^2$, etc. A detailed definition of δ_W is given in (10.2) *in the Supplementary Material* [47], *see also Remark* 5.1 *for the discussion about its dependence on d and n.*

7. Score tests. Let $y = (y_1, \ldots, y_n)$ be an i.i.d. sample from a p.d.f. or a p.m.f. $p(x)$. Let also $\mathcal{P} := \{p(x;\theta) : \theta \in \Theta \subseteq \mathbb{R}^d\}$ denote a known parametric family of probability distributions. The unknown function $p(x)$ does not necessarily belong to the parametric family $\mathcal{P}$, in other words the parametric model can be misspecified. Following the renown aphorism of Box [10] "All models are wrong, but some are useful," it is widely recognized that in general a (parametric) statistical model cannot be considered exactly correct. See, for example, White [44], Gustafson [23], Wit, Heuvel and Romeijn [45], Section 1.1.4 by McCullagh and Nelder [34], and p. 2 by Bickel and Doksum [8]. Hence, it is of particular importance to design methods of statistical inference that are *robust to model misspecification*. In this section, we propose *a bootstrap score test procedure* which is valid even in case when the parametric model $\mathcal{P}$ is misspecified.

Let $s(\theta) = s(\theta, y)$ and $I(\theta)$ denote the score function and the Fisher information matrix corresponding to the introduced parametric model

$$s(\theta) := \sum_{i=1}^n \partial \log p(y_i;\theta)/\partial\theta, \qquad I(\theta) := \operatorname{Var}\{s(\theta)\}.$$

We suppose that the standard regularity conditions on the parametric family $\mathcal{P}$ are fulfilled. Let $\theta_0 := \operatorname{argmin}_{\theta\in\Theta} \mathbb{E}\log(p(y_i)/p(y_i;\theta))$ denote the parameter which corresponds to the projection of $p(x)$ on the parametric family $\mathcal{P}$ w.r.t. the Kullback-Leibler divergence (also known as the relative entropy).

Consider a simple hypothesis $H_0 : \theta_0 = \theta'$. Rao's score test (by Rao [37]) for testing H_0 is based on the following test statistic and its convergence in distribution to χ_d^2:

$$R(\theta') := s(\theta')^T \{I(\theta')\}^{-1} s(\theta') \overset{d|H_0}{\to} \chi_d^2, \quad n \to \infty, \tag{7.1}$$

provided that matrix $I(\theta')$ is p.d. The sign $\overset{d|H_0}{\to}$ denotes convergence in distribution under H_0. Matrix $I(\theta')$ can be calculated explicitly for a known θ' if one assumes that $p(x) \in \mathcal{P}$, that is, if the parametric model is correct. However, if $p(x)$ does not necessarily belong to the considered parametric class $\mathcal{P}$, then neither $I(\theta')$ nor the probability distribution of $s(\theta')$ can be calculated in an explicit way under the general assumptions considered here. In this case, the Fisher information matrix $I(\theta)$ is typically estimated using the Hessian of the log-likelihood function $\sum_{i=1}^n \log p(y_i; \theta)$. However, the standardization with an empirical Fisher information may considerably reduce the power of the score test for a small sample size n (see Rao [38] and Freedman [21]).

Below we consider a bootstrap score test for testing simple hypothesis H_0, under possible misspecification of the parametric model. Denote

$$\tilde{R}(\theta') := \|s(\theta')/\sqrt{n}\|^2.$$

One can consider $s(\theta') = \sum_{i=1}^n X_i$, where random vectors $X_i := \partial \log p(y_i; \theta')/\partial\theta'$ are i.i.d. with $\mathbb{E}X_i = 0$ under H_0. Introduce the bootstrap approximations of $s(\theta')$ and $\tilde{R}(\theta')$:

$$s^*(\theta') := \sum_{i=1}^n X_i^*, \qquad R^*(\theta') := \|s^*(\theta')/\sqrt{n}\|^2,$$

where $\{X_i^*\}_{i=1}^n$ are sampled according to Efron's bootstrap scheme (5.1). Let also

$$t_\alpha^* := \inf\{t > 0 : (1-\alpha) \le \mathbb{P}^*(R^*(\theta') \le t)\}$$

denote $(1-\alpha)$-quantile of the bootstrap score statistic for arbitrary $\alpha \in (0, 1)$. Suppose that coordinates of vectors $X_i = \partial \log p(y_i; \theta')/\partial\theta'$ satisfy condition (5.2) with variance factor $\sigma_s^2 > 0$. Let also d and n be such that $\sigma_s^2 C_{1*} d/\sqrt{n} < \lambda_{\min}(I(\theta'))/n$. Then Theorem 5.1 implies the following statement, which characterizes accuracy of the bootstrap score test under H_0.

THEOREM 7.1 (Bootstrap score test). *If the conditions above are fulfilled, it holds*

$$|\mathbb{P}_{H_0}(\tilde{R}(\theta') > t_\alpha^*) - \alpha| \le \delta_R,$$

where δ_R is analogous to $\delta_{\mathscr{B}}$ up to the terms σ^2 and Σ. A detailed definition of δ_R is given in (10.3) *in the Supplementary Material* [47]; *see also Remark* 5.1 *for the discussion about its dependence on d and n.*

The following statement provides a finite sample version of Rao's score test based on (7.1) for testing simple hypothesis $H_0 : \theta_0 = \theta'$. Here we require also the finite 4-th moment of the score in order to apply the higher-order approximation from Theorem 2.1.

THEOREM 7.2 (Nonasymptotic version of Rao's score test). *Suppose that $p(x) \equiv p(x, \theta_0)$ for some $\theta_0 \in \Theta$, that is, there is no misspecification in the considered parametric model. Let also $\tilde{X}_i := \sqrt{n}\{I(\theta')\}^{-1/2} \partial \log p(y_i; \theta')/\partial\theta'$ denote the marginal standardized score for the i-th observation and $\tilde{\Sigma} := n^{-1} I(\theta')$. Suppose that $\mathbb{E}|\tilde{X}_i^{\otimes 4}| < \infty$, then the* *asymptotic poperty* (7.1) *for testing* $H_0 : \theta_0 = \theta'$ *can be represented in the finite sample form as follows*:

$$\begin{aligned}\sup_{\alpha\in(0,1)} &\left|\mathbb{P}_{H_0}\big(R(\theta') > q(\alpha;\chi^2_d)\big) - \alpha\right| \\ &\le (\sqrt{6}\beta^3)^{-1}\|\mathbb{E}(\tilde{X}_1^{\otimes 3})\|_{\mathrm{F}} n^{-1/2} \\ &\quad + 2C_{B,4}\|\tilde{\Sigma}^{-1}\|\|\tilde{\Sigma}\|\{(h_1(\beta) + (4\beta^4)^{-1})\mathbb{E}\|\tilde{X}_1\|^4 + d^2 + 2d\}^{1/2} n^{-1/2} \\ &\quad + (2\sqrt{6})^{-1}\{h_1(\beta)\mathbb{E}\|\tilde{X}_1\|^4 + h_2(\beta)(d^2+2d)\} n^{-1},\end{aligned}$$

where $q(\alpha;\chi^2_d)$ *denotes the* $(1-\alpha)$*-quantile of* χ^2_d *distribution. The inequality holds for any* $\beta \in (0,1)$, *functions* h_1, h_2 *are defined in* (2.2), *constant* $C_{B,4} \ge 9.5$ *is described in the statement of Theorem* 2.1.

Funding. Support by the National Science Foundation Awards CAREER DMS-2048028 and DMS-1712990 is gratefully acknowledged.

SUPPLEMENTARY MATERIAL

Supplement to "New Edgeworth-type expansions with finite sample guarantees" (DOI: 10.1214/22-AOS2192SUPP; .pdf). Supplementary material contains proofs of the main results from Sections 2–7, statements from [46] which are used in the proofs of main results, and the proof of optimality of the Gaussian anticoncentration bound over set $\mathscr{B}$.

REFERENCES

[1] ARLOT, S., BLANCHARD, G. and ROQUAIN, E. (2010). Some nonasymptotic results on resampling in high dimension. I. Confidence regions. *Ann. Statist.* **38** 51–82. MR2589316 https://doi.org/10.1214/08-AOS667

[2] BALL, K. (1993). The reverse isoperimetric problem for Gaussian measure. *Discrete Comput. Geom.* **10** 411–420. MR1243336 https://doi.org/10.1007/BF02573986

[3] BARBE, P. and BERTAIL, P. (1995). *The Weighted Bootstrap. Lecture Notes in Statistics* **98**. Springer, New York. MR2195545 https://doi.org/10.1007/978-1-4612-2532-4

[4] BELLONI, A., BUGNI, F. and CHERNOZHUKOV, V. (2018). Subvector inference in partially identified models with many moment inequalities. Preprint. Available at arXiv:1806.11466.

[5] BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117 https://doi.org/10.1016/S0378-3758(02)00094-0

[6] BERRY, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136. MR0003498 https://doi.org/10.2307/1990053

[7] BHATTACHARYA, R. N. and RAO, R. R. (2010). *Normal Approximation and Asymptotic Expansions. Classics in Applied Mathematics* **64**. SIAM, Philadelphia, PA. MR3396213 https://doi.org/10.1137/1.9780898719895.ch1

[8] BICKEL, P. J. and DOKSUM, K. A. (2015). *Mathematical Statistics—Basic Ideas and Selected Topics. Vol.* 1, 2nd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3445928

[9] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[10] BOX, G. E. P. (1976). Science and statistics. *J. Amer. Statist. Assoc.* **71** 791–799. MR0431440

[11] CHATTERJEE, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. MR2294976 https://doi.org/10.1214/009117906000000575

[12] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

[13] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040 https://doi.org/10.1007/s00440-014-0565-9

[14] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

[15] CRAMÉR, H. (1928). On the composition of elementary errors. *Scand. Actuar. J.* **1928** 13–74.

[16] EDGEWORTH, F. Y. (1896). The asymmetrical probability-curve. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **41** 90–99.

[17] EDGEWORTH, F. Y. (1905). The law of error. In *Pros. Camb. Philos. Soc* **20** 16–65.

[18] EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. MR0515681

[19] EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. MR1270903 https://doi.org/10.1007/978-1-4899-4541-9

[20] ESSEEN, C.-G. (1942). On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astron. Fys.* **28A** 19. MR0011909

[21] FREEDMAN, D. A. (2007). How can the score test be inconsistent? *Amer. Statist.* **61** 291–295. MR2411784 https://doi.org/10.1198/000313007X243061

[22] FRIENDLY, M., MONETTE, G. and FOX, J. (2013). Elliptic insights: Understanding statistical methods through elliptical geometry. *Statist. Sci.* **28** 1–39. MR3075337 https://doi.org/10.1214/12-STS402

[23] GUSTAFSON, P. (2001). On measuring sensitivity to parametric model misspecification. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 81–94. MR1811992 https://doi.org/10.1111/1467-9868.00277

[24] HALL, P. (1992). *The Bootstrap and Edgeworth Expansion. Springer Series in Statistics*. Springer, New York. MR1145237 https://doi.org/10.1007/978-1-4612-4384-7

[25] HOROWITZ, J. L. (2001). The bootstrap. *Handb. Econom.* **5** 3159–3228.

[26] KLIVANS, A. R., O'DONNELL, R. and SERVEDIO, R. A. (2008). Learning geometric concepts via Gaussian surface area. In *Foundations of Computer Science*, 2008. *FOCS'*08. *IEEE* 49*th Annual IEEE Symposium on* 541–550. IEEE Comput. Soc., Los Alamitos.

[27] KOLASSA, J. E. (1994). *Series Approximation Methods in Statistics. Lecture Notes in Statistics* **88**. Springer, New York. MR1295242 https://doi.org/10.1007/978-1-4757-4275-6

[28] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data. Springer Series in Statistics*. Springer, New York. MR2001447 https://doi.org/10.1007/978-1-4757-3803-2

[29] LINDEBERG, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15** 211–225. MR1544569 https://doi.org/10.1007/BF01494395

[30] LIU, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708. MR0964947 https://doi.org/10.1214/aos/1176351062

[31] LOPES, M. E. (2022). Central limit theorem and bootstrap approximation in high dimensions: $1/\sqrt{n}$ rates via implicit smoothing. *Ann. Statist.*, to appear. arXiv:2009.06004.

[32] MAMMEN, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.* **21** 255–285. MR1212176 https://doi.org/10.1214/aos/1176349025

[33] MCCULLAGH, P. (1987). *Tensor Methods in Statistics. Monographs on Statistics and Applied Probability*. CRC Press, London. MR0907286

[34] MCCULLAGH, P. and NELDER, J. A. (1983). *Generalized Linear Models. Monographs on Statistics and Applied Probability*. CRC Press, London. MR0727836 https://doi.org/10.1007/978-1-4899-3244-0

[35] PORTNOY, S. (1986). On the central limit theorem in R^p when $p \to \infty$. *Probab. Theory Related Fields* **73** 571–583. MR0863546 https://doi.org/10.1007/BF00324853

[36] PRÆSTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086. MR1245301

[37] RADHAKRISHNA RAO, C. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proc. Camb. Philos. Soc.* **44** 50–57. MR0024111

[38] RAO, C. R. (2005). Score test: Historical review and recent developments. In *Advances in Ranking and Selection, Multiple Comparisons, and Reliability. Stat. Ind. Technol.* 3–20. Birkhäuser, Boston, MA. MR2111500 https://doi.org/10.1007/0-8176-4422-9_1

[39] SHEVTSOVA, I. G. (2014). On absolute constants in inequalities of Berry–Esseen type. *Dokl. Akad. Nauk* **456** 650–654. MR3287912 https://doi.org/10.1134/s1064562414030338

[40] SKOVGAARD, I. M. (1986). On multivariate Edgeworth expansions. *Int. Stat. Rev.* **54** 169–186. MR0962933 https://doi.org/10.2307/1403142

[41] TCHEBYCHEFF, P. (1890). Sur deux théorèmes relatifs aux probabilités. *Acta Math.* **14** 305–315. MR1554802 https://doi.org/10.1007/BF02413327

[42] TROTTER, H. F. (1959). An elementary proof of the central limit theorem. *Arch. Math.* **10** 226–234. MR0108847 https://doi.org/10.1007/BF01240790

[43] WANG, M., DAO DUC, K., FISCHER, J. and SONG, Y. S. (2017). Operator norm inequalities between tensor unfoldings on the partition lattice. *Linear Algebra Appl.* **520** 44–66. MR3611456 https://doi.org/10.1016/j.laa.2017.01.017

[44] WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163 https://doi.org/10.2307/1912526

[45] WIT, E., VAN DEN HEUVEL, E. and ROMEIJN, J.-W. (2012). 'All models are wrong...': An introduction to model uncertainty. *Stat. Neerl.* **66** 217–236. MR2955417 https://doi.org/10.1111/j.1467-9574.2012.00530.x

[46] ZHILOVA, M. (2020). Nonclassical Berry–Esseen inequalities and accuracy of the bootstrap. *Ann. Statist.* **48** 1922–1939. MR4134780 https://doi.org/10.1214/18-AOS1802

[47] ZHILOVA, M. (2022). Supplement to "New Edgeworth-type expansions with finite sample guarantees." https://doi.org/10.1214/22-AOS2192SUPP