arXiv:1611.02686v2 [math.ST] 1 Jun 2020

NONCLASSICAL BERRY–ESSEEN INEQUALITIES AND ACCURACY OF THE BOOTSTRAP

BY MAYYA ZHILOVA*

School of Mathematics
Georgia Institute of Technology

We study accuracy of bootstrap procedures for estimation of quantiles of a smooth function of a sum of independent sub-Gaussian random vectors. We establish higher-order approximation bounds with error terms depending on a sample size and a dimension explicitly. These results lead to improvements of accuracy of a weighted bootstrap procedure for general log-likelihood ratio statistics. The key element of our proofs of the bootstrap accuracy is a multivariate higher-order Berry–Esseen inequality. We consider a problem of approximation of distributions of two sums of zero mean independent random vectors, such that summands with the same indices have equal moments up to at least the second order. The derived approximation bound is uniform on the sets of all Euclidean balls. The presented approach extends classical Berry–Esseen type inequalities to higher-order approximation bounds. The theoretical results are illustrated with numerical experiments.

1. Introduction. In this paper we study accuracy of bootstrap procedures for estimation of quantiles of statistics of the form $\|S_n\|$ or $f(S_n)$, where $\|\cdot\|$ denotes the ℓ_2-norm, and $f(\cdot): \mathbb{R}^p \mapsto \mathbb{R}$ is a twice continuously differentiable function with bounded second derivative,

$$S_n := n^{-1/2}\sum_{i=1}^{n}X_i$$

for independent sub-Gaussian random vectors $X_1, \ldots, X_n \in \mathbb{R}^p$ with positive definite covariance matrices $\mathrm{Var}(X_i)\ \forall\, i = 1, \ldots, n$. We consider the non-asymptotic setting, when the leading approximation errors depend on n and the dimension p explicitly. This setting allows to assess accuracy and limitations of a bootstrap approximation in terms of the dimension p and the sample size n. Estimation of distribution of statistics of the types $\|S_n\|$

or $f(S_n)$ is necessary for construction of confidence sets and hypothesis testing in some important statistical models and problems, such as linear regression model with unknown distribution of errors, general log-likelihood ratio statistic, construction of confidence sets for multivariate sample mean.

We focus on two basic bootstrapping procedures. The first method considered here is the Efron's bootstrap (introduced by Efron [14] in 1979), where the resampling is performed uniformly at random with replacement from the i.i.d. data $X_1, \ldots, X_n \in \mathbb{R}^p$. In this case the bootstrap samples $X_1^*, \ldots, X_n^*$ have the distribution $\mathbf{P}^*(X_j^* = X_i - \bar{X}) = 1/n \ \forall i, j = 1, \ldots, n$, where $\bar{X} = n^{-1}\sum_{i=1}^n X_i$, and $\mathbf{P}^*(\cdot) := \mathbf{P}(\cdot \mid X_1, \ldots, X_n)$. Define for the sum S_n its bootstrap version:

$$S_n^* := n^{-1/2}\sum_{i=1}^n X_i^*.$$

One of the main results of the paper is the following uniform approximation bound on the set $\mathscr{B}$ of all Euclidean balls in $\mathbb{R}^p$ which holds with high probability:

$$(1.1) \qquad \sup_{B \in \mathscr{B}} \left| \mathbf{P}\left(S_n \in B\right) - \mathbf{P}^*\left(S_n^* \in B\right) \right| \leq C_{\sigma,K,z}\left\{p^{K/(K-2)}/n\right\}^{1/2},$$

where $K \geq 3$ is a natural number, and constant $C_{\sigma,K,z}$ depends (up to log-terms) on K, on value C_z introduced in (2.4) in Section 2, and on constant $\sigma^2 > 0$ which comes from the following condition on the moment generating function of X_i: $\mathbf{E}\left\{\exp(\alpha^\top X_i)\right\} \leq \exp\left(\|\alpha\|^2 \sigma^2/2\right) \ \forall \alpha \in \mathbb{R}^p$ (see also Remark 4.3 in Section 4.3 for an asymptotic version of the statement).

The second of the considered methods is the weighted bootstrap. Here $X_1, \ldots, X_n \in \mathbb{R}^p$ are assumed to be zero mean, independent but not necessarily identically distributed. Introduce the following random variables

$$(1.2) \qquad \begin{array}{c} \varepsilon_1, \ldots, \varepsilon_n \in \mathbb{R}, \ \text{i.i.d., independent of } \{X_i\}_{i=1}^n, \\ \mathbf{E}\varepsilon_i = 0, \ \mathbf{E}(\varepsilon_i^2) = 1, \ \mathbf{E}(\varepsilon_i^3) = 1, \ \mathbf{E}(\varepsilon_i^4) < \infty. \end{array}$$

The weighted or the multiplier bootstrap approximation of S_n is:

$$(1.3) \qquad S_n^\circ := n^{-1/2}\sum_{i=1}^n X_i \varepsilon_i.$$

For this version of the bootstrap estimator we derive the following bound which holds with high probability:

$$(1.4) \qquad \sup_{B \in \mathscr{B}} \left| \mathbf{P}\left(S_n \in B\right) - \mathbf{P}^*\left(S_n^\circ \in B\right) \right| \leq C_{\sigma,z}(p^2/n)^{1/6}$$

for $p \leq C\sqrt{n}$, where constant $C_{\sigma,z}$ depends on value $\bar{C}_z$ introduced in (2.7) in Section 2, and on constants $\sigma_i^2 > 0$ which come from conditions on m.g.f.-s

of X_i for each $i = 1, \ldots, n$ (an asymptotic version of the statement is given in Remark 4.3, Section 4.3). Bounds (1.1) and (1.4) imply, in particular, that if the random vector X_1 is sub-Gaussian, and the ratio $p^{K/(K-2)}/n$ (or p^2/n for bound (1.4)) is rather small, then the bootstrap approximation is accurate. In addition, we give an example of $\{X_i\}_{i=1}^n$ which justifies that the condition $p = o(n)$ (or $p = o(n^{1/2})$ for the weighted bootstrap) as $n \to \infty$ is necessary for the consistency result $\sup_{B \in \mathscr{B}} \left| \mathbf{P}\left(S_n \in B\right) - \mathbf{P}^*\left(S_n^* \in B\right) \right| \xrightarrow{\mathbf{P}} 0$ (or $\sup_{B \in \mathscr{B}} \left| \mathbf{P}\left(S_n \in B\right) - \mathbf{P}^*\left(S_n^{\bullet} \in B\right) \right| \xrightarrow{\mathbf{P}} 0$ for the weighted bootstrap method) as $n \to \infty$.

An important feature of the present results is that they do not involve any asymptotic methods such as, for example, Edgeworth expansions that are frequently employed for studying the rates of convergence of bootstrap estimators. We develop a new non-asymptotic approach that allows to study higher-order accuracy of bootstrap in high-dimensional setting. The key element in the proofs of our theoretical results about bootstrapping is a multivariate Berry–Esseen inequality in a nonclassical form which might be interesting by itself.

We consider the problem of approximation of a probability distribution of the sum $S_n = n^{-1/2} \sum_{i=1}^n X_i$, where $X_i \in \mathbb{R}^p$ are independent random vectors such that $\mathbf{E} X_i = 0$ and $\mathbf{E}(\|X_i\|^K) < \infty$ for some $K \geq 3$. The approximating distribution corresponds to the sum $\tilde{S}_n := n^{-1/2} \sum_{i=1}^n Y_i$, where $Y_1, \ldots, Y_n \in \mathbb{R}^p$ are independent random vectors, independent of $\{X_i\}_{i=1}^n$ such that $\mathbf{E}(\|Y_i\|^K) < \infty$,

$$(1.5) \qquad \mathbf{E}(X_i^k) = \mathbf{E}(Y_i^k) \ \forall k = 1, \ldots, K-1,$$

and $Y_i = Z_i + U_i$ for some independent random vectors $Z_i, U_i \in \mathbb{R}^p$, where Z_i are normally distributed with $\mathbf{E} Z_i = 0$. Throughout the paper the condition $\mathbf{E}(X^k) = \mathbf{E}(Y^k) \ \forall k = 1, \ldots, K$ on the higher-order moments of random vectors $X = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$ and $Y = (y_1, \ldots, y_p)^\top \in \mathbb{R}^p$ denotes that for all degrees $k = 1, \ldots, K$ and for all indices $1 \leq i_1, \ldots, i_k \leq p$

$$(1.6) \qquad \mathbf{E}(x_{i_1} \ldots x_{i_k}) = \mathbf{E}(y_{i_1} \ldots y_{i_k}).$$

In Lemma 3.1 we show that if a cardinality of a support of X_i is sufficiently large, then the corresponding random vectors Z_i, U_i always exist. The probability distribution of such constructed random vector $\tilde{S}_n$ turns out to be a rather good approximation of a distribution of the initial sum S_n. One of the main results in the paper is the following uniform Berry–Esseen type

bound: for the set $\mathscr{B}$ of all Euclidean balls in $\mathbb{R}^p$ and for i.i.d. $\{X_i\}_{i=1}^n$

$$\sup_{B\in\mathscr{B}}\big|\mathbf{P}(S_n \in B) - \mathbf{P}(\tilde{S}_n \in B)\big|$$

(1.7) $$\leq C_{\Sigma,K}\big\{\mathbf{E}\left(\|X_1\|^K + \|Y_1\|^K\right)\big\}^{1/(K-2)}n^{-1/2},$$

where constant $C_{\Sigma,K}$ depends on K and on eigenvalues of $(\operatorname{Var} Z_1)^{-1}$. Bound (1.7) includes the classical Berry–Esseen inequality, where the approximating distribution is multivariate normal i.e. $Y_i \sim \mathcal{N}(0, \operatorname{Var} X_i)$ and $K = 3$. If $K > 3$, this bound exploits more information about coinciding moments, than the normal approximation does, which leads to a better accuracy.

Our proof of bound (1.7) is based on the work of Bentkus [4], where the author obtained a multivariate Berry–Esseen inequality involving the standard normal distribution, uniformly on the set of all Euclidean balls, and also on the set of all convex sets in $\mathbb{R}^p$. In this paper we extend the proof in the work of Bentkus [4] to the "quasi-normal" case, i.e. for the approximation with the sum $\tilde{S}_n$ of the convolutions $Y_i = Z_i + U_i$, where Z_i are normally distributed. This approach allows us to use both the properties of the normal distribution and the higher moments condition (1.5). Furthermore, if $\|X_1\|^2 \leq p$ a.s., then inequality (1.7) implies

$$\sup_{B\in\mathscr{B}}\big|\mathbf{P}(S_n \in B) - \mathbf{P}(\tilde{S}_n \in B)\big| \leq C_{\Sigma,K}\big\{p^{K/(K-2)}/n\big\}^{1/2}.$$

In Lemma 2.1 in Section 2, we show that for $K \geq 3$ the requirement $p = o(n^{(K-2)/K})$ as $n \to \infty$ is necessary for $\sup_{x\in\mathbb{R}} |\mathbf{P}(\|S_n\| \leq x) - \mathbf{P}(\|\tilde{S}_n\| \leq x)| \to 0$, $n \to \infty$ for some approximating distribution $\tilde{S}_n$, satisfying conditions of Theorem 2.1.

Now let us discuss how Berry–Esseen type bound (1.7) leads to the results (1.1), (1.4) about bootstrap. In the framework of the Efron's bootstrapping scheme, condition (1.5) is modified with concentration bounds for the higher-order bootstrap moments (equal to the empirical moments) $\mathbf{E}^*(X_i^{*k}) = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^k$ for $k = 2, \ldots, K-1$, where $\mathbf{E}^*(\cdot) := \mathbf{E}(\cdot \,|\, X_1, \ldots, X_n)$. For the case of the weighted bootstrap, condition (1.2) implies $\mathbf{E}^*(X_i^{\circ k}) = X_i^k \mathbf{E}(\varepsilon_i^k) = X_i^k$, $k = 2, 3$. In this way, the concentration properties of the empirical moments around the theoretical ones together with the higher-order Berry–Esseen bounds of the form (1.7) determine accuracy of the bootstrap procedures. Let us emphasize that the considered higher-order approximations play a key role for obtaining the improved accuracy of bootstrap procedures in terms of the ratio of p and n. For example, consider the weighted bootstrap procedure with a simplified condition on the random weights ε_i. If $\mathbf{E}\varepsilon_i = 0$ and $\mathbf{E}(\varepsilon_i^2) = 1$, then

(1.8) $$\mathbf{E}S_n = \mathbf{E}S_n^{\circ}, \quad \mathbf{E}(S_n S_n^\top) = \mathbf{E}(S_n^{\circ} S_n^{\circ\top}).$$

Using (1.8) and a normal approximation between probability distributions of $\|S_n\|$ and $\|S_n^{\bullet}\|$ (e.g. the results of Bentkus [4], or the inequalities by Spokoiny and Zhilova [35] for the non-i.i.d. case), one can obtain an approximation bound similar to (1.4), with an error term $\left(C_{\Sigma}^3 \mathbf{E}\left(\|X_1\|^3\right)/\sqrt{n}\right)^{1/4}$ which is less sharp than (1.4) in the ratio between p and n. Using also the condition $\mathbf{E}(\varepsilon_i^3) = 1$, we obtain

$$(1.9) \qquad \forall \alpha \in \mathbb{R}^p \ \mathbf{E}\{(\alpha^{\top} S_n)^3\} = \mathbf{E}\{(\alpha^{\top} S_n^{\bullet})^3\},$$

and this property leads to the improved error term in (1.4). In order to employ the information about the third moments, as in (1.9), one needs to use an approximation scheme that is more general than the normal approximation. For this purpose we establish the multivariate higher-order Berry–Esseen inequalities (Section 2).

The methods introduced in the paper allow to consider an important and a more general model, namely, the Smooth Function Model introduced by Bhattacharya and Ghosh [6] and Hall [16] (Chapter 2.4). In this model, the object of interest is $f(\mu)$, where $f : \mathbb{R}^p \mapsto \mathbb{R}$ is a smooth function and μ is an unknown expected value if X_i. The bootstrap estimators allow to approximate $f(\bar{X}) - f(\mu)$ in distribution, and, therefore, to establish a confidence set for $f(\mu)$. This also includes the case, when one aims at constructing a confidence set for μ in the form $f(\bar{X} - \mu)$. In Section 4 we establish the approximation bounds similar to (1.1) and (1.4) for the Smooth Function Model.

The weighted or the multiplier bootstrap procedure is useful in the situations, when it is required to resample a solution of estimating equations, or a maximum likelihood estimator, or in the case when the random summands $\{X_i\}_{i=1}^n$ are not necessarily identically distributed (see, e.g., Mammen [27], Chatterjee and Bose [9]). The present results for the weighted bootstrap lead also to an improvement of accuracy of a weighted bootstrap procedure for general log-likelihood ratio statistics under possible model misspecification. Spokoiny and Zhilova [35] considered the weighted bootstrap for estimation of quantiles of a log-likelihood ratio, they showed that if a parametric model is not severely misspecified, then the accuracy of bootstrap log-likelihood ratio quantiles corresponds to the accuracy of the normal approximation between statistics of the type $\|S_n\|$ and $\|S_n^{\bullet}\|$. Using inequality (1.4), we infer that the accuracy of the weighted bootstrap for log-likelihood ratio depends rather on accuracy of the Wilks-type bounds, than on the normal approximation. We employ this result for construction of likelihood-based confidence sets.

Below we give an overview of the existing literature about bootstrap accu-

racy. Resampling methods are widely used for statistical inference in various applications. The bootstrap is well-known for its good performance in the situations when the amount of data is small (see, e.g. Horowitz [19]), however, there are relatively few results about accuracy of the bootstrap in a non-asymptotic set-up. Most of the existing results are quite recent. Arlot et al. [2] studied generalized weighted bootstrap for construction of non-asymptotic confidence bounds in ℓ_r -norm ($r \in [1, \infty]$) for the mean value of high-dimensional random vectors with a symmetric and bounded (or with the normal) distribution. Chernozhukov et al. [10] established Gaussian approximation results, as well as accuracy of the multiplier and the Efron's bootstrap for maxima of sums of high-dimensional vectors in a very general set-up. Chernozhukov et al. [11] extended the results from maxima to general hyperrectangles and sparsely convex sets. The results of Chernozhukov et al. [10, 11] allow the dimension p grow as $O(\exp(Cn^c))$ for some constants $c, C > 0$. Spokoiny and Zhilova [35] considered the multiplier bootstrap for estimation of quantiles of a general log-likelihood ratio under model misspecification. Zhilova [40] extended this methodology for the simultaneous likelihood-based inference in the case of exponentially large number of models.

In the asymptotic high-dimensional setting when both the parameter dimension p and the sample size n are large, Bickel and Freedman [7], Mammen [25, 27] studied accuracy of the Efron's and the wild bootstrap for the linear regression model and for M-estimators; Chatterjee and Bose [9] studied generalized bootstrap for estimating equations also in high-dimensional asymptotic framework. Mammen [27] studied validity and higher-order accuracy of the wild bootstrap (or Wu's bootstrap, first proposed by Wu [38]) under the condition $\mathbf{E}(\varepsilon_i^3) = 1$ on the weights, in context of linear contrasts in high dimensional linear models and for bootstrapping F-tests. Liu [23] used the condition $\mathbf{E}(\varepsilon_i^3) = 1$ in order to obtain the second order accuracy of the wild bootstrap.

One of the basic ways of studying the properties of bootstrap procedures is to consider asymptotic approximations of distributions of an initial statistic and its bootstrap estimate, e.g. using central limit theorems or their refinements with Edgeworth expansions (see Præstgaard [29], Præstgaard and Wellner [30], Hall [16], Mammen [26], Barbe and Bertail [3], Shao and Tu [33], van der Vaart and Wellner [37], Janssen and Pauls [22], and references therein). Berry–Esseen type inequalities had been first used by Singh [34] and Liu [23] in the framework of bootstrap. Holmes and Reinert [17] established bootstrap consistency in various settings using Stein's method.

Below we discuss the literature about Berry–Esseen type bounds. The

problem of approximation of a probability distribution of the sum S_n belongs to the class of Central Limit Problems which has a long history of studies, see the paper by Loève [24] for a detailed overview. Ibragimov [21] studied convergence of a distribution of S_n in case of i.i.d. scalar summands, to the standard normal law, under the higher moments condition; the author obtained a higher-order accuracy using Edgeworth expansion. Zolotarev [43] introduced pseudomoments, which characterize closeness of moments of two distributions, for estimation of convergence rates in limit theorems; such limit theorems are called nonclassical. In the multivariate case, some of the first nonclassical results about normal approximation on closed convex sets had been obtained by Paulauskas [28], Rotar' [31] and Ul'yanov [36]. To the best of our knowledge, the problem of approximation of a probability distribution of S_n under the higher moments condition (1.5) and with an explicit dependence on the dimension p, had not been studied before.

Now let us summarize the contribution of this paper to the existing literature. In order to study the properties in a high-dimensional non-asymptotic setting, one needs to use new approaches and techniques. The methodology developed in this paper allows to consider higher-order properties of bootstrap methods in the modern set-up. To the best of our knowledge this had not been done in the earlier literature. The main theoretical tools, namely, the multivariate higher-order Berry–Esseen inequalities might be interesting by themselves. The approximation bounds established here allow to track the dependence of the error terms on the dimension, on the sample size, and on the moments of the considered distributions. We provide examples, showing that the obtained error rates cannot be improved under the considered conditions. In addition, we refined an accuracy of the weighted/multiplier bootstrap procedure for the general log-likelihood ratio statistics.

Structure of the paper. The results about accuracy of bootstrap rely on Berry-Essen type inequalities, for this reason we first present the latter results in Sections 2 and 3. Section 4 contains theoretical results about accuracy of the bootstrap. Sections A and B in the supplement [41] contain proofs of the statements from Sections 2 and 4 respectively. Section 5 presents results of numerical experiments.

Notation. $\|\cdot\|$ denotes the Euclidean norm for vectors and the operator norm for matrices or tensors; S_p^+ denotes the set of symmetric positive definite real-valued matrices of size $p \times p$; $\mathscr{B}$ is the set of all closed Euclidean balls in $\mathbb{R}^p$; $\mathbf{I}_p$ is the identity matrix of size $p \times p$; if X is a vector in $\mathbb{R}^p$, X^k stands for the tensor power $X^{\otimes k}$; for $f : \mathbb{R}^p \mapsto \mathbb{R}$ and $h \in \mathbb{R}^p$, $f^{(s)}(x)h^s$ denotes the higher-order directional derivative $(h^\top \nabla)^s f(x)$; C indicates a

positive generic constant unless specified otherwise.

2. Higher-order Berry–Esseen inequalities. Consider independent random vectors $X_1, \ldots, X_n \in \mathbb{R}^p$ such that $\forall i = 1, \ldots, n$ $\mathbf{E}X_i = 0$, $\mathrm{Var}(X_i) \in S_p^+$, $\mathbf{E}(\|X_i\|^K) < \infty$ for some integer $K \geq 3$. Let $Y_1, \ldots, Y_n \in \mathbb{R}^p$ be independent random vectors, and such that $\forall i = 1, \ldots, n$

$$(2.1) \qquad \begin{aligned} Y_i \text{ is independent of } X_1, \ldots, X_n, \ \mathbf{E}(\|Y_i\|^K) < \infty, \\ \mathbf{E}(X_i^k) = \mathbf{E}(Y_i^k) \ \forall k = 1, \ldots, K-1. \end{aligned}$$

A formal definition of the equality of the higher-order moments of vector-valued random variables (as in (2.1)) is given in (1.6). We assume also that $\forall i = 1, \ldots, n$

$$(2.2) \qquad \begin{aligned} \exists \text{ independent r.v. } Z_i, U_i \in \mathbb{R}^p \text{ s.t. } Y_i \overset{d}{=} Z_i + U_i, \\ \mathbf{E}Z_i = \mathbf{E}U_i = 0, \ Z_i \sim \mathcal{N}(0, \Sigma_{z,i}) \text{ for some } \Sigma_{z,i} \in S_p^+. \end{aligned}$$

Consider the following sums of mutually independent random vectors with zero mean:

$$(2.3) \qquad S_n := n^{-1/2} \sum_{i=1}^n X_i, \quad \tilde{S}_n := n^{-1/2} \sum_{i=1}^n Y_i.$$

We establish uniform approximation bounds between probability distributions of S_n and $\tilde{S}_n$ on the set $\mathscr{B}$ of all Euclidean balls in $\mathbb{R}^p$. Theorems 2.1 and 2.2 treat the cases when $\{X_i\}_{i=1}^n$ are i.i.d. and independent but not necessarily identically distributed vectors correspondingly. For the case of i.i.d. summands X_i (and, hence, i.i.d. Z_i) denote

$$(2.4) \qquad \Sigma_z := \Sigma_{z,i} = \mathrm{Var}(Z_i), \quad C_z := \|\Sigma_z^{-1/2}\|.$$

THEOREM 2.1. *Consider the random vectors $\{X_i\}_{i=1}^n$ introduced above, suppose that they are i.i.d., and that there exist i.i.d. approximating random vectors $\{Y_i\}_{i=1}^n$ meeting conditions (2.1) and (2.2). It holds for the sums S_n and $\tilde{S}_n$ defined in (2.3)*

$$\sup_{B \in \mathscr{B}} \left| \mathbf{P}(S_n \in B) - \mathbf{P}(\tilde{S}_n \in B) \right| \leq C_{\mathscr{B},iid} \frac{\left\{ C_z^K \mathbf{E}\left(\|X_1\|^K + \|Y_1\|^K\right) \right\}^{1/(K-2)}}{n^{1/2}},$$

where constant $C_{\mathscr{B},iid} > 0$ depends only on K; a detailed definition of $C_{\mathscr{B},iid}$ is given in the proof (see (A.52) in Section A.2 of the supplement [41]).

REMARK 2.1 (The case of the normal approximation). If the approximating random vectors Y_i are normally distributed, then $U_i \equiv 0$, $Y_i \sim \mathcal{N}(0, \mathrm{Var}(X_i))$, $\Sigma_z = \mathrm{Var}(X_i)$, and $C_z = \|\{\mathrm{Var}(X_i)\}^{-1/2}\|$. Furthermore, if $K = 3$ and Y_i are standard normal, then the bound in Theorem 2.1 is similar to the classical multivariate Berry–Esseen inequality by Bentkus [4]. If $K > 3$ and Y_i are normally distributed, the term $\|X_1\|$ enters the bound above with a better power, than in the classical case where $K = 3$. In this way, Theorem 2.1 extends the classical normal approximation result.

REMARK 2.2 (Dependence on C_z). The approximation bound in Theorem 2.1 depends on $C_z = \|\Sigma_z^{-1/2}\|$, where Σ_z is a covariance matrix of the normal component Z_i of the approximating distribution Y_i. In Lemma 3.1 (Section 3) we show that if a cardinality of a support of X_i is sufficiently large, then there exist random vectors U_i such that Σ_z is positive definite. Therefore, it holds $\lambda_\Sigma^{-1} \le C_z < \infty$, where λ_Σ^2 is the smallest eigenvalue of $\mathrm{Var}\, X_i$. In Lemma 3.2 we consider the case when the number of coinciding moments between X_i and Y_i is $K - 1 = 3$; we show that for any $c_0 \in (0, \lambda_\Sigma)$, there exists distribution $Y_i = Z_i + U_i$ such that $C_z < c_0^{-1}$. Hence C_z can be taken as a generic constant for $K = 4$. Moreover, if the coordinates of the vector X_i are mutually independent, then the problem of characterizing Σ_z and C_z becomes one-dimensional and, therefore, C_z does not depend on p in this case.

REMARK 2.3 (Accuracy of the approximation). The error term in Theorem 2.1 is proportional to $\{pn^{-(K-2)/K}\}^{1/(K-2)}$ if $\|X_1\|, \|Y_1\| \le \sqrt{p}$ a.s. In Lemma 2.1 below, we show that for $K \ge 3$ condition $p = o(n^{(K-2)/K})$ as $n \to \infty$ is necessary for $\sup_{x \in \mathbb{R}} |\mathbf{P}(\|S_n\| \le x) - \mathbf{P}(\|\tilde{S}_n\| \le x)| \to 0$, $n \to \infty$ under the conditions of Theorem 2.1.

LEMMA 2.1 (Necessity of the condition $p = o(n^{(K-2)/K})$). *Let* $\{X_i\}_{i=1}^n$ *be random vectors as in Theorem 2.1. Suppose that* $\mathbf{E}(\|X_i\|^{K+2}) < \infty$ *for an integer* $K > 3$. *There exist random vectors* $\{Y_i\}_{i=1}^n$ *satisfying conditions of Theorem 2.1, and s.t. that the condition* $p = o(n^{(K-2)/K})$ *for* $n \to \infty$ *is necessary for* $\sup_{x \in \mathbb{R}} |\mathbf{P}(\|S_n\| \le x) - \mathbf{P}(\|\tilde{S}_n\| \le x)| \to 0$ *as* $n \to \infty$.

REMARK 2.4. In the recent paper Zhai [39] considers a multivariate CLT in $\mathcal{W}_2$-distance. The author shows that if $X_1, \ldots, X_n \in \mathbb{R}^p$ are i.i.d. with mean zero and such that $\|X_i\| \le \beta$ a.s. for some constant $\beta > 0$, then

$$(2.5) \qquad \mathcal{W}_2(S_n, Z) \le 5\sqrt{p}\beta(1 + \log n)/\sqrt{n},$$

where $Z \sim \mathcal{N}(0, \mathrm{Var}\, X_1)$, and $\mathcal{W}_2$ is the 2-Wasserstein distance. This result implies, that if $\beta \leq C\sqrt{p}$, then $\mathcal{W}_2(S_n, Z) \leq Cp(1+\log n)/\sqrt{n}$. In Lemma 2.2 below, we consider a uniform bound on $\sup_{B \in \mathscr{B}} |\mathbf{P}(S_n \in B) - \mathbf{P}(Z \in B)|$, using the result (2.5). It turns out that under conditions (2.1), (2.2) for $K \geq 4$, the higher-order Berry–Esseen type inequality in Theorem 2.1 yields a better accuracy w.r.t. the sample size n, and w.r.t. the ratio between p and n. Indeed, inequality (2.6) below, which follows from the results of Zhai [39], has an error term of order $(p^2/n)^{1/3}$ (up to $\log n$). Whereas Theorem 2.1 (for the case $Y_i \sim \mathcal{N}(0, \mathrm{Var}\, X_i)$) provides a smaller error term of order $(p^{K/(K-2)}/n)^{1/2}$ for $K \geq 4$.

LEMMA 2.2. *Let* $X_1, \ldots, X_n \in \mathbb{R}^p$ *be i.i.d. random vectors, such that* $\mathrm{Var}\, X_i = \mathbf{I}_p$ *and* $\|X_i\| \leq \beta$ *a.s. for some constant* $\beta > 0$. *Let* $Z \sim \mathcal{N}(0, \mathbf{I}_p)$. *The results of Zhai [39] imply*

$$\sup_{B \in \mathscr{B}} |\mathbf{P}(S_n \in B) - \mathbf{P}(Z \in B)| \leq C_1 p^{1/3} \beta^{2/3} (1 + \log n)/n^{1/3}$$
$$(2.6) \qquad\qquad\qquad\qquad \leq C_2 p^{2/3}(1 + \log n)/n^{1/3},$$

where in the latter inequality one takes $\beta \leq c\sqrt{p}$. *Here* c, C_1, *and* C_2 *denote positive generic constants.*

Now let us consider the case when the random summands X_i are independent but not necessarily identically distributed (non-i.i.d.). Denote

$$(2.7) \qquad\qquad \bar{\Sigma}_z := n^{-1}\sum_{i=1}^{n}\Sigma_{z,i}, \quad \bar{C}_z := \|\bar{\Sigma}_z^{-1/2}\|.$$

THEOREM 2.2. *Consider random vectors* $\{X_i\}_{i=1}^{n}$ *introduced above, suppose that they are independent but not necessarily identically distributed, and that there exist independent approximating vectors* $\{Y_i\}_{i=1}^{n}$ *meeting conditions* (2.1) *and* (2.2). *It holds for the sums* S_n *and* $\tilde{S}_n$ *defined in* (2.3)

$$\sup_{B \in \mathscr{B}} \left|\mathbf{P}\left(S_n \in B\right) - \mathbf{P}\left(\tilde{S}_n \in B\right)\right| \leq C_{\mathscr{B},ind}\left\{\bar{C}_z^{K}\sum_{i=1}^{n}\mathbf{E}\left(\|X_i\|^{K} + \|Y_i\|^{K}\right)n^{-K/2}\right\}^{\frac{1}{K+1}},$$

where constant $C_{\mathscr{B},ind} > 0$ *depends only on* K, *it is defined in the proof (see* (A.57) *in Section A.3 of the supplement [41]).*

REMARK 2.5. The proof of Theorem 2.1 largely exploits the assumption that the summands $\{X_i\}_{i=1}^{n}$ are identically distributed, and it does not directly apply to the non-i.i.d. case. This causes a difference between the error terms in Theorems 2.2 and 2.1, however, the critical ratio of p and n (namely, $p^{K/(K-2)}/n$) in the error terms remains the same in both results. We leave an improvement of the power $1/(K+1)$ in the non-i.i.d. case for the future work.

REMARK 2.6.   The proofs of the Berry–Esseen type inequalities (Theorems 2.1 and 2.2) exploit the following properties of the set $\mathscr{B}$ (cf. Bentkus [4, 5] and Chernozhukov et al. [10, 11]):

- Invariance under rescaling and under taking shifts, i.e. if $B \in \mathscr{B}$, then $\forall x \in \mathbb{R}^p$ and $\forall a \in \mathbb{R}, a > 0$ it holds $aB + x \in \mathscr{B}$.
- Invariance under taking ε-neighborhood w.r.t. the ℓ_2-norm: consider an arbitrary $B \in \mathscr{B}$, $B = B_r(x_0) \coloneqq \{x \in \mathbb{R}^p : \|x - x_0\| \leq r\}$, then the ε-neighborhood of B reads as $B^\varepsilon = B_{r+\varepsilon}(x_0)$, if $r + \varepsilon \geq 0$, and $B^\varepsilon = \emptyset$ otherwise. Therefore $B^\varepsilon \in \mathscr{B}$.

The same properties hold for the set $\mathscr{H}$ of all half-spaces in $\mathbb{R}^p$. In proposition A.1 (Section A.3 of the supplement [41]), we consider a higher-order Berry–Esseen approximation for S_n uniformly over the set $\mathscr{H}$, similarly to Theorems 2.1 and 2.2.

Corollary 2.1 below follows directly from the previous theorems and the triangle inequality. It justifies a higher-order accuracy of approximation between two probability distributions with matching moments.

COROLLARY 2.1.   *Consider random vectors* $\{X_i\}_{i=1}^n$ *introduced above, and suppose that there exist independent approximating vectors* $\{Y_i\}_{i=1}^n$ *meeting conditions* (2.1) *and* (2.2). *Consider also independent random vectors* $X_1', \ldots, X_n' \in \mathbb{R}^p$, *that are independent of* $\{X_i\}_{i=1}^n$, $\{Y_i\}_{i=1}^n$, *and such that* $\forall i = 1, \ldots, n \; \mathbf{E}\left(\|X_i'\|^K\right) < \infty$, $\mathbf{E}\left(X_i^k\right) = \mathbf{E}\left(X_i'^k\right) \; \forall k = 1, \ldots, K-1$. *Let also* $S_n' \coloneqq n^{-1/2} \sum_{i=1}^n X_i'$, *and* $\Delta_n' \coloneqq \sup_{B \in \mathscr{B}} \left| \mathbf{P}\left(S_n \in B\right) - \mathbf{P}\left(S_n' \in B\right) \right|$.

1. *If conditions of Theorem* 2.1 *are fulfilled and* $\{X_i'\}_{i=1}^n$ *are i.i.d. Let also* $L_K \coloneqq \mathbf{E}\left(\|X_1\|^K + \|Y_1\|^K\right)$ *and* $L_K' \coloneqq \mathbf{E}\left(\|X_1'\|^K + \|Y_1\|^K\right)$, *then*

$$\Delta_n' \leq C_{\mathscr{B},iid} C_z^{K/(K-2)} \big\{ L_K^{1/(K-2)} + L_K'^{1/(K-2)} \big\} n^{-1/2}.$$

2. *If conditions of Theorem* 2.2 *are fulfilled and* $\{X_i'\}_{i=1}^n$ *are not necessarily identically distributed. Let also* $\bar{L}_K \coloneqq n^{-1} \sum_{i=1}^n \mathbf{E}(\|X_i\|^K + \|Y_i\|^K)$ *and* $\bar{L}_K' \coloneqq n^{-1} \sum_{i=1}^n \mathbf{E}(\|X_i'\|^K + \|Y_i\|^K)$, *then*

$$\Delta_n' \leq C_{\mathscr{B},ind} \bar{C}_z^{K/(K+1)} \big\{ \bar{L}_K^{1/(K+1)} + \bar{L}_K'^{1/(K+1)} \big\} n^{-0.5(K-2)/(K+1)}.$$

3. Properties of the approximating distribution.

LEMMA 3.1 (Existence of the approximating distribution).   *Let a random vector* X *be supported in a closed set* $A \subseteq \mathbb{R}^p$, *and let* X *be such that* $\mathbf{E} X = 0$, $\mathrm{Var}\, X \in S_p^+$, $\mathbf{E}(\|X\|^{K+2}) < \infty$ *for some integer* $K \geq 2$. *If* X *is*

continuously distributed, there exists a random vector $Y := Z + U$, such that $\mathbf{E}(\|Y\|^{K+2}) < \infty$, where $Z, U \in \mathbb{R}^p$ are independent, $Z \sim \mathcal{N}(0, \Sigma_z)$ for some $\Sigma_z \in S_p^+$, and $\mathbf{E}(X^k) = \mathbf{E}(Y^k)$ for all $k = 0, \ldots, K$. Furthermore, if X is a sub-Gaussian random vector, then there exists a sub-Gaussian approximating distribution Y satisfying the above conditions. If X has a discrete probability distribution supported on M points in $\mathbb{R}^p$ such that each coordinate of X is supported on at least m points in $\mathbb{R}$, then the lemma's statement holds for X when $M \geq 1 + (K + 2)m^{p-1}$.

PROOF OF LEMMA 3.1. Denote $m_k := \mathbf{E}(X^k)$, and $u_k := \mathbf{E}(U^k)$ for $k = 0, 1, 2, \ldots$ Conditioning on U leads to $\mathscr{L}(Y \mid U) = \mathcal{N}(U, \Sigma_z)$ and to the following system of linear equations:

$$m_0 = \mathbf{E}(Z + U)^0 = u_0, \quad m_2 = \mathbf{E}(Z + U)^2 = u_2 + \Sigma_z,$$

$$m_1 = \mathbf{E}(Z + U) = u_1, \quad m_3 = \mathbf{E}(Z + U)^3 = u_3,$$

$$m_K = \mathbf{E}(Z + U)^K = K! \sum_{l=0}^{[K/2]} S_{p\mathbb{1}_K} u_{K-2l} \otimes \mathrm{vec}(\Sigma_z)^l \{l!(K - 2l)!2^l\}^{-1},$$

where $S_{p\mathbb{1}_K}$ is the symmetrizer operator acting on the K-th tensor power of $\mathbb{R}^p$; this formula for the raw moments of the multivariate normal distribution is given in the work Holmquist [18]. The solution $\{u_k(\Sigma_z)\}_{k=0}^K$ of this system depends on Σ_z continuously. Moreover,

(3.1)   if $\Sigma_z = 0$, then $u_k(\Sigma_z) = m_k \quad \forall k = 0, \ldots, K$.

In order to prove the lemma's statement, it is sufficient to show that there exists $\Sigma_z \in S_p^+$, s.t. the solution $\{u_k(\Sigma_z)\}_{k=0}^K$ also solves the following multivariate truncated moment problem:

given a p-dimensional real multisequence $\{u_k(\Sigma_z)\}_{k=0}^K$, does there exist a positive Borel measure μ s.t. $\mathrm{supp}\,\mu \subseteq A$ and $\int_{\mathbb{R}^p} x^k d\mu(x) = u_k(\Sigma_z) \, \forall k = 0, \ldots, K$?

The work of Curto and Fialkow [13] provides necessary and sufficient conditions for solvability of multivariate truncated moment problems. Before stating these conditions we introduce some notation. Let $\mathcal{P}_K$ denote the space of polynomials: $\mathbb{R}^p \mapsto \mathbb{R}$, of degree $\leq K$, and with real coefficients. A polynomial $p = p(x) = \sum_{|\boldsymbol{i}| \leq K} a_{\boldsymbol{i}} x^{\boldsymbol{i}} \in \mathcal{P}_K$ is positive (or strictly positive) on A, if $p(x) \geq 0$ (or $p(x) > 0$) for all $x \in A$. Here $\boldsymbol{i} := (i_1, \ldots, i_p) \in \mathbb{N}_0^p$ denotes multi-index, $|\boldsymbol{i}| = \sum_{j=1}^p i_j$, and $x^{\boldsymbol{i}} := x_1^{i_1} \ldots x_p^{i_p}$. For a multisequence $\{u_{\boldsymbol{i}}\}_{|\boldsymbol{i}| \leq K}$ the Riesz functional $L : \mathcal{P}_K \mapsto \mathbb{R}$ is defined as $L(\sum_{|\boldsymbol{i}| \leq K} a_{\boldsymbol{i}} x^{\boldsymbol{i}}) :=$

$\sum_{|i| \leq K} a_i u_i$. If the truncated moment problem is soluble, we can write

(3.2) $$L(p) = \sum_{|i| \leq K} a_i u_i = \int_{\mathbb{R}^p} p(x) d\mu(x).$$

Curto and Fialkow [13] showed that a multisequence $\{u_i\}_{|i| \leq K}$ solves the multivariate truncated moment problem on the set A iff there exists an extension $\{\tilde{u}_i\}_{|i| \leq K+2}$ of $\{u_i\}_{|i| \leq K}$ (i.e. $\tilde{u}_i = u_i$ for all $i : |i| \leq K$), such that for the corresponding Riesz functional $\tilde{L}(\sum_{|i| \leq K+2} a_i x^i) := \sum_{|i| \leq K+2} a_i \tilde{u}_i$ it holds:

(3.3)   if $p \in \mathcal{P}_{K+2}$ and p is positive on A, then $\tilde{L}(p) \geq 0$.

Firstly, consider the case of a continuously distributed X. Due to the definitions of m_k and u_k, and by the theorem of Curto and Fialkow [13] there exists an extension $\{m_k\}_{k=0}^{K+2}$, s.t. its corresponding Riesz functional $\tilde{L}_m(\sum_{|i| \leq K+2} a_i x^i) := \sum a_i m_i$ satisfies (3.3). Moreover, if $p \in \mathcal{P}_{K+2}$ is s.t. $p(x) > 0 \; \forall x \in A$, then $\tilde{L}(p) > 0$; if $\tilde{L}(p) = 0$ for some $p \in \mathcal{P}_{K+2}$ nonnegative on A, then $\mathbf{P}(\forall x \in A \; p(x) = 0) = 1$. The extension $\{m_k\}_{k=0}^{K+2}$ leads to the extended sequence $\{\tilde{u}_k(\Sigma_z)\}_{k=0}^{K+2}$. Property (3.1), continuity of the solutions $\{\tilde{u}_k(\Sigma_z)\}_{k=0}^{K+2}$ w.r.t. Σ_z, and (3.2) imply that there exists some $\Sigma_z \in S_p^+$ s.t. the corresponding Riesz functional $\tilde{L}_u(p) := \sum_{|i| \leq K+2} a_i \tilde{u}_i(\Sigma_z) > 0$ for all $p = \sum_{|i| \leq K+2} a_i x^i \in \mathcal{P}_{K+2}$ such that $p > 0$ on A (here we use that S_p^+ is a dense subset of the set of symmetric positive semidefinite matrices in $\mathbb{R}^{p \times p}$; see e.g. Chapter 2.4 in Boyd and Vandenberghe [8]). This finalizes the proof for the continuous case.

Now let X have a discrete probability distribution. Let m denote the minimal cardinality of the supports of X's coordinates. Consider a polynomial $q \in \mathcal{P}_{K+2}$ of degree $K + 2$, such that $q \not\equiv 0$ on the support A. According to the Schwartz-Zippel Lemma by Schwartz [32] and Zippel [42], the number of zeros of q is $\leq (K + 2)m^{p-1}$. Hence, if the set A contains at least $1 + (K+2)m^{p-1}$ points, then for any non-zero and positive polynomial q, $\tilde{L}(q) > 0$, where $\tilde{L}$ is the Riesz functional, considered in the previous paragraph. This allows to apply here the arguments for the continuous case.

If X is sub-Gaussian, then $\left| m_k \gamma^k \right|^{1/k} \leq \left\{ \mathbf{E} | \gamma^\top X |^k \right\}^{1/k} \leq C\sqrt{k} \; \forall k \geq 1$, $\forall \gamma \in \mathbb{R}^p : \|\gamma\| = 1$. Since the solutions $u_k(\Sigma_z)$ are linearly dependent on $\{m_k\}$, the sub-Gaussian property holds for the moments $u_k(\Sigma_z)$ as well. $\square$

LEMMA 3.2 (Choice of C_z for 3 coinciding moments). *Let $X \in \mathbb{R}^p$ be a random vector satisfying conditions of Lemma 3.1 for $K = 3$. Let λ_Σ denote the smallest eigenvalue of $\Sigma := \mathrm{Var}(X)$. Then for any $c_0 \in (0, \lambda_\Sigma)$ there exist independent random vectors $Z, U \in \mathbb{R}^p$, such that for $Y = Z + U$, it*

holds $\mathbf{E}(\|Y\|^4) < \infty$, $\mathbf{E}(X^k) = \mathbf{E}(Y^k)$ *for* $k = 1, 2, 3$, *and* $Z \sim \mathcal{N}(0, \Sigma_z)$, *where* Σ_z *is some symmetric matrix with the smallest eigenvalue* $> c_0$.

PROOF OF LEMMA 3.2. Consider random variable $\varepsilon \in \mathbb{R}$ independent of X and s.t. $\mathbf{E}\varepsilon = 0, \mathbf{E}(\varepsilon^2) = \alpha, \mathbf{E}(\varepsilon^3) = 1$ for some $\alpha \in (0, 1)$. Such distribution exists by the criterion for solubility of the truncated Hamburger moment problem (see Curto and Fialkow [12]). Indeed, if the Hankel matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right)$ has a positive Hankel extension $\left(\begin{smallmatrix} 1 & 0 & \alpha \\ 0 & \alpha & 1 \\ \alpha & 1 & \beta \end{smallmatrix}\right)$ for some $\beta = \mathbf{E}(\varepsilon^4)$, then there exists such random variable ε.

Take $\tilde{X} := X\varepsilon$, then $\mathbf{E}\tilde{X} = 0, \mathbf{E}(\tilde{X}^2) = \alpha\Sigma, \mathbf{E}(\tilde{X}^3) = \mathbf{E}(X^3)$. By Lemma 3.1 $\exists$ r.v. $\tilde{Y} = \tilde{Z} + \tilde{U} \in \mathbb{R}^p$ s.t. $\tilde{Z}$ and $\tilde{U}$ are independent of each other, $\tilde{Z} \sim \mathcal{N}(0, \tilde{\Sigma}_z)$ for some $\tilde{\Sigma}_z \in S_p^+$, and $\mathbf{E}(\tilde{X}^k) = \mathbf{E}(\tilde{Y}^k)$ $\forall k = 0, 1, 2, 3$. Let $\tilde{Z}_1 \sim \mathcal{N}(0, (1-\alpha)\Sigma)$ be independent of all random vectors considered in the proof, take $Y := \tilde{Y} + \tilde{Z}_1 = \tilde{Z} + \tilde{Z}_1 + \tilde{U}$, then it holds $\mathbf{E}(Y) = 0, \mathbf{E}(Y^2) = \alpha\Sigma + (1-\alpha)\Sigma = \Sigma$, and $\mathbf{E}(Y^3) = \mathbf{E}(\tilde{Y}^3) = \mathbf{E}(X^3)$. The normal part of Y is $Z := \tilde{Z} + \tilde{Z}_1 \sim \mathcal{N}(0, (1-\alpha)\Sigma + \tilde{\Sigma}_z)$. Let λ_z denote the smallest eigenvalue of $\mathrm{Var}\, Z$, then $(1-\alpha)\lambda_\Sigma < \lambda_z < \lambda_\Sigma$, where $\lambda_\Sigma > 0$ is the smallest eigenvalue of Σ. Hence, taking $\alpha = 1 - c_0/\lambda_\Sigma$, we obtain the lemma's statement. $\square$

4. Validity and accuracy of the bootstrap procedures. Here we study accuracy of the Efron's and the weighted bootstrap procedures in various settings. We begin with the Efron's bootstrap in Section 4.1; Sections 4.2, 4.4 present the results for the weighted bootstrap.

4.1. *Efron's bootstrap.* Let $X_1, \ldots, X_n$ be i.i.d. random vectors with $\Sigma := \mathrm{Var}(X_i) \in S_p^+$, let also X_i be sub-Gaussian, i.e. it holds for some $\sigma^2 > 0$ and for all $\alpha \in \mathbb{R}^p$

(4.1) $$\mathbf{E}\big\{\exp(\alpha^\top X_i)\big\} \leq \exp\big(\|\alpha\|^2 \sigma^2/2\big).$$

Assume that there exist i.i.d. random vectors $Y_1, \ldots, Y_n$ satisfying (2.1), (2.2) for some integer $K \geq 3$. Introduce resampled variables $X_1^*, \ldots, X_n^*$ with zero mean, according to the Efron's bootstrap methodology (Efron [14], Efron and Tibshirani [15]): $\mathbf{P}^*(X_i^* = X_j - \bar{X}) = 1/n$ $\forall i, j = 1, \ldots, n$, where $\bar{X} = n^{-1}\sum_{i=1}^n X_i$, and $\mathbf{P}^*(\cdot) = \mathbf{P}(\cdot \mid X_1, \ldots, X_n)$. In this way, $\{X_j^*\}_{j=1}^n$ are i.i.d., $\mathbf{E}^*(X_j^*) = 0$, and $\mathbf{E}^*(X_j^{*k}) = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^k$, for $k \geq 1$. The bootstrap approximation of the sum S_n is $S_n^* := n^{-1/2}\sum_{i=1}^n X_i^*$. Denote $C_X := \|\Sigma^{-1/2}\|$. By this definition, $C_X < C_z$. Assume also that a p.d.f. of X is bounded with a constant $c_f > 0$. In the statements in Section 4, including the theorems below, we use notation from the previous Section 2, e.g. constant $C_{\mathscr{B},iid}$. Let also $\tilde{C}_{x,k} := \big(1 + 2\sqrt{x/p} + 2x/p\big)^{k/2}$.

THEOREM 4.1 (Accuracy of the bootstrap for S_n on the set $\mathscr{B}$). *Suppose that the above conditions are fulfilled, then the following uniform bound holds on the set $\mathscr{B}$ of all Euclidean balls with probability $\geq 1 - 6e^{-x}$ for $x > 0$:*

$$\sup_{B \in \mathscr{B}} \left| \mathbf{P}\left(S_n \in B\right) - \mathbf{P}^*\left(S_n^* \in B\right) \right|$$
$$\leq \Delta^*_{\mathscr{B},iid} := C_{\mathscr{B},iid} \left\{ C_z^K \mathbf{E}\left(\|X_1\|^K + \|Y_1\|^K \right) \right\}^{1/(K-2)} n^{-1/2}$$
$$+ C_{\mathscr{B},iid} (2\tilde{C}_{x,K})^{1/(K-2)} (C_z \sigma p^{1/2})^{K/(K-2)} n^{-1/2} + R_{n,K},$$

where $R_{n,K} \leq C\sqrt{p/n}\, C_z^2 \sigma^{K+1} C_X^{K-1} C_{x,K}$, and $C_{x,K}$ is defined in (B.27); a detailed definition of $R_{n,K}$ is given in (B.9), (B.10) (Section B.1 in the supplement [41]).

REMARK 4.1.   The error term $\Delta^*_{\mathscr{B},iid}$ in the above result consists of two parts: one part corresponds to the higher-order Berry–Esseen type inequalities, another part $R_{n,K}$ comes from concentration bounds for higher-order empirical moments of X_i. If the ratio $p^{K/(K-2)}/n$ is small, then the first part is small as well. Furthermore, $\forall K \geq 3\ p^{K/(K-2)}/n \geq p/n$. In Lemma 4.1 (in Section 4.3), we consider an example where the condition $p/n = o(1)$ for $n \to \infty$ is required for the bootstrap consistency.

In the following theorem we study accuracy of the Efron's bootstrap procedure for the Smooth Function Model introduced by Bhattacharya and Ghosh [6] and Hall [16] (Chapter 2.4). In this model the object of interest is $f(\mu)$, where $f : \mathbb{R}^p \mapsto \mathbb{R}$ is a smooth function and μ is an unknown expected value if X_i. The bootstrap estimators allow to approximate $f(\bar{X}) - f(\mu)$ in distribution, and, therefore, to establish a confidence set for $f(\mu)$. This also includes the case, when we aim at constructing a confidence set for μ in the form $f(\bar{X} - \mu)$. Consider i.i.d. $X_1, \ldots, X_n \in \mathbb{R}^p$ with mean μ and sub-Gaussian tail behavior, i.e. condition (4.1) holds for $X_i - \mu$. Let $f : \mathbb{R}^p \mapsto \mathbb{R}$ be at least twice continuously differentiable function, s.t. $\forall h \in \mathbb{R}^p\ \sup_{x \in \mathbb{R}^p} |f^{(2)}(x)h^2| \leq C_{f,2}\|h\|^2$ for some constant $C_{f,2} > 0$. Assume also that $f'(\mu) \neq 0$, and $\|f'(\mu)\| > C_{f,l}\sqrt{p}$ for some constant $C_{f,l} > 0$. Denote the resampled i.i.d. data $X_1^*, \ldots, X_n^*$ and the bootstrap empirical mean as follows:

$$\mathbf{P}^*(X_i^* = X_j) = 1/n \quad \forall i,j = 1, \ldots, n, \text{ and } \bar{X}^* := n^{-1}\sum_{i=1}^n X_i^*.$$

Theorem 4.2 shows that the c.d.f. of $f(\bar{X}) - f(\mu)$ is uniformly well approximated by the c.d.f. of $f(\bar{X}^*) - f(\bar{X})$ conditioned on $\{X_i\}_{i=1}^n$.

THEOREM 4.2 (Accuracy of the bootstrap for the Smooth Function Model). *Let the above assumptions and conditions of Theorem 4.1 be fulfilled. It holds with probability $\geq 1 - 6e^{-x}$ for $x > 0$:*

$$\sup_{t \in \mathbb{R}} \left| \mathbf{P}\left(f(\bar{X}) - f(\mu) \leq t\right) - \mathbf{P}^*\left(f(\bar{X}^*) - f(\bar{X}) \leq t\right) \right|$$
$$\leq \Delta^*_{f,iid} := 2C_{f,2}C_z\sigma^2\tilde{C}_{x,2}C_{f,l}^{-1}(p/n)^{1/2} + C_{\mathscr{B},iid}C_z^{K/(K-2)}C_{M,K}n^{-1/2}$$
$$+ C_{\mathscr{B},iid}(2\tilde{C}_{x,K})^{1/(K-2)}\left\{C_z\sigma\right\}^{K/(K-2)}n^{-1/2} + R_{1,n,K},$$

where $C_{M,K} := \left[\|\mathbf{E}(X_1 - \mu)^K\| + \|\mathbf{E}(Y_1 - \mu)^K\|\right]^{1/(K-2)}$, the term $\tilde{C}_{x,K}$ is described in the previous statement, and $R_{1,n,K} \leq Cn^{-1/2}C_z^2\sigma^{K+1}C_X^{K-1}C_{x,K}$ is defined in (B.14) and (B.14) (Section B.1 in the supplement [41]).

COROLLARY 4.1. *Consider the following upper quantile functions of the bootstrap approximations: $Q_2^*(\alpha) := \inf\left\{t \in \mathbb{R} : \mathbf{P}^*\left(\|S_n^*\| > t\right) \leq \alpha\right\}$, $Q_f^*(\alpha) := \inf\left\{t \in \mathbb{R} : \mathbf{P}^*\left(f(\bar{X}^*) - f(\bar{X}) > t\right) \leq \alpha\right\}$ for $\alpha \in (0,1)$. Let $x > 0$. Theorems 4.1 and 4.2 imply the following two bounds:*

$$\left| \mathbf{P}\left(\|S_n\| > Q_2^*(\alpha)\right) - \alpha \right| \leq 2\Delta^*_{\mathscr{B},iid} + 6e^{-x},$$
$$\left| \mathbf{P}\left(f(\bar{X}) - f(\mu) > Q_f^*(\alpha)\right) - \alpha \right| \leq 2\Delta^*_{f,iid} + 6e^{-x}.$$

4.2. *Weighted bootstrap.* Let $X_1, \ldots, X_n$ be independent random vectors with $\Sigma_i := \mathrm{Var}(X_i) \in S_p^+$, let also $\{X_i\}_{i=1}^n$ be sub-Gaussian, i.e. it holds for some $\sigma_i^2 > 0$, $\forall \alpha \in \mathbb{R}^p$, and $\forall i = 1, \ldots, n$ $\mathbf{E}\left\{\exp(\alpha^\top X_i)\right\} \leq \exp\left(\|\alpha\|^2\sigma_i^2/2\right)$. Denote $\bar{\sigma}_k^2 := n^{-1}\sum_{i=1}^n \sigma_i^{2k}$. Assume that there exist i.i.d. random vectors $Y_1, \ldots, Y_n$ satisfying (2.1) and (2.2) for $K = 4$. Assume also that p.d.f.-s of X_i are bounded with a constant $c_f > 0$. The bootstrap random weights $\varepsilon_1, \ldots, \varepsilon_n$, are taken as in (1.2). These are some examples of such random weights (here $z_i \sim \mathcal{N}(0,1)$, independent of e_i, c_i, b_i): $(1 - 2^{-2/3})^{1/2}z_i + 2^{-1/3}(e_i - 1)$ for $e_i \sim exp(1)$; $2^{-1/2}z_i + 2^{-1}(c_i - 1)$ for $c_i \sim \chi_1^2$; $(1 - 3^{-2/3})^{1/2}z_i + 3^{-1/3}2(b_i - 0.5)$ for $b_i \sim Bernoulli(0.5)$. More examples of the bootstrap weights satisfying (1.2) can be found in the works of Liu [23] and Mammen [27].

The weighted bootstrap approximation of the sum S_n is $S_n^\circ := n^{-1/2}\sum_{i=1}^n X_i\varepsilon_i$. The probability distribution of S_n° is taken conditioned on $\{X_i\}_{i=1}^n$.

THEOREM 4.3 (Accuracy of the weighted bootstrap for S_n on the set $\mathscr{B}$). *Let the above conditions be fulfilled, then it holds with probability $\geq 1 - 6e^{-x}$ for $x > 0$:*

$$\sup_{B \in \mathscr{B}} \left| \mathbf{P}\left(S_n \in B\right) - \mathbf{P}^*\left(S_n^\circ \in B\right) \right|$$
$$\leq \Delta^\circ_{\mathscr{B},w,ind} := C_{\mathscr{B},w,ind}\left\{\bar{C}_z^4\sum_{i=1}^n\mathbf{E}\left(\|X_i\|^4 + \|Y_i\|^4\right)/n^2\right\}^{1/5} + R_{2,3},$$

where for $p \leq C\sqrt{n}$ it holds $R_{2,3} \leq C(\bar{C}_z \vee \bar{\sigma}_1^4 \tilde{C}_{x,4}\{1 + \mathbf{E}(\varepsilon_i^4)\})(p/\sqrt{n})^{1/3}$; a detailed definition of $R_{2,3}$ is given in (B.17), and constant $C_{\mathscr{B},w,ind} > 0$ is defined in (B.16) (Section B.2 in the supplement [41]).

REMARK 4.2.   Lemma 3.2 implies (see also Remark 2.2) that for the case $K = 4$, $\bar{C}_z$ in Theorem 2.2 can be taken as a generic constant independent of the dimension p. The result in Theorem 4.3 relies on the Berry–Esseen type inequality in Theorem 2.2, and, therefore, we can take $\bar{C}_z = const$ in the above statement.

COROLLARY 4.2.   *Consider the following upper quantile function of the approximating sum obtained using the weighted bootstrap: $Q_{2,w}^{\circ}(\alpha) \coloneqq \inf\{t \in \mathbb{R} : \mathbf{P}^*\left(\|S_n^{\circ}\| > t\right) \leq \alpha\}$, $\alpha \in (0,1)$. Theorem 4.3 implies the following bound*

$$\left|\mathbf{P}\left(\|S_n\| > Q_{2,w}^{\circ}(\alpha)\right) - \alpha\right| \leq 2\Delta_{\mathscr{B},w,ind}^{\circ} + 6e^{-x} \text{ for } x > 0.$$

4.3. *Some remarks about accuracy of the bootstrap procedures.*

REMARK 4.3 (Theorems 4.1, 4.3 in the asymptotic form).   If the conditions of Theorem 4.1 are fulfilled and C_z is dimension-free, then taking $x = \log(2n)$, using the Borel-Cantelli lemma, and the deviation inequality for $\|X\|^2$ by Hsu et al. [20] (see also Section B.4 in the supplement [41]), we have with probability one

$$\sup_{B \in \mathscr{B}} \left|\mathbf{P}\left(S_n \in B\right) - \mathbf{P}^*\left(S_n^* \in B\right)\right|$$
$$= O\left(\{p^{K/(K-2)}/n\}^{1/2}\{1 + \log(n)/p\}^{\frac{K}{2(K-2)}} + (p/n)^{1/2}\{1 + \log(Kn)/p\}^{\frac{(K-1)}{2}}\right)$$

for $n \to \infty$ and $K \geq 3$. Similarly, given the conditions of Theorem 4.3, it holds with probability one

$$\sup_{B \in \mathscr{B}} \left|\mathbf{P}\left(S_n \in B\right) - \mathbf{P}^*\left(S_n^{\circ} \in B\right)\right|$$
$$= O\left((p/\sqrt{n})^{1/3}\{1 + \log(n)/p^{7/6} + (p^2/n)^{1/5}\{1 + \log(n)/p\}^{2/5}\right.$$
$$\left. + (p/\sqrt{n})^{2/3}\{1 + \log(n)/p\}^2\right) \text{ for } p \leq Cn \text{ and } n \to \infty.$$

Theorem 4.3 implies that if the ratio p^2/n (up to $\log n$) is small, then the weighted bootstrap approximation has a good accuracy. Lemma 4.1 below shows that the condition $p^2/n = o(1)$ for $n \to \infty$ is necessary for the weighted bootstrap consistency.

LEMMA 4.1 (Necessary conditions on p and n for the bootstrap consistency).   *Let $X_i \sim \mathcal{N}(0, \mathbf{I}_p)$, $i = 1, \ldots, n$ be i.i.d. random vectors. Consider S_n, S_n^*, and S_n° as in Theorems 4.1 and 4.3. Then*

(a) $p = o(n)$ for $n \to \infty$ is necessary for $\sup_{x \in \mathbb{R}} |\mathbf{P}^*(\|S_n^*\|^2 \leq x) - \mathbf{P}(\|S_n\|^2 \leq x)| \xrightarrow{\mathbf{P}} 0,$

(b) $p^2 = o(n)$ for $n \to \infty$ is necessary for $\sup_{x \in \mathbb{R}} |\mathbf{P}^*(\|S_n^\bullet\|^2 \leq x) - \mathbf{P}(\|S_n\|^2 \leq x)| \xrightarrow{\mathbf{P}} 0.$

REMARK 4.4. In Theorem 4.3, the bootstrap weights $\varepsilon_1, \ldots, \varepsilon_n$ satisfy the 3-rd moment condition (1.2). This is very similar to taking $K = 4$ in Theorem 4.1 for the Efron's bootstrap. It is not possible to continue the sequence of moments (1.2) like $\mathbf{E}(\varepsilon_i^4) = 1, \ldots,$ since the corresponding Hankel matrix $\left(\begin{smallmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{smallmatrix} \right)$ fails the criterion for solubility of the Hamburger moment problem (see, e.g., Akhiezer [1]). Together with Lemma 4.1 and the preceding results in Section 4, this implies that under the conditions of Theorem 4.1, the Efron's bootstrap yields a better accuracy w.r.t. the ratio between p and n, than the considered weighted bootstrap scheme.

REMARK 4.5. In Theorems 4.1-4.3, the sub-Gaussian tail behavior (condition (4.1)) is required in order to apply concentration bounds for the higher-order moments of X_i (see Section B.4 in the supplement [41]). In the asymptotic set-up, one can relax this condition, assuming, e.g. boundedness of the K-th moments of $X_i, i = 1, \ldots, n.$

4.4. *Weighted bootstrap for log-likelihood ratio statistics.* Here we consider a weighted (or a multiplier) bootstrap procedure for estimation of quantiles of log-likelihood ratio statistics. Before describing the procedure and formulating a theoretical result, we give some necessary definitions.

Let $\boldsymbol{y} = (y_1, \ldots, y_n)$ denote the data sample, $y_1, \ldots, y_n$ are i.i.d. random observations from a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Introduce some known parametric family $\{\mathbf{P}_\theta\} \coloneqq \{\mathbf{P}_\theta \ll \mu_0, \theta \in \Theta \subseteq \mathbb{R}^p\}$, here μ_0 is a σ-finite measure on $(\Omega, \mathcal{F})$ which dominates all $\mathbf{P}_\theta$ for $\theta \in \Theta$. The true data distribution $\mathbf{P}$ is not assumed to belong to the family $\{\mathbf{P}_\theta\}$, thus our analysis includes the case when the parametric family $\{\mathbf{P}_\theta\}$ is misspecified. $\{\mathbf{P}_\theta\}$ induces the following (quasi)log-likelihood function for the sample $\boldsymbol{y} \colon L(\theta) = L(\theta, \boldsymbol{y}) \coloneqq \log\left(\frac{d\mathbf{P}_\theta}{d\mu_0}(\boldsymbol{y}) \right)$. The target parameter θ_0 is defined by projecting the true probability distribution $\mathbf{P}$ on the parametric family $\{\mathbf{P}_\theta\}$, using Kullback-Leibler divergence: $\theta_0 \coloneqq \operatorname{argmin}_{\theta \in \Theta} \mathrm{KL}(\mathbf{P}, \mathbf{P}_\theta) = \operatorname{argmax}_{\theta \in \Theta} \mathbf{E}L(\theta)$. The (quasi) maximum likelihood estimate (MLE) is defined as $\tilde{\theta} \coloneqq \operatorname{argmax}_{\theta \in \Theta} L(\theta)$. Let $Q_L(\alpha)$ denote the upper quantile function of square root of the two times log-likelihood ratio statistic: $Q_L(\alpha) \coloneqq \inf\{t \geq 0 : \mathbf{P}\big(L(\tilde{\theta}) - L(\theta) > t^2/2\big) \leq \alpha\}$. $Q_L(\alpha)$ is a critical value of the

likelihood-based confidence set $\mathscr{E}(\alpha)$:

$$(4.2) \quad \mathscr{E}(t) := \{\theta : L(\tilde{\theta}) - L(\theta) \le t^2/2\}, \quad \mathbf{P}\{\theta_0 \in \mathscr{E}(Q_L(\alpha))\} \ge 1 - \alpha.$$

Probability distribution of $L(\tilde{\theta}) - L(\theta_0)$ depends on the unknown parameter θ_0 and $\mathbf{P}$, hence, in general, quantiles of $L(\tilde{\theta}) - L(\theta_0)$ are also unknown.

Consider the weighted (or the multiplier) bootstrap procedure which allows to estimate the distribution of $L(\tilde{\theta}) - L(\theta_0)$. Let $u_1, \ldots, u_n$ be i.i.d. random variables:

$$u_i := \varepsilon_i + 1, \text{ for } \varepsilon_i \text{ defined in (1.2), independent of } \boldsymbol{y}.$$

The bootstrap log-likelihood function $L^{\circ}(\theta)$ equals to the initial one $L(\theta)$ weighted with the random bootstrap weights u_i:

$$L^{\circ}(\theta) := \sum_{i=1}^n \log\left(\frac{d\mathbf{P}_\theta}{d\mu_0}(y_i)\right) u_i.$$

Recall that $\mathbf{P}^*(\cdot) := \mathbf{P}(\cdot \,|\, \{y_i\}_{i=1}^n)$ and $\mathbf{E}^*(\cdot) := \mathbf{E}(\cdot \,|\, \{y_i\}_{i=1}^n)$. It holds $\mathbf{E}^* L^{\circ}(\theta) = L(\theta)$, therefore, $\tilde{\theta} = \operatorname{argmax}_{\theta \in \Theta} L(\theta) = \operatorname{argmax}_{\theta \in \Theta} \mathbf{E}^* L^{\circ}(\theta)$, and the MLE $\tilde{\theta}$ can be considered as a bootstrap analogue of the unknown target parameter θ_0. The bootstrap likelihood ratio statistic is defined as

$$L^{\circ}(\tilde{\theta}^{\circ}) - L^{\circ}(\tilde{\theta}) := \sup_{\theta \in \Theta} L^{\circ}(\theta) - L^{\circ}(\tilde{\theta}).$$

$L^{\circ}(\tilde{\theta}^{\circ}) - L^{\circ}(\tilde{\theta})$ can be computed for each i.i.d. sample of the bootstrap weights $u_1, \ldots, u_n$, thus we can calculate empirical probability distribution function of $L^{\circ}(\tilde{\theta}^{\circ}) - L^{\circ}(\tilde{\theta})$ and estimate its quantiles. Denote

$$(4.3) \qquad Q_L^{\circ}(\alpha) := \inf\{t \ge 0 : \mathbf{P}^*\big(L^{\circ}(\tilde{\theta}^{\circ}) - L^{\circ}(\theta) > t^2/2\big) \le \alpha\}.$$

Theorem 4.4 below provides a two-sided bound on the coverage error of the likelihood confidence set (4.2) based on the bootstrap quantile $Q_L^{\circ}(\alpha)$. Let us introduce some additional notation before stating the theorem. Denote $\ell_i(\theta) := \log\left(\frac{d\mathbf{P}_\theta}{d\mu_0}(y_i)\right)$, $d_0^2 := -\mathbf{E}\ell_1''(\theta_0)$, here $\ell_i'(\theta) := \nabla_\theta \ell_i(\theta)$. Take $X_i := d_0^{-1}\ell_i'(\theta_0)$. By previous definitions, such defined $\{X_i\}_{i=1}^n$ are i.i.d with zero mean. Moreover, if conditions from Section B.3 in the supplement [41] are fulfilled, then $\mathbf{E}(\|X_i\|^4) < \infty$. Let $Y_1, \ldots, Y_n$ be i.i.d. vectors meeting conditions (2.1) and (2.2) for $K = 4$, and $C_{z,L} := \|\{\operatorname{Var}(Z_i)\}^{-1/2}\|$. Now we are ready to formulate the following

THEOREM 4.4.   *If the conditions from Section B.3 in the supplement [41] are fulfilled, then it holds with probability $\geq 1 - 10e^{-x}$ for $x > 0$*

$$\sup_{t \geq 0} \left| \mathbf{P} \left\{ L(\tilde{\theta}) - L(\theta_0) \leq t \right\} - \mathbf{P}^* \left\{ L^{\circledcirc}(\tilde{\theta}^{\circledcirc}) - L^{\circledcirc}(\tilde{\theta}) \leq t \right\} \right| \leq \Delta_L, \text{ and}$$

$$\left| \mathbf{P} \left\{ \theta_0 \notin \mathscr{E}(Q_L^{\circledcirc}(\alpha)) \right\} - \alpha \right| \leq 2\Delta_L + 10e^{-x}, \text{ where}$$

$$(4.4) \qquad \Delta_L \leq C_{\mathscr{B},w,ind} \left\{ C_{z,L}^4 \mathbf{E} \left(\|d_0^{-1}\ell_1'(\theta_0)\|^4 + \|Y_1\|^4 \right)/n \right\}^{1/5}$$
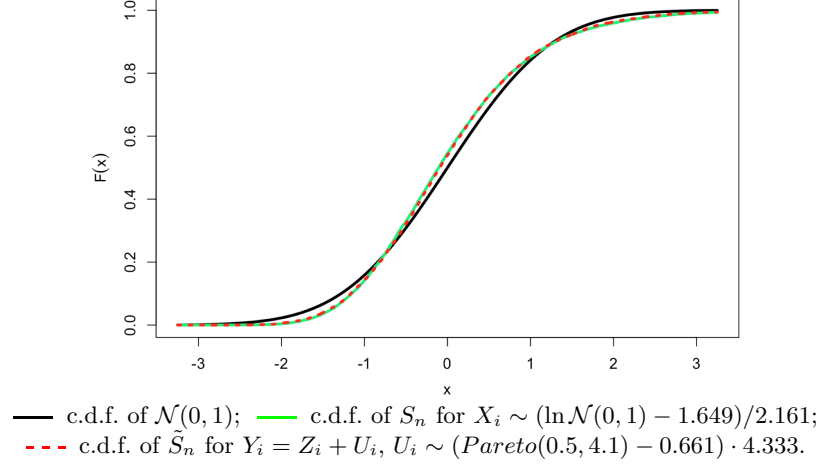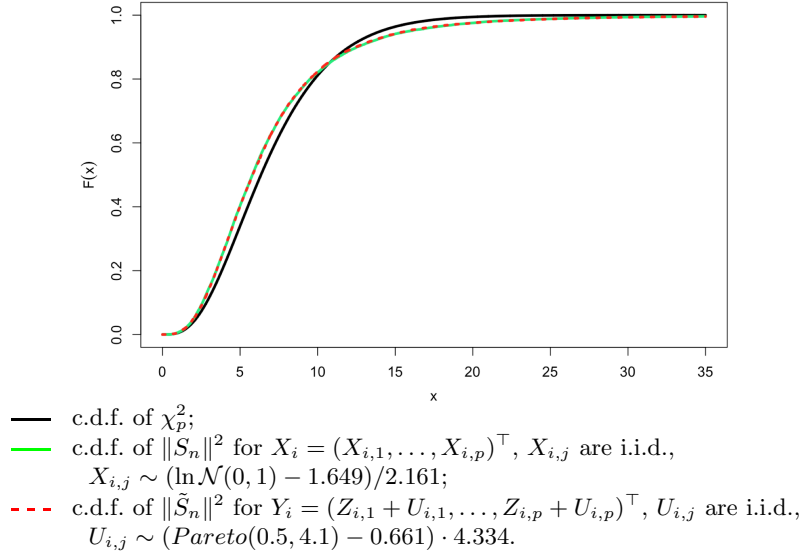$$+ R_{L,2,3} + C_{z,L}C(p+x)n^{-1/2},$$

where for $p \leq C\sqrt{n}$ it holds $R_{L,2,3} \leq C(C_{z,L} \vee (\nu\mathfrak{a})^4 \tilde{C}_{x,4}\{1+\mathbf{E}(\varepsilon_i^4)\}) \, (p/\sqrt{n})^{1/3}$; is defined in (B.17); a more detailed definition of the error term Δ_L is given in (B.20), $C_{\mathscr{B},w,ind}$ is defined in (B.16) (Section B.2 in the supplement [41]).

REMARK 4.6.   The third term in bound (4.4) comes from Wilks-type approximations for the likelihood ratios $L(\tilde{\theta}) - L(\theta_0)$ and $L^{\circledcirc}(\tilde{\theta}^{\circledcirc}) - L^{\circledcirc}(\tilde{\theta})$ (see the proof in Section B.3 in the supplement [41] for more details); the first two terms in (4.4) come from the Berry–Esseen type inequality justifying the weighted bootstrap procedure on the set $\mathscr{B}$ (Theorem 4.3). Due to the assumed sub-Gaussian tail behavior of $d_0^{-1}\ell_i'(\theta_0)$, the first term is bounded from above with $Cpn^{-1/2}$ with large probability. Thus, in Theorem 4.4 both Wilks-type bound and the higher-order Berry–Esseen type inequality yield similar ratios between p and n in the error of approximation Δ_L.

5. Numerical experiments.   This section presents results of simulation studies, illustrating accuracy of the considered Berry–Esseen bounds and the bootstrap procedures.

5.1. *Berry–Esseen inequality.*   Figure 1 shows the c.d.f.-s of S_n, $\tilde{S}_n$ and $\mathcal{N}(0,1)$ for the sample size $n = 50$, dimension $p = 1$ and number $K - 1 = 3$ of equal moments of S_n and $\tilde{S}_n$. Similarly Figure 2 shows c.d.f.-s of $\|S_n\|^2$, $\|\tilde{S}_n\|^2$ and χ_p^2 for $n = 50$, $p = 7$ and $K - 1 = 3$. Distributions of X_i and Y_i are described in the bottom of each of the Figures 1 and 2. The c.d.f.-s are obtained from $15 \cdot 10^3$ i.i.d. samples. Both figures agree with the theoretical results about the higher order Berry–Esseen bounds: the latter approximation has a better accuracy than the Gaussian approximation.

5.2. *Bootstrap.*   Here we examine accuracy of the bootstrap procedures for $\|S_n\|$ (described in Section 4) by computing coverage probabilities using bootstrap quantiles $Q_2^{\circledcirc}(\alpha)$. All the results are collected in Table 1. Columns n, p, $\mathscr{L}(\varepsilon_i)$, $\mathscr{L}(X_{i,j})$ show the sample size, the dimension, the distribution of the bootstrap weights ε_i, and the distribution of $X_{i,j}$, where i.i.d. coordinates $X_{i,j}$ are s.t. $X_i = (X_{i,1}, \ldots, X_{i,p})^\top$. Nominal coverage probabilities

FIG 1. *Distribution functions of S_n and $\tilde{S}_n$ for $n = 50$, $p = 1$, $K = 4$.*

— c.d.f. of $\mathcal{N}(0, 1)$; — c.d.f. of S_n for $X_i \sim (\ln \mathcal{N}(0, 1) - 1.649)/2.161$;
- - - c.d.f. of $\tilde{S}_n$ for $Y_i = Z_i + U_i$, $U_i \sim (Pareto(0.5, 4.1) - 0.661) \cdot 4.333$.

FIG 2. *Distribution functions of $\|S_n\|^2$ and $\|\tilde{S}_n\|^2$ for $n = 50$, $p = 7$, $K = 4$.*

— c.d.f. of χ_p^2;
— c.d.f. of $\|S_n\|^2$ for $X_i = (X_{i,1}, \ldots, X_{i,p})^\top$, $X_{i,j}$ are i.i.d.,
   $X_{i,j} \sim (\ln \mathcal{N}(0, 1) - 1.649)/2.161$;
- - - c.d.f. of $\|\tilde{S}_n\|^2$ for $Y_i = (Z_{i,1} + U_{i,1}, \ldots, Z_{i,p} + U_{i,p})^\top$, $U_{i,j}$ are i.i.d.,
   $U_{i,j} \sim (Pareto(0.5, 4.1) - 0.661) \cdot 4.334$.

$1 - \alpha$ are given in the second row $0.975, 0.95, \ 0.90, 0.85, \ldots, 0.50$. All the rest numbers represent frequencies of the event $\{\|S_n\| \leq Q^\bullet(\alpha)\}$, computed for different n, p, α, $\mathscr{L}(\varepsilon_i)$, and $\mathscr{L}(X_{i,j})$, from $7 \cdot 10^3$ i.i.d. samples $\{X_i\}_{i=1}^n$ and $\{\varepsilon_i\}_{i=1}^n$. We consider three types of the bootstrap weights: first one $\varepsilon_i = z_i + u_i$, with $u_i \sim (Bernoulli(b) - b)\sigma_u$, $b = 0.276, \sigma_u \approx 2.235$, and $z_i \sim \mathcal{N}(0, \sigma_z^2)$, $\sigma_z^2 \approx 0.038$, for this case $\mathbf{E}\varepsilon_i = 0$, $\mathbf{E}(\varepsilon_i^2) = \mathbf{E}(\varepsilon_i^3) = 1$,

therefore ε_i meet conditions (1.2). The second type is $\varepsilon_i \sim \mathcal{N}(0, 1)$, in this case $\mathbf{E}(\varepsilon_i^3) \neq 1$, and the approximation accuracy corresponds to the classical normal approximation with a larger error term. In this numerical experiment we check, whether the additional condition $\mathbf{E}(\varepsilon_i^3) = 1$ improves numerical performance of the weighted bootstrap for $\|S_n\|$. The third type of the weights *Multinom.* corresponds to the multinomial distribution $Multinomial(n; 1/n, \ldots, 1/n)$, i.e., to the classical Efron's bootstrap scheme. Table 1 confirms the higher-order properties of the bootstrap schemes for most of the computed coverage probabilities.

TABLE 1
Coverage probabilities $\mathbf{P}\big(\|S_n\| \leq Q_2^\circ(\alpha)\big)$

n	p	$\mathscr{L}(X_{i,j})$	$\mathscr{L}(\varepsilon_i)$	Confidence levels							
				0.975	**0.95**	**0.90**	**0.85**	**0.80**	**0.70**	**0.60**	**0.50**
400	40	$\chi_1^2 - 1$	$\mathscr{L}(z_i + u_i)$	0.982	0.957	0.910	0.855	0.804	0.701	0.595	0.491
			$\mathcal{N}(0, 1)$	0.984	0.960	0.914	0.862	0.810	0.704	0.597	0.495
			$Multinom.$	0.983	0.960	0.916	0.864	0.812	0.702	0.593	0.492
		$Pareto^*$	$\mathscr{L}(z_i + u_i)$	0.984	0.964	0.917	0.865	0.813	0.704	0.593	0.490
			$\mathcal{N}(0, 1)$	0.986	0.972	0.925	0.873	0.821	0.707	0.589	0.480
			$Multinom.$	0.989	0.969	0.927	0.875	0.822	0.710	0.591	0.475
		$\ln \mathcal{N}^*(2.5)$	$\mathscr{L}(z_i + u_i)$	0.996	0.987	0.958	0.912	0.863	0.711	0.555	0.416
			$\mathcal{N}(0, 1)$	0.998	0.992	0.973	0.934	0.880	0.725	0.543	0.387
			$Multinom.$	0.998	0.994	0.967	0.914	0.847	0.678	0.511	0.390
150	15	$\chi_1^2 - 1$	$\mathscr{L}(z_i + u_i)$	0.983	0.958	0.907	0.855	0.807	0.703	0.596	0.492
			$\mathcal{N}(0, 1)$	0.986	0.965	0.915	0.863	0.811	0.706	0.595	0.485
			$Multinom.$	0.986	0.964	0.912	0.855	0.826	0.683	0.576	0.468
		$Pareto^*$	$\mathscr{L}(z_i + u_i)$	0.985	0.967	0.920	0.869	0.807	0.695	0.585	0.472
			$\mathcal{N}(0, 1)$	0.990	0.974	0.931	0.882	0.820	0.697	0.580	0.459
			$Multinom.$	0.988	0.973	0.926	0.866	0.804	0.683	0.561	0.449
		$\ln \mathcal{N}^*(2.5)$	$\mathscr{L}(z_i + u_i)$	0.992	0.978	0.936	0.889	0.830	0.674	0.514	0.386
			$\mathcal{N}(0, 1)$	0.995	0.987	0.956	0.910	0.851	0.693	0.507	0.357
			$Multinom.$	0.996	0.987	0.949	0.891	0.818	0.656	0.494	0.349
50	5	$\chi_1^2 - 1$	$\mathscr{L}(z_i + u_i)$	0.985	0.961	0.906	0.853	0.798	0.688	0.582	0.483
			$\mathcal{N}(0, 1)$	0.988	0.969	0.915	0.862	0.804	0.688	0.572	0.466
			$Multinom.$	0.985	0.959	0.900	0.845	0.785	0.667	0.555	0.454
		$Pareto^*$	$\mathscr{L}(z_i + u_i)$	0.983	0.960	0.911	0.852	0.795	0.675	0.560	0.460
			$\mathcal{N}(0, 1)$	0.986	0.967	0.923	0.866	0.804	0.673	0.546	0.432
			$Multinom.$	0.984	0.960	0.908	0.844	0.777	0.650	0.528	0.424
		$\ln \mathcal{N}^*(1.5)$	$\mathscr{L}(z_i + u_i)$	0.977	0.956	0.903	0.839	0.775	0.638	0.532	0.411
			$\mathcal{N}(0, 1)$	0.983	0.965	0.920	0.858	0.795	0.645	0.506	0.382
			$Multinom.$	0.980	0.958	0.901	0.833	0.763	0.621	0.493	0.383

Here $Pareto^*$ and $\ln \mathcal{N}^*(\sigma^2)$ denote zero mean distributions $Pareto(0.5, 4.1) - 0.661$ and $\ln \mathcal{N}(0, \sigma^2) - e^{\sigma^2/2}$ correspondingly.

SUPPLEMENTARY MATERIAL

Supplement to "Nonclassical Berry–Esseen inequalities and accuracy of the bootstrap"

(). The supplementary material contains proofs of the results from Sections 2 and 4.

References.

[1] Akhiezer, N. I. (1965). *The classical moment problem and some related questions in analysis*, volume 5. Oliver & Boyd.

[2] Arlot, S., Blanchard, G., and Roquain, E. (2010). Some nonasymptotic results on resampling in high dimension. I. Confidence regions. *The Annals of Statistics*, 38(1):51–82.

[3] Barbe, P. and Bertail, P. (1995). *The weighted bootstrap*, volume 98. Springer.

[4] Bentkus, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *Journal of Statistical Planning and Inference*, 113(2):385–402.

[5] Bentkus, V. (2005). A Lyapunov-type bound in R^d. *Theory of Probability & Its Applications*, 49(2):311–323.

[6] Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist*, 6(2):434–451.

[7] Bickel, P. J. and Freedman, D. A. (1983). Bootstrapping regression models with many parameters. *Festschrift for Erich L. Lehmann*, pages 28–48.

[8] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

[9] Chatterjee, S. and Bose, A. (2005). Generalized bootstrap for estimating equations. *The Annals of Statistics*, 33(1):414–436.

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

[11] Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309–2352.

[12] Curto, R. E. and Fialkow, L. A. (1991). Recursiveness, positivity, and truncated moment problems. *Houston Journal of Mathematics*, 17(4):603–635.

[13] Curto, R. E. and Fialkow, L. A. (2008). An analogue of the Riesz–Haviland theorem for the truncated moment problem. *Journal of Functional Analysis*, 255(10):2709–2731.

[14] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, pages 1–26.

[15] Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

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

[17] Holmes, S. and Reinert, G. (2004). Stein's method for the bootstrap. In *Stein's Method*, pages 93–132. Institute of Mathematical Statistics.

[18] Holmquist, B. (1988). Moments and cumulants of the multivariate normal distribution. *Stochastic Analysis and Applications*, 6(3):273–278.

[19] Horowitz, J. L. (2001). The bootstrap. *Handbook of econometrics*, 5:3159–3228.

[20] Hsu, D., Kakade, S. M., and Zhang, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab*, 17(52):1–6.

[21] Ibragimov, I. A. (1966). On the accuracy of Gaussian approximation to the distribution functions of sums of independent variables. *Theory of Probability & Its Applications*, 11(4):559–579.

[22] Janssen, A. and Pauls, T. (2003). How do bootstrap and permutation tests work? *The Annals of Statistics*, 31(3):768–806.

[23] Liu, R. Y. (1988). Bootstrap Procedures under some Non-I.I.D. Models. *The Annals of Statistics*, 16(4):1696–1708.

[24] Loève, M. (1950). Fundamental limit theorems of probability theory. *The Annals of Mathematical Statistics*, 21(3):321–338.

[25] Mammen, E. (1989). Asymptotics with increasing dimension for robust regression with applications to the bootstrap. *The Annals of Statistics*, pages 382–400.

[26] Mammen, E. (1992). *When does bootstrap work?*, volume 77. Springer.

[27] Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21(1):255–285.

[28] Paulauskas, V. (1975). An estimate of the remainder term in the multidimensional central limit theorem. *Lithuanian Mathematical Journal*, 15(3):484–493.

[29] Præstgaard, J. (1990). Bootstrap with general weights and multiplier central limit theorems. *Technical Report195, Department of Statistics, University of Washington*.

[30] Præstgaard, J. and Wellner, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *The Annals of Probability*, pages 2053–2086.

[31] Rotar', V. (1978). Non-classical estimates of the rate of convergence in the multidimensional central limit theorem. I. *Theory of Probability & Its Applications*, 22(4):755–772.

[32] Schwartz, J. T. (1980). Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM (JACM)*, 27(4):701–717.

[33] Shao, J. and Tu, D. (1995). *The jackknife and bootstrap*. Springer.

[34] Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics*, pages 1187–1195.

[35] Spokoiny, V. and Zhilova, M. (2015). Bootstrap confidence sets under model misspecification. *The Annals of Statistics*, 43(6):2653–2675.

[36] Ul'yanov, V. V. (1979). On more precise convergence rate estimates in the central limit theorem. *Theory of Probability & Its Applications*, 23(3):660–663.

[37] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical processes*. Springer, New York.

[38] Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14(4):1261–1295+.

[39] Zhai, A. (2018). A high-dimensional CLT in $\mathcal{W}_2$-distance with near optimal convergence rate. *Probability Theory and Related Fields*, 170(3):821–845.

[40] Zhilova, M. (2015). Simultaneous likelihood-based bootstrap confidence sets for a large number of models. *arXiv:1506.05779*.

[41] Zhilova, M. (2020). Supplement to "Nonclassical Berry–Esseen inequalities and accuracy of the bootstrap".

[42] Zippel, R. (1979). Probabilistic algorithms for sparse polynomials. *Symbolic and algebraic computation*, pages 216–226.

[43] Zolotarev, V. M. (1965). On the closeness of the distributions of two sums of independent random variables. *Theory of Probability & Its Applications*, 10(3):472–479.

School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332-0160 USA
E-mail: mzhilova@math.gatech.edu