Non-classical Berry-Esseen inequality and accuracy of the weighted bootstrap

Mayya Zhilova

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

Abstract: We study accuracy of a weighted bootstrap procedure for estimation of quantiles of Euclidean norm of a sum of independent random vectors with zero mean and bounded fourth moment. We establish higher-order approximation bounds with error terms depending explicitly on a sample size and a dimension. 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 Berry-Esseen inequality in a non-classical form. 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 set of all Euclidean balls in $\mathbb{R}^p$. This approximation is an extension of a Gaussian one. The theoretical results are illustrated with numerical experiments.

MSC 2010 subject classifications: Primary 62E17, 62F40; secondary 62F25.
Keywords and phrases: Multivariate Berry-Esseen inequality, dependence on dimension, weighted bootstrap, multiplier bootstrap, likelihood-based confidence set, linear regression model.

1. Introduction

In this paper we study accuracy of a weighted (or a multiplier) bootstrap procedure for estimation of quantiles of statistics of the form $\|S_n\|$, where $\|\cdot\|$ denotes ℓ_2-norm in $\mathbb{R}^p$, and

$$S_n \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^{n} X_i$$

for independent random vectors $X_1, \ldots, X_n \in \mathbb{R}^p$ such that $\forall\, i = 1, \ldots, n$

$$\mathbf{E} X_i = 0,\ \mathbf{E}(\|X_i\|^4) < \infty, \text{ and } \operatorname{Var}(X_i) \text{ is positive definite.}$$

We consider the setting when the sample size n is bounded, and approximation errors depend on n and dimension p explicitly. This allows to assess accuracy and limitations of bootstrap approximation in terms of dimension and sample size. Estimation of distribution of statistics of the type $\|S_n\|$ 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 elliptical confidence sets for multivariate sample mean.

The weighted bootstrap procedure for $\|S_n\|$ is defined as follows. Introduce the random variables

$$\begin{gathered}\varepsilon_1,\dots,\varepsilon_n\in\mathbb{R}\,,\text{ i.i.d., independent of }\{X_i\}\,,\\ \mathbf{E}\varepsilon_i=0,\ \mathbf{E}(\varepsilon_i^2)=1,\ \mathbf{E}(\varepsilon_i^3)=1,\ \mathbf{E}(\varepsilon_i^4)<\infty.\end{gathered} \tag{1.1}$$

Define for the sum S_n its bootstrap version:

$$S_n^{\circ}\overset{\text{def}}{=}\frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^{n}X_i\varepsilon_i. \tag{1.2}$$

The unknown quantiles of the initial statistic $\|S_n\|$ are approximated with the quantiles of $\|S_n^{\circ}\|$ conditioned on the sample $X_1,\dots,X_n$. Denote the upper quantile function of $\|S_n^{\circ}\|$ as

$$Q^{\circ}(\alpha)\overset{\text{def}}{=}\inf\left\{t\in\mathbb{R}:\mathbf{P}^{\circ}\left(\|S_n^{\circ}\|>t\right)\le\alpha\right\}, \tag{1.3}$$

where $\alpha\in(0,1)$, and $\mathbf{P}^{\circ}(\cdot)\overset{\text{def}}{=}\mathbf{P}(\cdot\,|\,X_1,\dots,X_n)$. One of the main results of the paper is the following approximation bound between $\mathbf{P}\left(\|S_n\|>Q^{\circ}(\alpha)\right)$ and α: if $X_1,\dots,X_n$ are i.i.d., then

$$\left|\mathbf{P}\left(\|S_n\|>Q^{\circ}(\alpha)\right)-\alpha\right|\le C_{\Sigma}\frac{\sqrt{\mathbf{E}\left(\|X_1\|^4\right)\mathbf{E}(\varepsilon_1^4)}}{\sqrt{n}}, \tag{1.4}$$

where the constant C_{Σ} depends on the largest eigenvalue of the matrix $(\operatorname{Var}X_1)^{-1}$. This bound implies, in particular, that if the random vector X_1 is sub-gaussian, then the ratio $p/\sqrt{n}$ has to be small in order to keep the bootstrap approximation accurate. This complies with the results by [7] for residual bootstrap, the authors showed that the bootstrap least squares estimate in high dimensional linear regression model converges in Mallow's distance to the original estimator if $p^2/n\to 0$. [28] extended these results for M-estimators.

Conditions (1.1) on the bootstrap weights play an important role for obtaining the accuracy in approximating bound (1.4). [26] used the condition $\mathbf{E}(\varepsilon_i^3)=1$ in order to obtain the second order accuracy of the wild bootstrap (or Wu's bootstrap, first proposed by [50]) approximation to the least squares estimate in a linear regression model. [30] studied validity and higher-order accuracy of the wild bootstrap 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. Here we impose condition $\mathbf{E}(\varepsilon_i^3)=1$ in order to obtain a higher-order accuracy as well. Consider the first two moments of the bootstrap sum (1.2) w.r.t. the joint distribution of $\{X_i\}$ and $\{\varepsilon_i\}$. By $\mathbf{E}\varepsilon_i=0,\ \mathbf{E}(\varepsilon_i^2)=1$ it holds

$$\mathbf{E}S_n=\mathbf{E}S_n^{\circ},\quad\mathbf{E}(S_nS_n^{\top})=\mathbf{E}(S_n^{\circ}S_n^{\circ\top}). \tag{1.5}$$

Using (1.5) and normal approximation between distributions of $\|S_n\|$ and $\|S_n^{\circ}\|$ (e.g. the results of [4]), one can obtain an approximation bound similar to (1.4), with an error term $C_{\Sigma}^{3/2}\mathbf{E}\left(\|X_1\|^3\right)\mathbf{E}(\varepsilon_1^3)/\sqrt{n}$, 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

$$\forall \alpha \in \mathbb{R}^p \ \mathbf{E}\{(\alpha^\top S_n)^3\} = \mathbf{E}\{(\alpha^\top S_n^{\circ})^3\}, \tag{1.6}$$

and this property leads to the improved error term in (1.4). In order to employ the information about the third moments, as in (1.6), one needs to use an approximation, which is more general than the normal one. For this purpose we consider a multivariate Berry-Esseen type bound. Before introducing the latter result, let us mention that approximation (1.4) leads also to an improvement of accuracy of a weighted bootstrap procedure for general log-likelihood ratio statistics. [47] considered weighted bootstrap for estimation of quantiles of 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 normal approximation between statistics of the type $\|S_n\|$ and $\|S_n^{\circ}\|$. Using inequality (1.4), we infer that the accuracy of weighted bootstrap for log-likelihood ratio depends rather on accuracy of Wilks-type bounds, then on normal approximation.

The key element in the proofs of our theoretical results about accuracy of the bootstrap is a multivariate Berry-Esseen inequality in a non-classical form, which might be interesting by itself. We consider a problem of approximation of probability distribution of the sum $S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$, where $X_i \in \mathbb{R}^p$ are independent random vectors such that $\mathbf{E}X_i = 0$ and $\mathbf{E}\left(\|X_i\|^K\right) < \infty$ for some $K \geq 3$. The approximating distribution corresponds to the following sum

$$\widetilde{S}_n \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^n Y_i,$$

where $Y_1,\dots,Y_n \in \mathbb{R}^p$ are independent random vectors, independent of $\{X_i\}_{i=1}^n$, such that $\mathbf{E}\left(\|Y_i\|^K\right) < \infty$,

$$\mathbf{E}\bigl(X_i^k\bigr) = \mathbf{E}\bigl(Y_i^k\bigr) \ \forall k = 1,\dots,K-1, \tag{1.7}$$

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

$$\mathbf{E}(x_{i_1}\dots x_{i_k}) = \mathbf{E}(y_{i_1}\dots y_{i_k}). \tag{1.8}$$

In Lemma 2.1 (in Section 2) we show that if X_i has a continuous probability distribution, then the corresponding random vectors Z_i, U_i always exist. However, in general, the continuity condition is not necessary for existence of Z_i and U_i. The probability distribution of such constructed random vector $\widetilde{S}_n$ turns out to be a rather good approximation of the 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. $X_1, \dots, X_n$

$$\sup_{B \in \mathscr{B}} \bigl| \mathbf{P}(S_n \in B) - \mathbf{P}(\widetilde{S}_n \in B) \bigr| \le C_{K,\Sigma} \frac{\bigl\{ \mathbf{E}\left(\|X_1\|^K + \|Y_1\|^K \right) \bigr\}^{1/(K-2)}}{n^{1/2}}, \quad (1.9)$$

where constant $C_{K,\Sigma}$ depends on K and on the largest eigenvalue of the matrix $(\operatorname{Var} Z_1)^{-1}$. We study also the case of independent but not necessarily identically distributed summands X_i. Bound (1.9) includes the classical Berry-Esseen inequality, when the approximating distribution is purely Gaussian, 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 Gaussian approximation does, which leads to a better accuracy.

Our proof of bound (1.9) is based on the work of [4], where the author obtained the multivariate Berry-Esseen inequality in purely standard Gaussian case, uniformly on the set of all Euclidean balls, and also on the set of all convex sets in $\mathbb{R}^p$. The results by [4] have the best known dependence of the approximation error on dimension. In this paper we extend the proof of [4] to the "quasi-Gaussian" case, i.e. for the approximation with the sum $\widetilde{S}_n$ 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 Gaussian distribution and the higher moments condition (1.7).

Below we give an overview of the existing literature and discuss contribution of this paper to it. Weighted bootstrap is a general version of the classical Efron's bootstrap (proposed by [13]). According to the latter method, bootstrap approximation of an empirical measure of a random sample is constructed by weighing the empirical measure with multinomial random weights, conditioned on the sample. Later [31] extended this scheme for general exchangeable random weights. Let us mention the papers [38, 34, 35] as some of the first works about general weighted bootstrap. Let us also refer to the book by [3] and to the paper by [22] for exhaustive literature reviews about this topic. 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 the books [17, 29, 43]). Accuracy of bootstrap procedures is usually studied using Edgeworth expansions or Berry-Esseen-type inequalities (the latter technique had been first used by [44] and [26] in the framework of bootstrap).

In the asymptotic high-dimensional setting when both the parameter dimension p and the sample size n are large, [7, 28, 30] studied accuracy of Efron's and wild bootstrap for linear regression model; [8] studied generalized bootstrap for estimating equations also in high-dimensional asymptotic framework. Most of the non-asymptotic results (i.e. without using asymptotic arguments w.r.t. n) about bootstrap are quite recent. [1] 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 Gaussian) distribution. [10] studied Gaussian approximation and multiplier bootstrap for maxima of sums of high-dimensional vectors in a very general set-up. [11] extended the results from maxima to general hyperractangles and sparsely convex sets.

In the approximation bound (1.4) obtained in this work we do not use any asymptotic arguments. Moreover, this bound has a better accuracy than a Gaussian approximation does. This justifies that weighted bootstrap can outperform Gaussian approximation if the bootstrap weights are properly chosen (for example, as in (1.1)). We apply this result for construction of confidence sets for least squares estimate in a linear regression model, and for likelihood-based confidence sets.

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. In particular, convergence of a distribution of S_n in context of convergence of its moments had been considered by P. Chebyshev, A. Markov, P. Lévy (see the paper by [27] for detailed overview). [21] studied convergence of distribution of S_n in case of i.i.d. scalar summands, to standard normal law, under higher moments condition; the author obtained a higher-order accuracy using Edgeworth expansion. [51] introduced pseudomoments, which characterize closeness of moments of two distributions, for estimation of convergence rates in limit theorems. These characteristics turned out to be very useful for refining the classical limit theorems, without imposing condition of uniform asymptotic negligibility of the summands. Such limit theorems are called non-classical. In the multivariate finite-dimensional case some of the first non-classical results about normal approximation on closed convex sets had been obtained by [33, 37, 48]. [32, 40, 41, 14, 9, 15, 4, 5, 10, 11] studied normal approximation in finite-dimensional space. Let us refer to the books [6, 52, 42] for comprehensive overview of earlier results on these topics. The results of [4] have the best known dependence on dimension among the Berry-Esseen type bounds for $\|S_n\|$.

To the best of our knowledge, the problem of approximation of probability distribution of S_n under the higher moments condition (1.7) and with explicit dependence on p, had not been studied before.

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 Section 2. Section 3 contains theoretical results regarding bootstrap accuracy. In Sections 3.2 and 3.3 we consider the weighted bootstrap for the linear regression model and for log-likelihood ratio statistics correspondingly. Sections A and B contain proofs of results from Sections 2 and 3 respectively. Section 4 presents results of numerical experiments.

Notation

$\|\cdot\|$ denotes Euclidean norm for vectors and spectral norm for matrices; S_p^+ denotes space of symmetric positive definite real-valued matrices of size $p \times p$; $\mathscr{B}$ is the set of all 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 $X^{\otimes k}$; C indicates positive generic constant unless specified otherwise.

2. Non-classical Berry-Esseen inequality

Consider independent random vectors $X_1,\dots,X_n \in \mathbb{R}^p$ such that $\forall i = 1,\dots,n$ $\mathbf{E}X_i = 0$, $\operatorname{Var}(X_i) \in S_p^+$, $\mathbf{E}(\|X_i\|^K) < \infty$ for some integer $K \geq 3$. Let $Y_1,\dots,Y_n \in \mathbb{R}^p$ be independent random vectors, and such that $\forall i = 1,\dots,n$

$$\begin{gathered} Y_i \text{ is independent of } X_1,\dots,X_n, \\ \mathbf{E}\left(\|Y_i\|^K\right) < \infty, \\ \mathbf{E}\big(X_i^k\big) = \mathbf{E}\big(Y_i^k\big)\ \forall k = 1,\dots,K-1, \end{gathered} \tag{2.1}$$

A formal definition of the equality of the higher-order moments of vector-valued random variables (as in (2.1)) is given in (1.8). We assume also that $\forall i = 1,\dots,n$

$$\begin{gathered} \exists \text{ independent r.v. } Z_i, U_i \in \mathbb{R}^p \text{, such that} \\ Y_i \overset{d}{=} Z_i + U_i,\ \mathbf{E}Z_i = \mathbf{E}U_i = 0, \\ Z_i \sim \mathcal{N}(0, \varSigma_{\boldsymbol{z},i}) \text{ for some } \varSigma_{\boldsymbol{z},i} \in S_p^+ . \end{gathered} \tag{2.2}$$

In Lemma 2.1 below it is shown that continuity of a probability distribution of X_i is sufficient for existence of the r.v. Y_i described above. Consider the following sums of mutually independent zero mean random vectors:

$$S_n \overset{\text{def}}{=} \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^n X_i, \quad \widetilde{S}_n \overset{\text{def}}{=} \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^n Y_i. \tag{2.3}$$

We establish unform approximation bounds between probability distributions of S_n and $\widetilde{S}_n$ on the set $\mathscr{B}$ of all Euclidean balls in $\mathbb{R}^p$. Theorem 2.1 treats the case when $\{X_i\}_{i=1}^n$ are i.i.d.; the case of independent but not necessarily identically distributed vectors $\{X_i\}_{i=1}^n$ is considered in Theorem 2.2.

Let us introduce some additional notation before stating the first result. $M > 0$ is a generic constant, in the proof in Section A.2 we show that one can take $M \geq 72.5$. $\widetilde{C}_{\mathscr{B}} > 0$ is an isoperimetric constant of the set $\mathscr{B}$ corresponding to the standard Gaussian measure in $\mathbb{R}^p$, namely, for $Z_0 \sim \mathcal{N}(0, \mathbf{I}_p)$ and $\forall\, r, \varepsilon > 0$

$$\mathbf{P}\left(r \leq \|Z_0\| \leq r + \varepsilon\right) \leq \widetilde{C}_{\mathscr{B}} \varepsilon; \tag{2.4}$$

due to results of [2], the constant $\widetilde{C}_{\mathscr{B}}$ is dimension-free (see Section A.1 for more detail). In the statement of Theorem 2.1 we use $C_{\mathscr{B}} \overset{\text{def}}{=} \max\{1, \widetilde{C}_{\mathscr{B}}\}$. In the proof of Theorem 2.1 we use function $\varphi(x) : \mathbb{R}^p \mapsto [0,1]$, which is at least $K-1$ times continuously differentiable approximation of the indicator function $1\!\!1\{x \in B_{r+1} \setminus B_r\}$, where $B_r \in \mathscr{B}$ is some Euclidean ball of radius r; constants $\widetilde{C}_\phi, \widetilde{C}_{\phi,1}$ enter upper bounds on supremum norms of the higher-order derivatives $\varphi^{(K-2)}(x)$ and $\varphi^{(K-1)}(x)$ (see Lemma A.3 in Section A.2). In the statement of Theorem 2.1 we use $C_\phi \stackrel{\text{def}}{=} \max\{1, \widetilde{C}_\phi, \widetilde{C}_{\phi,1}\}$. For the case of i.i.d. summands X_i (and hence i.i.d. Z_i) denote $\varSigma_{\boldsymbol{z}} \stackrel{\text{def}}{=} \varSigma_{\boldsymbol{z},i} = \operatorname{Var}(Z_i)$. Since $\varSigma_{\boldsymbol{z}} \in S_p^+$, we can define $C_{\boldsymbol{z}} \stackrel{\text{def}}{=} \|\varSigma_{\boldsymbol{z}}^{-1/2}\|$.

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 $\widetilde{S}_n$ defined in* (2.3)

$$\sup_{B\in\mathscr{B}} \left|\mathbf{P}\left(S_n \in B\right) - \mathbf{P}\big(\widetilde{S}_n \in B\big)\right| \le M C_{\mathscr{B}} C_\phi \frac{\left\{C_{\boldsymbol{z}}^K \mathbf{E}\left(\|X_1\|^K + \|Y_1\|^K\right)\right\}^{1/(K-2)}}{n^{1/2}}.$$

Remark 2.1 (The case of Gaussian approximation)*.* If the approximating random vectors Y_i have purely Gaussian distribution, then $U_i \equiv 0$, $Y_i \sim \mathcal{N}(0, \operatorname{Var}(X_i))$, $\varSigma_{\boldsymbol{z}} = \operatorname{Var}(X_i)$, and $C_{\boldsymbol{z}} = \|\{\operatorname{Var}(X_i)\}^{-1/2}\|$. Furthermore, if $K=3$ and Y_i are Gaussian, then the bound in Theorem 2.1 is similar to the classical multivariate Berry-Esseen inequality by [4]. If $K>3$ and Y_i are Gaussian, the term $\|X_1\|$ enters the bound above with a better power, than in the classical case where $K=3$.

Remark 2.2 (Accuracy of the approximation)*.* As it is mentioned in Remark 2.1, in case when $K=3$ and $Y_i \sim \mathcal{N}(0, \mathbf{I}_p)$, Theorem 2.1 is almost identical to the results by [4], which are the best known in dependence on p. Moreover, [32] proved the lower bound $\varDelta_n \ge C\mathbf{E}(\|X_1\|^3)n^{-1/2}$ for the class of all convex sets in $\mathbb{R}^p$. We conjecture that the optimal error term in Theorem 2.1 is $\le M C_{\mathscr{B}} C_\phi C_{\boldsymbol{z}}^K \mathbf{E}(\|X_1\|^K + \|Y_1\|^K) n^{-(K-2)/2}$; this complies with the results of [21]. We leave this improvement of the accuracy for the future work.

Remark 2.3 (Dependence on $C_{\boldsymbol{z}}$)*.* The approximation bound in Theorem 2.1 depends on $C_{\boldsymbol{z}} = \|\varSigma_{\boldsymbol{z}}^{-1/2}\|$, where $\varSigma_{\boldsymbol{z}}$ is a covariance matrix of the Gaussian part Z_i of the approximating distribution Y_i . In Lemma 2.1 below we show that if X_i are continuously distributed, then there exist random vectors U_i such that $\varSigma_{\boldsymbol{z}}$ is positive definite. Therefore, it holds $0 < \varSigma_{\boldsymbol{z}} \le \operatorname{Var}(X_i)$ and $\|\{\operatorname{Var}(X_i)\}^{-1/2}\| \le C_{\boldsymbol{z}} < \infty$. Bounding the value $C_{\boldsymbol{z}}$ from above is important for better understanding of the considered approximation, we leave this problem for the future work as well.

Now let us consider the case when the random summands X_i are not necessarily identically distributed. Denote $\overline{\varSigma}_{\boldsymbol{z}} \stackrel{\text{def}}{=} n^{-1}\sum_{i=1}^n \varSigma_{\boldsymbol{z},i}$ and $\overline{C}_{\boldsymbol{z}} \stackrel{\text{def}}{=} \|\overline{\varSigma}_{\boldsymbol{z}}^{-1/2}\|$. Constant $C_{\phi,2}$ is similar to $\widetilde{C}_\phi, \widetilde{C}_{\phi,1}$ in the previous theorem. $C_{\phi,2}$ enters an upper bound on supremum norm of $\widetilde{\varphi}^{(K)}(x)$, where $\widetilde{\varphi}$ is K times continuously differentiable approximation of the indicator function $1\!\!1\{x \in B_{r+1} \setminus B_r\}$ for $B_{r+1}, B_r \in \mathscr{B}$ (see Section A.3).

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 $\widetilde{S}_n$ defined in (2.3)*

$$\sup_{B\in\mathscr{B}}\left|\mathbf{P}\left(S_n\in B\right)-\mathbf{P}\bigl(\widetilde{S}_n\in B\bigr)\right|$$
$$\leq M_K\left\{C_{\phi,2}\widetilde{C}_{\mathscr{B}}^K\right\}^{1/(K+1)}\left\{\frac{\overline{C}_{\boldsymbol{z}}^K\sum_{i=1}^n\mathbf{E}\left(\|X_i\|^K+\|Y_i\|^K\right)}{n^{K/2}}\right\}^{1/(K+1)},$$

where $M_K\stackrel{\text{def}}{=}2\left\{(K-1)!\right\}^{-1/(K+1)}<1.7$ for $K\geq 3$, and $\widetilde{C}_{\mathscr{B}}$ is the isoperimetric constant introduced in (2.4).

Corollary 2.1 below follows directly from the previous theorems and triangle inequality.

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 $X_1',\dots,X_n'\in\mathbb{R}^p$, that are independent of $\{X_i\}_{i=1}^n$, $\{Y_i\}_{i=1}^n$, and such that $\forall i=1,\dots,n$*

$$\begin{gathered}\mathbf{E}\left(\|X_i'\|^K\right)<\infty,\\ \mathbf{E}\bigl(X_i^k\bigr)=\mathbf{E}\bigl(X_i'^k\bigr)\ \forall k=1,\dots,K-1.\end{gathered}\tag{2.5}$$

Let also

$$S_n'\stackrel{\text{def}}{=}\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i',\quad \varDelta_n'\stackrel{\text{def}}{=}\sup_{B\in\mathscr{B}}\left|\mathbf{P}\left(S_n\in B\right)-\mathbf{P}\bigl(S_n'\in B\bigr)\right|.$$

1. *If conditions of Theorem 2.1 are fulfilled and $\{X_i'\}_{i=1}^n$ are i.i.d., then*

$$\varDelta_n'\leq MC_{\mathscr{B}}C_\phi\frac{\left\{C_{\boldsymbol{z}}^K\mathbf{E}\left(\|X_1\|^K+\|Y_1\|^K\right)\right\}^{1/(K-2)}}{n^{1/2}}+MC_{\mathscr{B}}C_\phi\frac{\left\{C_{\boldsymbol{z}}^K\mathbf{E}\left(\|X_1'\|^K+\|Y_1\|^K\right)\right\}^{1/(K-2)}}{n^{1/2}}.$$

2. *If conditions of Theorem 2.2 are fulfilled and $\{X_i'\}_{i=1}^n$ are not necessarily identically distributed, then*

$$\varDelta_n'\leq M_K\left\{C_{\phi,2}\widetilde{C}_{\mathscr{B}}^K\right\}^{1/(K+1)}\left\{\frac{\overline{C}_{\boldsymbol{z}}^K\sum_{i=1}^n\mathbf{E}\left(\|X_i\|^K+\|Y_i\|^K\right)}{n^{K/2}}\right\}^{1/(K+1)}+M_K\left\{C_{\phi,2}\widetilde{C}_{\mathscr{B}}^K\right\}^{1/(K+1)}\left\{\frac{\overline{C}_{\boldsymbol{z}}^K\sum_{i=1}^n\mathbf{E}\left(\|X_i'\|^K+\|Y_i\|^K\right)}{n^{K/2}}\right\}^{1/(K+1)}.$$

In the following theorem we consider values of a function $f : \mathbb{R}^p \mapsto \mathbb{R}$, which is at least K times continuously differentiable on $\mathbb{R}^p$. The statement shows how well the value $\mathbf{E} f(S_n)$ can be approximated with $\mathbf{E} f(S_n')$ upon condition (2.5).

Theorem 2.3. *Consider independent random vectors* $X_1, X_1', \ldots, X_n, X_n' \in \mathbb{R}^p$ *such that* $\forall i = 1, \ldots, n$ $\mathbf{E} X_i = \mathbf{E} X_i' = 0$, $\operatorname{Var}(X_i), \operatorname{Var}(X_i') \in S_p^+$, $\mathbf{E}(\|X_i\|^K), \mathbf{E}(\|X_i'\|^K) < \infty$, *and*

$$\mathbf{E}\big(X_i^k\big) = \mathbf{E}\big({X_i'}^k\big) \ \forall k = 1, \ldots, K-1. \tag{2.6}$$

Let a function $f : \mathbb{R}^p \mapsto \mathbb{R}$ *be at least* K *times continuously differentiable; constant* $C_f > 0$ *is such that* $\sup_x |f^{(K)}(x) h^K| \leq C_f \|h\|^K$ *for all* $h \in \mathbb{R}^p$. *Then it holds for* $S_n' \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i'$

$$\Big|\mathbf{E} f\left(S_n\right) - \mathbf{E} f(S_n')\Big| \leq \sum\nolimits_{i=1}^n C_f \frac{\mathbf{E}\left(\|X_i\|^K + \|X_i'\|^K\right)}{K! n^{K/2}}.$$

If $X_1, \ldots, X_n$ *are identically distributed, and so are* $X_1', \ldots, X_n'$, *then*

$$\Big|\mathbf{E} f\left(S_n\right) - \mathbf{E} f(S_n')\Big| \leq C_f \frac{\mathbf{E}\left(\|X_1\|^K + \|X_1'\|^K\right)}{K! n^{(K-2)/2}}.$$

Lemma 2.1 (Existence of the approximating distribution)**.** *Let a random vector* $X \in \mathbb{R}^p$ *have continuous probability distribution, such that* $\mathbf{E} X = 0$, $\operatorname{Var} X \in S_p^+$, *and* $\mathbf{E}(\|X\|^K) < \infty$ *for some integer* $K \geq 2$. *Then there exists a random vector* $Y \stackrel{\text{def}}{=} Z + U$, *where* $Z, U \in \mathbb{R}^p$ *are independent,* $Z \sim \mathcal{N}(0, \Sigma_{\boldsymbol{z}})$ *for some* $\Sigma_{\boldsymbol{z}} \in S_p^+$, *and* $\mathbf{E}(X^k) = \mathbf{E}(Y^k)$ *for all* $k = 0, \ldots, K$.

Proof of Lemma 2.1. Denote

$$m_k \stackrel{\text{def}}{=} \mathbf{E}(X^k), \quad u_k \stackrel{\text{def}}{=} \mathbf{E}(U^k) \text{ for } k = 0, 1, 2, \ldots \tag{2.7}$$

Conditioning on U leads to $\mathscr{L}(Y \,\big|\, U) = \mathcal{N}(U, \Sigma_{\boldsymbol{z}})$ and to the following system of linear equations:

$$\begin{aligned}
m_0 &= \mathbf{E}(Z+U)^0 &&= u_0, \\
m_1 &= \mathbf{E}(Z+U) &&= u_1, \\
m_2 &= \mathbf{E}(Z+U)^2 &&= u_2 + \Sigma_{\boldsymbol{z}}, \\
m_3 &= \mathbf{E}(Z+U)^3 &&= u_3, \\
&\vdots \\
m_K &= \mathbf{E}(Z+U)^K &&= K! \sum\nolimits_{l=0}^{[K/2]} S_{p\mathbf{I}_K} \frac{u_{K-2l} \otimes \operatorname{vec}(\Sigma_{\boldsymbol{z}})^l}{l!(K-2l)! 2^l},
\end{aligned}$$

where $S_{p\mathbf{I}_K}$ 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 of [20]. The solution $\{u_k(\varSigma_{\boldsymbol{z}})\}_{k=0}^{K}$ of this system depends on $\varSigma_{\boldsymbol{z}}$ continuously. Moreover,

$$\text{if } \varSigma_{\boldsymbol{z}} = 0\,, \text{ then } u_k(\varSigma_{\boldsymbol{z}}) = m_k \ \ \forall\, k = 0, \ldots, K. \tag{2.8}$$

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

> given a p-dimensional real multisequence $\{u_k(\varSigma_{\boldsymbol{z}})\}_{k=0}^{K}$, does there exist a positive Borel measure μ on $\mathbb{R}^p$ s.t. $\int_{\mathbb{R}^p} x^k d\mu(x) = u_k(\varSigma_{\boldsymbol{z}})\ \forall\, k = 0, \ldots, K$?

The work of [12] 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 space of polynomials of degree $\le K$ with real coefficients. A polynomial $p = p(x) = \sum_{|i|\le K} a_i x^i \in \mathcal{P}_K$ is non-negative, if $p \ge 0$ for all $x \in \mathbb{R}^p$. Here $i \stackrel{\text{def}}{=} (i_1, \ldots, i_p) \in \mathbb{N}_0^p$ denotes multi-index, $|i| = \sum_{j=1}^{p} i_j$, and $x^i \stackrel{\text{def}}{=} x_1^{i_1} \ldots x_p^{i_p}$. For a multisequence $\{u_i\}_{|i|\le K}$ the Riesz functional $L: \mathcal{P}_K \mapsto \mathbb{R}$ is defined as $L(\sum_{|i|\le K} a_i x^i) \stackrel{\text{def}}{=} \sum_{|i|\le K} a_i u_i$. If the Hamburger moment problem is soluble, we can write

$$L(p) = \sum\nolimits_{|i|\le K} a_i u_i = \int_{\mathbb{R}^p} p(x) d\mu(x). \tag{2.9}$$

[12] showed that a multisequence $\{u_i\}_{|i|\le K}$ solves the multivariate Hamburger truncated moment problem iff there exists an extension $\{\widetilde{u}_i\}_{|i|\le K+2}$ of $\{u_i\}_{|i|\le K}$ (i.e. $\widetilde{u}_i = u_i$ for all $|i| \le K$), such that for the corresponding Riesz functional $\widetilde{L}(\sum a_i x^i) \stackrel{\text{def}}{=} \sum_{|i|\le K+2} a_i \widetilde{u}_i$ it holds:

$$\text{if } p \in \mathcal{P}_{K+2} \text{ and } p \text{ is non-negative, then } \widetilde{L}(p) \ge 0. \tag{2.10}$$

Let us consider the moment sequence $\{m_k\}_{k=0}^{K}$ defined in (2.7). By the theorem of [12] there exists an extension $\{\widetilde{m}_k\}_{k=0}^{K+2}$, s.t. its corresponding Riesz functional $\widetilde{L}_m(\sum_{|i|\le K+2} a_i x^i) \stackrel{\text{def}}{=} \sum a_i \widetilde{m}_i$ satisfies (2.10). The extension $\{\widetilde{m}_k\}_{k=0}^{K+2}$ leads to the extended sequence $\{\widetilde{u}_k(\varSigma_{\boldsymbol{z}})\}_{k=0}^{K+2}$. Property (2.8), continuity of the solutions $\{\widetilde{u}_k(\varSigma_{\boldsymbol{z}})\}_{k=0}^{K+2}$ w.r.t. $\varSigma_{\boldsymbol{z}}$, and (2.9) imply that there exists some $\varSigma_{\boldsymbol{z}} \in S_p^+$ s.t. the corresponding Riesz functional $\widetilde{L}_u(p) \stackrel{\text{def}}{=} \sum_{|i|\le K+2} a_i \widetilde{u}_i(\varSigma_{\boldsymbol{z}}) > 0$ for all $p = \sum_{|i|\le K+2} a_i x^i \in \mathcal{P}_{K+2}$ such that $p > 0$. Due to continuity of $\mathscr{L}(X)$, $\mathbf{P}(X = x_0) = 0$ for any $x_0 \in \mathbb{R}^p$ such that $p(x_0) = \sum_{|i|\le K+2} a_i x_0^i = 0$, which finishes the proof. $\square$

3. Validity and accuracy of the weighted bootstrap

3.1. Weighted bootstrap for $\|S_n\|$

Consider independent random vectors $X_1, \ldots, X_n \in \mathbb{R}^p$ with $\mathbf{E}X_i = 0$, $\operatorname{Var}(X_i) \in S_p^+$, and $\mathbf{E}(\|X_i\|^4) < \infty$. The bootstrap random weights $\varepsilon_1, \ldots, \varepsilon_n$, are taken as in (1.1). Below are some examples of such random weights (here $z_i \sim \mathcal{N}(0,1)$, independent of e_i, c_i, b_i given below):

$$\begin{gathered}
\sqrt{1-2^{-2/3}}z_i + 2^{-1/3}(e_i - 1), \text{ for } e_i \sim exp(1);\\
\frac{1}{\sqrt{2}}z_i + \frac{1}{2}(c_i - 1), \text{ for } c_i \sim \chi_1^2;\\
\sqrt{1-3^{-2/3}}z_i + \frac{2}{3^{1/3}}(b_i - 0.5), \text{ for } b_i \sim Bernoulli(0.5).
\end{gathered}$$

More examples of the bootstrap weights satisfying (1.1) can be found in the works of [26] and [30].

The bootstrap approximation of the sum S_n is its weighted version S_n° defined in (1.2). The probability distribution of S_n° is taken conditioned on $\{X_i\}_{i=1}^n$. Theorems 3.1 and 3.2 show that the bootstrap quantile function $Q^{\circ}(\alpha)$ of $\|S_n^{\circ}\|$ defined in (1.3) is a rather good approximation of the true unknown quantile function $Q(\alpha)$:

$$Q(\alpha) \stackrel{\text{def}}{=} \inf\{t \in \mathbb{R} : \mathbf{P}(\|S_n\| > t) \leq \alpha\} \text{ for } \alpha \in (0,1). \tag{3.1}$$

In the statements in Section 3, including the theorems presented below, we use notation from the previous Section 2, in particular, values $M, \widetilde{C}_{\mathscr{B}}, C_\phi, C_{\boldsymbol{z}}$ were introduced before Theorem 2.1.

Theorem 3.1. *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) for $K = 4$. It holds for $\alpha \in (0,1)$*

$$\left|\mathbf{P}\big(\|S_n\| > Q^{\circ}(\alpha)\big) - \alpha\right| \leq \varDelta_1 + \varDelta_{1,c},$$

where

$$\begin{aligned}
\varDelta_1 &\stackrel{\text{def}}{=} 2MC_{\mathscr{B}}C_\phi C_{\boldsymbol{z}}^2 \frac{\left\{\mathbf{E}\left(\|X_1\|^4\right)\mathbf{E}(\varepsilon_1^4) + \mathbf{E}\left(\|Y_1\|^4\right)\right\}^{1/2}}{n^{1/2}},\\
\varDelta_{1,c} &\stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } \|S_n\| \text{ and } \|S_n^{\circ}\| \text{ are continuously distributed;}\\ \varDelta_1(1 + 2\widetilde{C}_{\mathscr{B}}C_{\boldsymbol{z}}), & \text{otherwise.}\end{cases}
\end{aligned}$$

Theorem 3.2. *Let the conditions of Theorem 3.1 be fulfilled, except that $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ are not assumed to be identically distributed. It holds for $\alpha \in (0,1)$*

$$\left|\mathbf{P}\big(\|S_n\| > Q^{\circ}(\alpha)\big) - \alpha\right| \leq \varDelta_2 + \varDelta_{2,c},$$

where

$$\Delta_2 \stackrel{\text{def}}{=} 2.8 \left\{ C_{\phi,2} \widetilde{C}_{\mathscr{B}}^4 \right\}^{1/5} \left\{ \frac{\overline{C}_{\boldsymbol{z}}^4 \sum_{i=1}^n \left[\mathbf{E}\left(\|X_i\|^4 \right) \mathbf{E}(\varepsilon_1^4) + \mathbf{E}\left(\|Y_i\|^4 \right) \right]}{n^2} \right\}^{1/5},$$

$$\Delta_{2,c} \stackrel{\text{def}}{=} \begin{cases} 0, & \textit{if } \|S_n\| \textit{ and } \|S_n^{\circ}\| \textit{ are continuously distributed;} \\ \Delta_2(1 + 2\widetilde{C}_{\mathscr{B}} \overline{C}_{\boldsymbol{z}}), & \textit{otherwise.} \end{cases}$$

3.2. *Weighted bootstrap for linear regression model*

Let $\boldsymbol{y} \stackrel{\text{def}}{=} (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$ be a data sample with

$$y_i = \Psi_i^\top \theta^* + \epsilon_i, \tag{3.2}$$

$\theta^* \in \mathbb{R}^p$ is unknown, $\Psi_i \in \mathbb{R}^p$ are deterministic regressors such that the matrix $\Psi\Psi^\top$ is invertible, where $\Psi \stackrel{\text{def}}{=} (\Psi_1, \ldots, \Psi_n) \in \mathbb{R}^{p \times n}$. The random errors $\epsilon_1, \ldots, \epsilon_n$ are independent, $\mathbf{E}\epsilon_i = 0$, $\mathbf{E}(\epsilon_i^4) < \infty$, for all $i = 1, \ldots, n$, and $\operatorname{Var}(\epsilon_i) > 0$ are unknown. Let $\boldsymbol{\epsilon} \stackrel{\text{def}}{=} (\epsilon_1, \ldots, \epsilon_n)^\top$. The least squares estimate of the parameter θ^* reads as

$$\widetilde{\theta} = (\Psi\Psi^\top)^{-1}\Psi\boldsymbol{y} = \theta^* + (\Psi\Psi^\top)^{-1}\Psi\boldsymbol{\epsilon}.$$

Consider the normalized quadratic loss

$$T^2 \stackrel{\text{def}}{=} \|(\Psi\Psi^\top)^{1/2}(\widetilde{\theta} - \theta^*)\|^2 = \|(\Psi\Psi^\top)^{-1/2} \textstyle\sum_{i=1}^n \Psi_i \epsilon_i\|^2.$$

The weighted bootstrap estimate of T can be written as

$$T^{\circ} = \|(\Psi\Psi^\top)^{-1/2} \textstyle\sum_{i=1}^n \Psi_i \epsilon_i \varepsilon_i\|, \tag{3.3}$$

where the random i.i.d. bootstrap weights $\varepsilon_1, \ldots, \varepsilon_n$ are independent of $\{\epsilon_i\}_{i=1}^n$ and meet conditions (1.1). For the linear regression model the weighted bootstrap estimate given in (3.3) coincides with the wild bootstrap estimate introduced by [50] (see also [26, 18, 30]).

Denote similarly to $Q^{\circ}(\alpha)$ in (1.3)

$$Q_T^{\circ}(\alpha) \stackrel{\text{def}}{=} \inf\left\{ t \in \mathbb{R} : \mathbf{P}^{\circ}\left(T^{\circ} > t \right) \le \alpha \right\},$$

where $\alpha \in (0,1)$, and $\mathbf{P}^{\circ}(\cdot) \stackrel{\text{def}}{=} \mathbf{P}(\cdot \,|\, \epsilon_1, \ldots, \epsilon_n)$. Let $Y_1, \ldots, Y_n$ be independent approximating vectors meeting conditions (2.1) and (2.2) for $K = 4$ and for $X_i \coloneqq (\Psi\Psi^\top)^{-1/2}\Psi_i\epsilon_i$. Applying Theorem 3.2 to T and $Q_T^{\circ}(\alpha)$, we obtain the following confidence statement

Theorem 3.3. *Consider the linear regression model (3.2) with random errors $\{\epsilon_i\}$ and approximating vectors $\{Y_i\}_{i=1}^n$ described above. It holds*

$$\left| \mathbf{P}\left(\|(\Psi\Psi^\top)^{1/2}(\widetilde{\theta} - \theta^*)\| > Q_T^{\circ}(\alpha) \right) - \alpha \right| \le \Delta_{2,T} + \Delta_{2,T,c},$$

where

$$\Delta_{2,T} \stackrel{\text{def}}{=} 2.8 \left\{C_{\phi,2}\widetilde{C}^4_{\mathscr{B}}\right\}^{1/5} \left\{ \frac{\overline{C}^4_{\boldsymbol{z}} \sum_{i=1}^n \left[\mathbf{E}\left(\|(\Psi\Psi^\top)^{-1/2}\Psi_i\|^4\right)\mathbf{E}(\epsilon_i^4)\mathbf{E}(\varepsilon_1^4)\right.}{n^2} \right.$$
$$\left. + \frac{\left.\mathbf{E}\left(\|Y_i\|^4\right)\right]}{n^2} \right\}^{1/5},$$

$$\Delta_{2,T,c} \stackrel{\text{def}}{=} \begin{cases} 0, & \textit{if } T \textit{ and } T^{\circ} \textit{ are continuously distributed;} \\ \Delta_{2,T}(1+2\widetilde{C}_{\mathscr{B}}\overline{C}_{\boldsymbol{z}}), & \textit{otherwise.} \end{cases}$$

3.3. Weighted bootstrap for log-likelihood ratio statistics

Here we consider a weighed bootstrap procedure for estimation of quantiles of log-likelihood ratio statistics. Before describing the procedure and formulating the 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 $\{\mathbf{P}_\theta\} \stackrel{\text{def}}{=} \{\mathbf{P}_\theta \ll \mu_0,\, \theta\in\Theta\subset\mathbb{R}^p\}$, 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 $\boldsymbol{y}$:

$$L(\theta) = L(\theta,\boldsymbol{y}) \stackrel{\text{def}}{=} \log\left(\frac{d\mathbf{P}_\theta}{d\mu_0}(\boldsymbol{y})\right).$$

The target parameter θ^* is defined by projecting the true probability distribution $\mathbf{P}$ on the parametric family $\{\mathbf{P}_\theta\}$, using Kullback-Leibler divergence:

$$\theta^* \stackrel{\text{def}}{=} \operatorname{argmin}_{\theta\in\Theta}\operatorname{KL}(\mathbf{P},\mathbf{P}_\theta) = \operatorname{argmax}_{\theta\in\Theta}\mathbf{E}L(\theta).$$

The (quasi) maximum likelihood estimate (MLE) is defined as

$$\widetilde{\theta} \stackrel{\text{def}}{=} \operatorname{argmax}_{\theta\in\Theta} L(\theta).$$

Let $Q_L(\alpha)$ denote the upper quantile function of square root of the two times log-likelihood ratio statistic:

$$Q_L(\alpha) \stackrel{\text{def}}{=} \inf\left\{t\ge 0 : \mathbf{P}\left(L(\widetilde{\theta}) - L(\theta) > t^2/2\right) \le \alpha\right\}.$$

$Q_L(\alpha)$ is a critical value of the likelihood-based confidence set $\mathscr{E}(\alpha)$:

$$\mathscr{E}(t) \stackrel{\text{def}}{=} \left\{\theta : L(\widetilde{\theta}) - L(\theta) \le t^2/2\right\}, \tag{3.4}$$
$$\mathbf{P}\left\{\theta^* \in \mathscr{E}(Q_L(\alpha))\right\} > 1-\alpha.$$

Distribution of $L(\widetilde{\theta}) - L(\theta^*)$ depends on the unknown parameter θ^* and $\mathbf{P}$, hence, in general, quantiles of $L(\widetilde{\theta}) - L(\theta^*)$ are also unknown.

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

$$u_i \stackrel{\text{def}}{=} \varepsilon_i + 1, \text{ for } \varepsilon_i \text{ defined in (1.1), independent of } \boldsymbol{y}.$$

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

$$L^{\circ}(\theta) \stackrel{\text{def}}{=} \sum\nolimits_{i=1}^{n} \log\left(\frac{d\mathbf{P}_{\theta}}{d\mu_0}(y_i)\right) u_i.$$

Let $\mathbf{P}^{\circ}(\cdot) \stackrel{\text{def}}{=} \mathbf{P}(\cdot \,|\, \boldsymbol{y})$ and $\mathbf{E}^{\circ}(\cdot) \stackrel{\text{def}}{=} \mathbf{E}(\cdot \,|\, \boldsymbol{y})$. It holds $\mathbf{E}^{\circ} L^{\circ}(\theta) = L(\theta)$, therefore,

$$\widetilde{\theta} = \operatorname{argmax}_{\theta \in \Theta} L(\theta) = \operatorname{argmax}_{\theta \in \Theta} \mathbf{E}^{\circ} L^{\circ}(\theta),$$

and the MLE $\widetilde{\theta}$ can be considered as a bootstrap analogue of the unknown target parameter θ^*. The bootstrap likelihood ratio statistic is defined as

$$L^{\circ}(\widetilde{\theta}^{\circ}) - L^{\circ}(\widetilde{\theta}) \stackrel{\text{def}}{=} \sup\nolimits_{\theta \in \Theta} L^{\circ}(\theta) - L^{\circ}(\widetilde{\theta}).$$

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

$$Q_L^{\circ}(\alpha) \stackrel{\text{def}}{=} \inf\left\{t \geq 0 : \mathbf{P}^{\circ}\left(L^{\circ}(\widetilde{\theta}^{\circ}) - L^{\circ}(\theta) > t^2/2\right) \leq \alpha\right\}. \tag{3.5}$$

Theorem 3.4 below provides a two-sided bound on the coverage error of the likelihood confidence set (3.4) based on the bootstrap quantile $Q_L^{\circ}(\alpha)$. Let us introduce some additional notation before stating the theorem. Denote $\ell_i(\theta) \stackrel{\text{def}}{=} \log\left(\frac{d\mathbf{P}_{\theta}}{d\mu_0}(y_i)\right)$, $d_0^2 \stackrel{\text{def}}{=} -\mathbf{E}\ell_1''(\theta^*)$, here $\ell_i'(\theta) \stackrel{\text{def}}{=} \nabla_{\theta}\ell_i(\theta)$. Take $X_i \coloneqq d_0^{-1}\ell_i'(\theta^*)$. By previous definitions, such defined $\{X_i\}_{i=1}^n$ are i.i.d with zero mean. Moreover, if conditions from Section B.1 are fulfilled, then $\mathbf{E}(\|X_i\|^4) < \infty$. Let $Y_1, \dots, Y_n$ be i.i.d. vectors meeting conditions (2.1) and (2.2) for $K = 4$, and $C_{\boldsymbol{z},L} \stackrel{\text{def}}{=} \|\{\operatorname{Var}(Z_i)\}^{-1/2}\|$. Now we are ready to formulate the following

Theorem 3.4. *If the conditions from Section B.1 are fulfilled, then it holds with probability* $\geq 1 - 10\mathrm{e}^{-x}$

$$\left|\mathbf{P}\left\{\theta^* \notin \mathscr{E}(Q_L^{\circ}(\alpha))\right\} - \alpha\right| \leq \varDelta_L + \varDelta_{L,c},$$

where

$$\varDelta_L \leq 2MC_{\mathscr{B}}C_{\phi}C_{\boldsymbol{z},L}^2 \frac{\left\{\mathbf{E}\left(\|d_0^{-1}\ell_i'(\theta^*)\|^4\right)\mathbf{E}(\varepsilon_1^4) + \mathbf{E}\left(\|Y_1\|^4\right)\right\}^{1/2}}{n^{1/2}} \tag{3.6}$$
$$+ C_{\mathscr{B}}C_{\boldsymbol{z},L}C(p + x)n^{-1/2},$$

here C is a generic constant ≥ 0; a more detailed definition of the error term $\varDelta_L$ is given in (B.8)*, Section B;*

$$\varDelta_{L,c} \stackrel{\text{def}}{=} \begin{cases} 0, & \textit{if } L(\widetilde{\theta}) - L(\theta^*) \textit{ and } L^{\circ}(\widetilde{\theta}^{\circ}) - L^{\circ}(\widetilde{\theta}) \\ & \textit{are continuously distributed;} \\ 2\varDelta_L, & \textit{otherwise.} \end{cases}$$

Remark 3.1. The second term in the bound (3.6) comes from Wilks-type approximations for the likelihood ratios $L(\widetilde{\theta}) - L(\theta^*)$ and $L^{\circ}(\widetilde{\theta}^{\circ}) - L^{\circ}(\widetilde{\theta})$ (see the proof in Section B.1 for more detail); the first term in (3.6) comes from Berry-Esseen type inequality for i.i.d summands (Theorem 2.1), and is similar to the error term in Theorem 3.1. The imposed conditions in Section B.1 include sub-exponential tail behavior of ε_i and $d_0^{-1}\ell_i'(\theta^*)$, therefore, the first term (3.6) is bounded from above with $Cpn^{-1/2}$ on a set of exponentially large probability. Thus, in Theorem 3.4 both Wilks-type bound and Berry-Esseen type inequality yield similar impacts of p and n in the error of approximation $\varDelta_L$.

4. Numerical experiments

This section presents results of simulation studies, illustrating accuracy of the considered Berry-Esseen bounds and weighted bootstrap procedure.

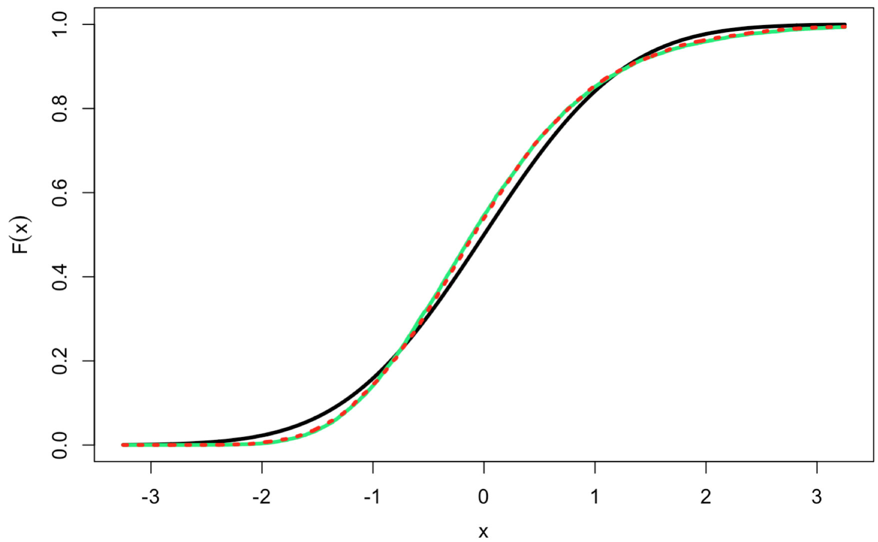
4.1. Berry-Esseen inequality

Figure 1 shows c.d.f.-s of S_n, $\widetilde{S}_n$ and $\mathcal{N}(0,1)$ for sample size $n = 50$, dimension $p = 1$ and number $K - 1 = 3$ of equal moments of S_n and $\widetilde{S}_n$. Similarly Figure 2 shows c.d.f.-s of $\|S_n\|^2$, $\|\widetilde{S}_n\|^2$ 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 one.

4.2. Bootstrap

Here we examine accuracy of the weighted bootstrap for $\|S_n\|$ (described in Section 3) by computing coverage probabilities using bootstrap quantiles $Q^{\circ}(\alpha)$. All the results are collected in Table 1. Columns n, p, $\mathscr{L}(\varepsilon_i)$, $\mathscr{L}(X_{i,j})$ show the sample size, dimension, distribution of the bootstrap weights ε_i, and 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 $1 - \alpha$ 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 $\{\|S_n\| \leq Q^{\circ}(\alpha)\}$, 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 two types of the bootstrap weights: first one $\varepsilon_i = z_i + u_i$, with $u_i \sim (Bernoulli(b) - b)\sigma_u$, $b = 0.276, \sigma_u \approx$

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



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 (1.1). The second type is $\varepsilon_i \sim \mathcal{N}(0,1)$, in this case $\mathbf{E}(\varepsilon_i^3) \neq 1$, and an approximation accuracy corresponds to 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 $\|S_n\|$. Table 1 confirms this property for most of the computed coverage probabilities, which agrees with the theoretical results.

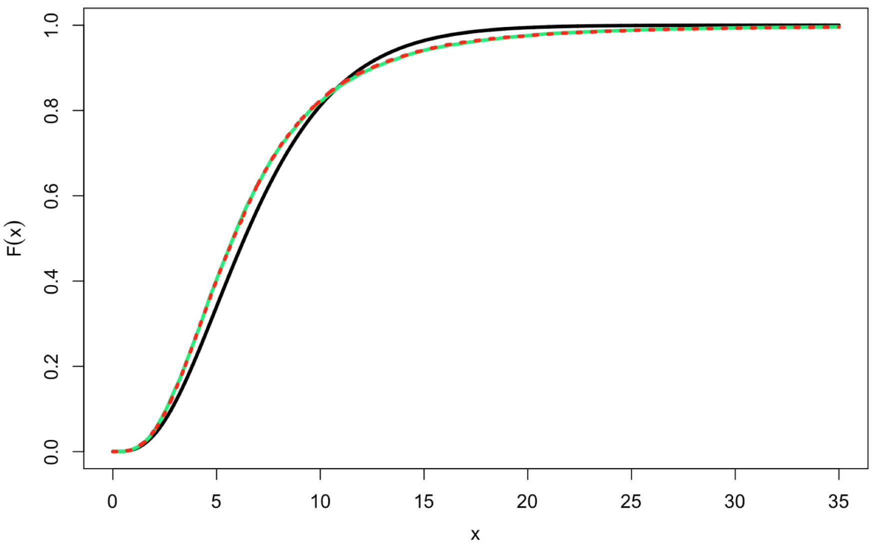
Appendix A: Proof of the Non-Classical Berry-Esseen inequality

In this section we prove Theorems 2.1, 2.2, and 2.3. In Section A.1 we provide some bounds on Gaussian surface area of Euclidean balls and ellipsiods in $\mathbb{R}^p$, these bounds are used in the proofs of the theorems from Sections 2 and 3.

A.1. Gaussian surface area of ellipsoids in $\mathbb{R}^p$

In this section we collect bounds on Gaussian surface area (GSA) of Euclidean balls and ellipsoids in $\mathbb{R}^p$. These bounds are required for the proofs of Theorems 2.1 and 2.2. The following lemma shows that the GSA of any ball or an ellipsoid in $\mathbb{R}^p$ is bounded with a constant independent of dimention p. This properties are well known (see the works [39, 2, 24, 23]), we give here the proof for the sake of the text's completeness.

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



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

Table 1
Coverage probabilities $\mathbf{P}\big(\|S_n\| \le Q^\circ(\alpha)\big)$

				Confidence levels							
n	p	$\mathscr{L}(X_{i,j})$	$\mathscr{L}(\varepsilon_i)$	**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
		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
		$\ln \mathcal{N}^*(2.5)$	$\mathscr{L}(z_i + u_i)$	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
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
		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
		$\ln \mathcal{N}^*(2.5)$	$\mathscr{L}(z_i + u_i)$	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
50	5	$\chi_1^2 - 1$	$\mathscr{L}(z_i + u_i)$	0.985	0.963	0.906	0.850	0.794	0.694	0.587	0.483
			$\mathcal{N}(0,1)$	0.988	0.970	0.917	0.861	0.801	0.692	0.578	0.467
		Pareto*	$\mathscr{L}(z_i + u_i)$	0.981	0.959	0.906	0.849	0.790	0.681	0.571	0.463
			$\mathcal{N}(0,1)$	0.984	0.966	0.919	0.860	0.800	0.678	0.555	0.435
		$\ln \mathcal{N}^*(1.5)$	$\mathscr{L}(z_i + u_i)$	0.977	0.955	0.898	0.838	0.778	0.649	0.532	0.422
			$\mathcal{N}(0,1)$	0.981	0.964	0.914	0.858	0.794	0.654	0.515	0.388

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

Lemma A.1 (GSA of Euclidean balls and ellipsoids)**.** *There exists a generic constant* $\widetilde{C}_{\mathscr{B}} > 0$ *such that for all* $a \in \mathbb{R}^p$ *and* $r \geq 0$

$$\int_{\|x-a\|=r} \phi(x)dx \leq \widetilde{C}_{\mathscr{B}}, \tag{A.1}$$

where $\phi(x) \stackrel{\text{def}}{=} (2\pi)^{-p/2}\mathrm{e}^{-\|x\|^2/2}$ *is the standard normal density. Moreover, for any matrix* $\varSigma \in S_p^+$ *and* $\phi_{\varSigma}(x) \stackrel{\text{def}}{=} \{(2\pi)^p \det(\varSigma)\}^{-1/2}\, \mathrm{e}^{-\|\varSigma^{-1/2}x\|^2/2}$*, it holds*

$$\int_{\|x\|=r} \phi_{\varSigma}(x+a)dx \leq \frac{\widetilde{C}_{\mathscr{B}}}{\sqrt{\lambda_{\min}}},$$

where $\lambda_{\min} > 0$ *is the smallest eigenvalue of the covariance matrix* $\varSigma$*.*

Proof of Lemma A.1. At first let us consider the case $a = 0$.

$$\begin{aligned}\int_{\|x\|=r} \phi(x)dx &= \frac{1}{(2\pi)^{p/2}}\mathrm{e}^{-r^2/2}\int_{\|x\|=r} dx \\ &= \frac{1}{(2\pi)^{p/2}}\mathrm{e}^{-r^2/2} r^{p-1} p \frac{\pi^{p/2}}{\Gamma(p/2+1)} \qquad \text{(A.2)} \\ &\leq const. \qquad \text{(A.3)}\end{aligned}$$

Expression (A.2) is maximized in $r = \sqrt{p-1}$, which implies (A.3).

Now consider the balls with an arbitrary center $a \in \mathbb{R}^p$. Let $Z \sim \mathcal{N}(0, \mathbf{I}_p)$, the following expression corresponds to the density function of the r.v. $\|Z+a\|^2$, which follows the noncentral chi-squared distribution:

$$f(x;p,a) = \sum\nolimits_{k=0}^{\infty} \frac{\mathrm{e}^{-\|a\|^2/2}(\|a\|^2/2)^k}{k!} f(x; p+2k, 0), \tag{A.4}$$

where $f(x;p+2k,0)$ is a probability density function of the chi-squared distribution with $p+2k$ degrees of freedom. Equation (A.4) together with the bound (A.3) imply inequality (A.1).

For the case of ellipsoids with an arbitrary center $a \in \mathbb{R}^p$ we assume w.l.o.g. that the covariance matrix $\varSigma$ is diagonal. Let $Z \sim \mathcal{N}(0, \mathbf{I}_p)$. Due to the results of [36] and [19] the c.d.f. of the random variable $\|\varSigma^{1/2}Z + a\|^2$ (which is a weighted sum of independent noncentral chi-squared r.v.) reads as

$$F(x;p,a,\varSigma) = \sum\nolimits_{k=1}^{\infty} q_k F(x/\lambda_{\min}; p+2k, 0, \mathbf{I}_p), \tag{A.5}$$

where $F(x;p+2k,0,\mathbf{I}_p)$ is a c.d.f. of a chi-squared distribution with $p+2k$ degrees of freedom, c is a positive number, and $q_k = q_k(p,a,\varSigma) > 0$ are coefficients which depend on p, a and $\varSigma$, s.t. $\sum_{k=1}^{\infty} q_k = 1$. Equation (A.5) together with bound (A.3) complete the proof. □

Consider an arbitrary Euclidean ball $B = B_r(x_0) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^p : \|x - x_0\| \le r\}$. Denote for some number $\varepsilon \in \mathbb{R}$

$$B^{\varepsilon} \stackrel{\text{def}}{=} \begin{cases} B_{r+\varepsilon}(x_0), & \text{if } r+\varepsilon \ge 0, \\ \emptyset, & \text{otherwise.} \end{cases} \tag{A.6}$$

Lemma A.1 implies the following statement.

Lemma A.2. *Let $Z \sim \mathcal{N}(a, \Sigma_{\boldsymbol{z}})$ for some $a \in \mathbb{R}^p$ and $\Sigma_{\boldsymbol{z}} \in S_p^{+}$. It holds for $C_{\boldsymbol{z}} \stackrel{\text{def}}{=} \|\Sigma_{\boldsymbol{z}}^{-1/2}\|$ and for arbitrary $\varepsilon > 0$*

$$\sup_{B \in \mathscr{B}} \mathbf{P}\bigl(Z \in B^{\varepsilon} \setminus B\bigr) \le \varepsilon \widetilde{C}_{\mathscr{B}} C_{\boldsymbol{z}}.$$

A.2. Auxiliary statements and proof of Theorem 2.1

Here we extend the proof of [4] to our setting of "quasi-Gaussian" higher order approximation. The proof uses smoothing of a characteristic function of a Euclidean ball, and induction w.r.t. n. Let us introduce some necessary statements before proving Theorem 2.1. Lemma A.4 follows from a more general Lemma 2.1 in the work of [4], and we do not give its proof here.

Taylor's formula

Below is the Taylor's formula, which will be used further in the proof: for a sufficiently smooth function $f : \mathbb{R}^p \mapsto \mathbb{R}$ and $x, h \in \mathbb{R}^p$

$$\begin{aligned} f(x+h) &= f(x) + f'(x)h + \cdots + \frac{1}{s!} f^{(s)}(x) h^s \\ &\quad + \frac{1}{s!} \mathbf{E}(1-\tau)^s f^{(s+1)}(x + \tau h) h^{s+1}, \end{aligned} \tag{A.7}$$

here we use the notation $f^{(s)}(x)h^s = (h^{\top}\nabla)^s f(x)$; τ is a random variable uniformly distributed on the interval $[0,1]$, independent of all other random variables; the above formula for the remainder term follows directly from the remainder's integral form.

Lemma A.3 (Properties of a smoothing function φ)**.** *Take an arbitrary Euclidean ball B in $\mathbb{R}^p$, then for any $\varepsilon > 0$ there exists a function φ (which depends only on ε and B) such that*

$$0 \le \varphi(x) \le 1, \quad \varphi(x) = \begin{cases} 1, & x \in B; \\ 0, & x \notin B^{\varepsilon}, \end{cases} \tag{A.8}$$

and for all $x, h \in \mathbb{R}^p$, *and some constants* $\widetilde{C}_\phi, \widetilde{C}_{\phi,1} > 0$

$$\left|\varphi^{(K-2)}(x)h^{K-2}\right| \le \frac{\widetilde{C}_\phi \|h\|^{K-2}}{(\varepsilon/2)^{K-2}} \, \mathbb{I}\{x \in B^\varepsilon \setminus B\}, \text{ and} \tag{A.9}$$

$$\begin{aligned}&\left|\left\{\varphi^{(K-2)}(x) - \varphi^{(K-2)}(y)\right\}h^{K-2}\right| \\ &\le \|x-y\| \frac{\widetilde{C}_{\phi,1}\|h\|^{K-2}}{(\varepsilon/2)^{K-1}} \left(\mathbb{I}\{x \in B^\varepsilon \setminus B\} + \mathbb{I}\{y \in B^\varepsilon \setminus B\}\right).\end{aligned} \tag{A.10}$$

Furthermore, we can choose φ *to have the form*

$$\varphi(x) = \phi(\widetilde{\rho}(x)/\widetilde{\varepsilon}), \tag{A.11}$$

where $\phi : \mathbb{R} \mapsto [0,1]$ *is* $K-1$ *times continuously differentiable non-negative non-increasing function such that*

$$\int_{\mathbb{R}} |\phi'(t)| \, dt = 1, \tag{A.12}$$

function $\widetilde{\rho}(x) : \mathbb{R}^p \mapsto \mathbb{R}$ *and number* $\widetilde{\varepsilon} > 0$ *are such that*

$$\begin{gathered}\widetilde{\rho}(x) = 0 \text{ for } x \in B, \ \widetilde{\rho}(x) > 0 \text{ for } x \notin B, \\ \widetilde{\rho}(x)/\widetilde{\varepsilon} \le 1 \text{ for } x \in B^\varepsilon, \ \widetilde{\rho}(x)/\widetilde{\varepsilon} > 1 \text{ for } x \notin B^\varepsilon.\end{gathered}$$

Proof of Lemma A.3. Let $\phi(t)$ be a sufficiently smooth approximation of a step function (e.g. based on higher-order polynomials)

$$0 \le \phi(x) \le 1, \quad \phi(x) = \begin{cases} 1, & x \le 0; \\ 0, & x \ge 1, \end{cases} \tag{A.13}$$

such that the lemma's conditions are fulfilled. Let the ball $B = B_r(x_0)$ have center x_0 and radius r. Function $\widetilde{\rho}(x)$ can be taken as follows

$$\widetilde{\rho}(x) = \begin{cases} \|x - x_0\|^2 - r^2, & x \notin B; \\ 0, & x \in B. \end{cases}$$

Let also $\widetilde{\varepsilon} = \varepsilon^2 + 2r\varepsilon$, then $\widetilde{\rho}(x)/\widetilde{\varepsilon} = 0$ for $x \in B$, and

$$0 < \widetilde{\rho}(x)/\widetilde{\varepsilon} \le 1 \text{ for } x \in B^\varepsilon \setminus B, \quad \widetilde{\rho}(x)/\widetilde{\varepsilon} > 1 \text{ for } x \in \mathbb{R}^p \setminus B^\varepsilon.$$

Properties (A.9) and (A.10) follow from the representation (A.11). Indeed, definition (A.13) implies that $\phi'(x) \ne 0$ iff $x \in (0,1)$, therefore, $\phi'(x) = \phi'(x)\, \mathbb{I}\{x \in (0,1)\}$. Moreover

$$\begin{aligned}\|\varphi'(x)\| &\le \|\phi'\|_\infty \frac{2\|x - x_0\|}{\widetilde{\varepsilon}} \, \mathbb{I}\left\{\widetilde{\rho}(x)/\widetilde{\varepsilon} \in (0,1)\right\} \\ &\le \frac{2\|\phi'\|_\infty}{\varepsilon} \, \mathbb{I}\left\{x \in B^\varepsilon \setminus B\right\}.\end{aligned}$$

Property (A.9) is derived by further differentiation of $\varphi'(x)$. Inequality (A.10) is derived similarly, using also Taylor's formula. □

Lemma A.4 (Smoothing lemma)**.** *Let* ε *be a positive number and* $B \in \mathscr{B}$. *Let function* $\varphi : \mathbb{R}^p \mapsto \mathbb{R}$ *satisfy* (A.8) *from Lemma A.3. It holds for arbitrary random variables* $X, Y \in \mathbb{R}^p$ *:*

$$\begin{aligned}&\sup_{B\in\mathscr{B}} |\mathbf{P}(X \in B) - \mathbf{P}(Y \in B)| \\ &\leq \sup_{B\in\mathscr{B}} |\mathbf{E}\varphi(X) - \mathbf{E}\varphi(Y)| + \sup_{B\in\mathscr{B}} \mathbf{P}\bigl(Y \in B^{\varepsilon} \setminus B\bigr),\end{aligned}$$

the ball B^{ε} *was introduced in* (A.6)*.*

Lemma A.5 (Some bounds used in the proof)**.** *Let random vector* X *be an i.i.d. copy of* X_i *from Theorem 2.1, let also* $\Sigma_{\boldsymbol{z}}, C_{\boldsymbol{z}}$ *be as in Theorem 2.1, then the following properties hold*

$$\forall\, k \geq 2 \quad C_{\boldsymbol{z}}^k \mathbf{E}\bigl(\|X\|^k\bigr) \geq 1; \tag{A.14}$$

$$\forall\, 2 \leq j \leq k \quad C_{\boldsymbol{z}}^j \mathbf{E}\bigl(\|X\|^j\bigr) \leq C_{\boldsymbol{z}}^k \mathbf{E}\bigl(\|X\|^k\bigr); \tag{A.15}$$

$$\forall\, j \leq K-1 \quad \mathbf{E}\left(|Z_0^{\top} \Sigma_{\boldsymbol{z}}^{-1/2} X| \|X\|^j\right) \leq C_{\boldsymbol{z}} \mathbf{E}(\|X\|^{j+1}); \tag{A.16}$$

$$\forall\, 1 \leq j \leq K \quad \mathbf{E} \int_{\mathbb{R}^p} \left| f_0^{(j)}(x) \left\{ \Sigma_{\boldsymbol{z}}^{-1/2} X \right\}^j \right| dx \leq \sqrt{j!} C_{\boldsymbol{z}}^j \mathbf{E}(\|X\|^j). \tag{A.17}$$

where $Z_0 \sim \mathcal{N}(0, \mathbf{I}_p)$, *independent from* X, *and* $f_0(x) \stackrel{\text{def}}{=} (2\pi)^{-p/2} \mathrm{e}^{-\|x\|^2/2}$ *denotes p.d.f. of the multivariate standard normal distribution in* $\mathbb{R}^p$.

Proof of Lemma A.5. Bounds (A.14), (A.16) follow from Hölder's and Cauchy-Schwarz inequalities. Indeed

$$\begin{aligned} C_{\boldsymbol{z}}^k \mathbf{E}\bigl(\|X\|^k\bigr) &= \|\Sigma_{\boldsymbol{z}}^{-1/2}\|^k \mathbf{E}\bigl(\|X\|^k\bigr) \\ &\geq \|\Sigma^{-1/2}\|^k \mathbf{E}\bigl(\|X\|^k\bigr) \\ &\geq \bigl\{\mathbf{E}\bigl(\|\Sigma^{-1/2} X\|^2\bigr)\bigr\}^{k/2} = p^{k/2} \geq 1; \\ \mathbf{E}\left(|Z_0^{\top} \Sigma_{\boldsymbol{z}}^{-1/2} X| \|X\|^j\right) &= \mathbf{E}\left\{\|X\|^j \mathbf{E}\bigl[|Z_0^{\top} \Sigma_{\boldsymbol{z}}^{-1/2} X| \bigm| X\bigr]\right\} \\ &\leq \mathbf{E}\left\{\|X\|^j \sqrt{\mathbf{E}\bigl[|Z_0^{\top} \Sigma_{\boldsymbol{z}}^{-1/2} X|^2 \bigm| X\bigr]}\right\} \\ &\leq \mathbf{E}(\|X\|^{j+1}) \|\Sigma_{\boldsymbol{z}}^{-1/2}\|. \end{aligned}$$

Inequality (A.15) is implied by Hölder's inequality and by the previous bound (A.14)

$$C_{\boldsymbol{z}}^j \mathbf{E}\bigl(\|X\|^j\bigr) \leq \bigl\{C_{\boldsymbol{z}}^k \mathbf{E}\bigl(\|X\|^k\bigr)\bigr\}^{j/k} \leq C_{\boldsymbol{z}}^k \mathbf{E}\bigl(\|X\|^k\bigr).$$

Now we check the property (A.17). Let $H_j(x)$ denote a multivariate Hermite polynomial defined by the multivariate analogue of Rodrigues's formula $f_0^{(j)}(x) = (-1)^j H_j(x) f_0(x)$. Then by the orthogonal property of Hermite polynomials (see e.g. [16]) and Hölder's inequality, it holds

$$\mathbf{E} \int_{\mathbb{R}^p} \left| f_0^{(j)}(x) X^j \right| dx \leq \mathbf{E} \left\{ \int_{\mathbb{R}^p} \left[H_j(x) X^j \right]^2 f_0(x) dx \right\}^{1/2} \leq \mathbf{E}\left(\|X\|^j\right) \sqrt{j!}.$$

$\square$

Proof of Theorem 2.1. Denote

$$\beta \stackrel{\text{def}}{=} \mathbf{E}\left(\|X_1\|^K + \|Y_1\|^K\right).$$

With this notation, the theorem's statement reads as

$$\Delta_n \stackrel{\text{def}}{=} \sup_{B\in\mathscr{B}} \left|\mathbf{P}\left(S_n \in B\right) - \mathbf{P}\bigl(\widetilde{S}_n \in B\bigr)\right| \le M C_{\mathscr{B}} C_\phi \frac{\left\{C_{\boldsymbol{z}}^K \beta\right\}^{1/(K-2)}}{n^{1/2}}. \tag{A.18}$$

Below we use induction w.r.t. n.

Induction basis. Bound (A.18) holds for all n s.t. $1 \le n \le n'$, where $n' \ge 1$ is some natural number

Indeed, due to property (A.14), and since $C_{\mathscr{B}}, C_\phi \ge 1$, $K \ge 3$, it holds

$$1 \le C_{\mathscr{B}} C_\phi \left\{C_{\boldsymbol{z}}^K \beta/2\right\}^{1/(K-2)},$$

therefore, since $\Delta_n \le 1$, one can take $n \le n' = \left(M C_{\mathscr{B}} C_\phi \left\{C_{\boldsymbol{z}}^K \beta/2\right\}^{1/(K-2)}\right)^2 \ge M^2$ for the induction basis. In the end of the proof we show that $M \ge 72.5$. In the next steps of the proof we consider $n > n'$.

Induction step

Assume that the following bound holds for all $l = 1, \dots, n-1$:

$$\Delta_l \le M C_{\mathscr{B}} C_\phi \frac{\left\{C_{\boldsymbol{z}}^K \beta\right\}^{1/(K-2)}}{l^{1/2}}. \tag{A.19}$$

Our goal is to show that (A.19) is also true for $l = n$. The first step is to apply the smoothing lemma A.4 to Δ_n. Let $\varepsilon > 0$ be some fixed number, it holds

$$\begin{aligned}
\Delta_n &\stackrel{\text{def}}{=} \sup_{B\in\mathscr{B}} \left|\mathbf{P}\left(S_n \in B\right) - \mathbf{P}\bigl(\widetilde{S}_n \in B\bigr)\right| \\
&\le \sup_{B\in\mathscr{B}} \left|\mathbf{E}\varphi(S_n) - \mathbf{E}\varphi(\widetilde{S}_n)\right| + \sup_{B\in\mathscr{B}} \mathbf{P}\bigl(Y \in B^\varepsilon \setminus B\bigr) \\
&\le \sup_{B\in\mathscr{B}} \left|\mathbf{E}\varphi(S_n) - \mathbf{E}\varphi(\widetilde{S}_n)\right| + \varepsilon C_{\mathscr{B}} C_{\boldsymbol{z}},
\end{aligned} \tag{A.20}$$

where the function φ is taken from Lemmas A.3 and A.4, and the last inequality follows from Lemma A.2 and the following property of the approximating sum $\widetilde{S}_n$

$$\mathscr{L}\Bigl(\widetilde{S}_n - \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^n U_i \,\big|\, U_1, \dots, U_n\Bigr) = \mathcal{N}(0, \Sigma_{\boldsymbol{z}}). \tag{A.21}$$

Below we represent the difference $\mathbf{E}\varphi(S_n) - \mathbf{E}\varphi(\widetilde{S}_n)$ as a telescopic sum and divide this sum into four parts, then we derive a bound for each of this parts, and collect all the bounds together in the end of the proof. Let random vectors X, Y, Z, U be i.i.d. copies of X_i, Y_i, Z_i, U_i correspondingly. Denote

$$\overline{X} \stackrel{\text{def}}{=} \frac{X}{\sqrt{n}}, \quad \overline{X}_i \stackrel{\text{def}}{=} \frac{X_i}{\sqrt{n}}, \quad \overline{Y} \stackrel{\text{def}}{=} \frac{Y}{\sqrt{n}}, \quad \overline{Y}_i \stackrel{\text{def}}{=} \frac{Y_i}{\sqrt{n}}. \tag{A.22}$$

Let also

$$W_k \stackrel{\text{def}}{=} \sum\nolimits_{i=1}^{k-1} \overline{Y}_i + \sum\nolimits_{i=k+1}^{n} \overline{X}_i \quad \text{for } k = 2, \ldots, n,$$
$$W_1 \stackrel{\text{def}}{=} \overline{X}_2 + \cdots + \overline{X}_n, \quad W_n \stackrel{\text{def}}{=} \overline{Y}_1 + \cdots + \overline{Y}_{n-1},$$

thus

$$S_n = \overline{X}_1 + W_1, \quad \widetilde{S}_n = W_n + \overline{Y}_n, \quad W_k + \overline{Y} \stackrel{w}{=} W_{k+1} + \overline{X},$$

here W_k is independent of $\overline{X}_k, \overline{Y}_k$ and $\overline{X}, \overline{Y}$; we will use this property further in the proofs. Denote

$$\gamma_k \stackrel{\text{def}}{=} \left|\mathbf{E}\varphi\left(W_k + \overline{X}\right) - \mathbf{E}\varphi\left(W_k + \overline{Y}\right)\right|, \tag{A.23}$$

therefore

$$\begin{aligned}
&\left|\mathbf{E}\varphi(S_n) - \mathbf{E}\varphi(\widetilde{S}_n)\right| \\
&= \left|\mathbf{E}\varphi\left(W_1 + \overline{X}\right) - \mathbf{E}\varphi\left(W_n + \overline{Y}\right)\right| \\
&= \left|\mathbf{E}\varphi\left(W_1 + \overline{X}\right) - \mathbf{E}\varphi\left(W_1 + \overline{Y}\right) + \mathbf{E}\varphi\left(W_1 + \overline{Y}\right) - \mathbf{E}\varphi\left(W_n + \overline{Y}\right)\right| \\
&= \left|\mathbf{E}\varphi\left(W_1 + \overline{X}\right) - \mathbf{E}\varphi\left(W_1 + \overline{Y}\right) + \mathbf{E}\varphi\left(W_2 + \overline{X}\right) - \mathbf{E}\varphi\left(W_n + \overline{Y}\right)\right| \\
&\leq \sum\nolimits_{k=1}^{n} \gamma_k.
\end{aligned}$$

Define $n_0 \stackrel{\text{def}}{=} [n/2]$, let m be some natural number $2 \leq m \leq n_0$. Let us split the sum from the last inequality as follows:

$$\sum_{k=1}^{n} \gamma_k = \gamma_1 + \sum_{k=2}^{m} \gamma_k + \sum_{k=m+1}^{n_0} \gamma_k + \sum_{k=n_0+1}^{n} \gamma_k. \tag{A.24}$$

Induction step, part 1

We shall prove that

$$\begin{aligned}
\gamma_1 &\stackrel{\text{def}}{=} \left|\mathbf{E}\varphi(W_1 + \overline{X}) - \mathbf{E}\varphi(W_1 + \overline{Y})\right| \\
&\leq \frac{1.5\widetilde{C}_\phi C_{\boldsymbol{z}}^2}{(K-2)!} \frac{\beta}{(\varepsilon/2)^{(K-2)} n^{(K-2)/2}} \left\{ 2MC_{\mathscr{B}}C_\phi \frac{\left\{C_{\boldsymbol{z}}^K \beta\right\}^{1/(K-2)}}{n^{1/2}} + \varepsilon C_{\mathscr{B}} C_{\boldsymbol{z}} \right\}.
\end{aligned} \tag{A.25}$$

It holds

$$
\begin{aligned}
&\mathbf{E}\varphi(W_1+\overline{X}) - \frac{1}{s!}\sum_{s=0}^{K-3}\mathbf{E}\left\{\varphi^{(s)}(W_1)\overline{X}^s\right\} \\
&= \frac{1}{(K-3)!}\mathbf{E}\left\{(1-\tau)^{K-3}\varphi^{(K-2)}(W_1+\tau\overline{X})\overline{X}^{K-2}\right\} && \text{(A.26)}\\
&\le \frac{\widetilde{C}_\phi}{(K-3)!(\varepsilon/2)^{K-2}}\mathbf{E}\left(\|\overline{X}\|^{K-2}(1-\tau)^{K-3}\,\mathbb{I}\left\{W_1+\tau\overline{X}\in B^\varepsilon\setminus B\right\}\right) && \text{(A.27)}\\
&\le \frac{\widetilde{C}_\phi}{(K-2)!(\varepsilon/2)^{K-2}}\mathbf{E}\left(\|\overline{X}\|^{K-2}\right)\left\{2\varDelta_{n-1}+\varepsilon\alpha_1 C_{\mathscr{B}}C_{\boldsymbol{z}}\right\} && \text{(A.28)}\\
&\le \frac{\widetilde{C}_\phi C_{\boldsymbol{z}}^2}{(K-2)!(\varepsilon/2)^{K-2}}\frac{\mathbf{E}\left(\|X\|^K\right)}{n^{(K-2)/2}}\left\{2\varDelta_{n-1}+\varepsilon\alpha_1 C_{\mathscr{B}}C_{\boldsymbol{z}}\right\} && \text{(A.29)}
\end{aligned}
$$

here (A.26) follows from the Taylor formula (A.7) taken with $s = K-3$; (A.27) follows from Lemma A.3; (A.29) follows from (A.15); inequality (A.28) is derived below:

$$
\begin{aligned}
&\mathbf{E}\left(\|\overline{X}\|^{K-2}(1-\tau)^{K-3}\,\mathbb{I}\left\{W_1+\tau\overline{X}\in B^\varepsilon\setminus B\right\}\right)\\
&= \mathbf{E}\left\{\|\overline{X}\|^{K-2}(1-\tau)^{K-3}\mathbf{P}\left(W_1+\tau\overline{X}\in B^\varepsilon\setminus B\,\middle|\,\tau,\overline{X}\right)\right\}\\
&\le \frac{1}{K-2}\mathbf{E}\left(\|\overline{X}\|^{K-2}\right)\sup_{z\in\mathbb{R}^p}\mathbf{P}\left(W_1\in B^\varepsilon\setminus B+z\right)\\
&\le \frac{1}{K-2}\mathbf{E}\left(\|\overline{X}\|^{K-2}\right)\sup_{B\in\mathscr{B}}\mathbf{P}\left(W_1\in B^\varepsilon\setminus B\right)\\
&= \frac{1}{K-2}\mathbf{E}\left(\|\overline{X}\|^{K-2}\right)\sup_{B\in\mathscr{B}}\mathbf{P}\left(S_{n-1}/\alpha_1\in B^\varepsilon\setminus B\right)\\
&\le \frac{1}{K-2}\mathbf{E}\left(\|\overline{X}\|^{K-2}\right)\left\{2\varDelta_{n-1}+\varepsilon\alpha_1 C_{\mathscr{B}}C_{\boldsymbol{z}}\right\}, && \text{(A.30)}
\end{aligned}
$$

where $\alpha_1 \stackrel{\text{def}}{=} \sqrt{n/(n-1)}$; inequality (A.30) is implied by induction assumption (A.19), property (A.21) and Lemma A.2, indeed

$$
\begin{aligned}
\mathbf{P}\left(S_{n-1}\in B^{\varepsilon\alpha_1}\right)-\mathbf{P}\left(S_{n-1}\in B\right) &\le 2\varDelta_{n-1}+\mathbf{P}\left(\widetilde{S}_{n-1}\in B^{\varepsilon\alpha_1}\setminus B\right)\\
&\le 2\varDelta_{n-1}+\varepsilon\alpha_1 C_{\mathscr{B}}C_{\boldsymbol{z}}.
\end{aligned} \tag{A.31}
$$

Similar bounds hold for $\overline{Y}$, therefore, using independence of W_1 of $\overline{X}$ and $\overline{Y}$, and condition $\mathbf{E}\big(\overline{X}^k\big)=\mathbf{E}\big(\overline{Y}^k\big)\ \forall\, k=0,\ldots,K-1$, we infer

$$
\begin{aligned}
&\left|\mathbf{E}\varphi(W_1+\overline{X})-\mathbf{E}\varphi(W_1+\overline{Y})\right|\\
&\le \frac{\widetilde{C}_\phi C_{\boldsymbol{z}}^2}{(K-2)!}\frac{\beta}{(\varepsilon/2)^{(K-2)}n^{(K-2)/2}}\left\{2\varDelta_{n-1}+\varepsilon C_{\mathscr{B}}C_{\boldsymbol{z}}\sqrt{n/(n-1)}\right\},
\end{aligned}
$$

which implies (A.25) due to induction assumption (A.19).

Induction step, part 2

Below we show that

$$
\begin{aligned}
&\gamma_2 + \dots + \gamma_m \\
&\le \frac{4\sqrt{2}\left(\sqrt{m-1}-1\right)\widetilde{C}_{\phi,1}C_{\boldsymbol{z}}}{(\varepsilon/2)^{(K-1)}K!}\frac{\beta}{n^{(K-1)/2}}\left(2MC_{\mathscr{B}}C_{\phi}\frac{\left\{C_{\boldsymbol{z}}^{K}\beta\right\}^{1/(K-2)}}{n^{1/2}} + \varepsilon C_{\mathscr{B}}C_{\boldsymbol{z}}\right).
\end{aligned}
\tag{A.32}
$$

Let us fix some integer k s.t. $2 \le k \le n-1$. Introduce similarly to (A.22)

$$
\overline{Z} \stackrel{\text{def}}{=} \frac{Z}{\sqrt{n}}, \quad \overline{U} \stackrel{\text{def}}{=} \frac{U}{\sqrt{n}}, \quad \overline{Z}_i \stackrel{\text{def}}{=} \frac{Z_i}{\sqrt{n}}, \quad \overline{U}_i \stackrel{\text{def}}{=} \frac{U_i}{\sqrt{n}},
$$

therefore, $\overline{Y} = \overline{Z} + \overline{U}$, $\overline{Y}_i = \overline{Z}_i + \overline{U}_i$, and

$$
\begin{aligned}
\sum_{i=1}^{k-1}\overline{Y}_i = \sum_{i=1}^{k-1}\overline{Z}_i + \sum_{i=1}^{k-1}\overline{U}_i &\stackrel{w}{=} (k-1)^{1/2}\overline{Z} + \sum_{i=1}^{k-1}\overline{U}_i \\
&\stackrel{w}{=} \frac{(k-1)^{1/2}}{\sqrt{n}}\Sigma_{\boldsymbol{z}}^{1/2}Z_0 + \sum_{i=1}^{k-1}\overline{U}_i,
\end{aligned}
$$

where $Z_0 \sim \mathcal{N}(0, \mathbf{I}_p)$, independent of all $\overline{U}_i, \overline{X}_i, \overline{X}$. Denote

$$
\begin{aligned}
\overline{X}_{sum} \stackrel{\text{def}}{=} \sum\nolimits_{i=k+1}^{n}\overline{X}_i, \quad \overline{U}_{sum} \stackrel{\text{def}}{=} \sum\nolimits_{i=1}^{k-1}\overline{U}_i, \\
1/\theta_k \stackrel{\text{def}}{=} \sqrt{(k-1)/n}.
\end{aligned}
\tag{A.33}
$$

Using all the notation introduced above, we have:

$$
\begin{aligned}
W_k + \overline{X} &\stackrel{w}{=} \Sigma_{\boldsymbol{z}}^{1/2}Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} + \overline{X}, \\
W_k + \overline{Y} &\stackrel{w}{=} \Sigma_{\boldsymbol{z}}^{1/2}Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} + \overline{Y}.
\end{aligned}
\tag{A.34}
$$

Let $f_0(x)$ denote p.d.f. of multivariate standard normal distribution: $f_0(x) \stackrel{\text{def}}{=} (2\pi)^{-p/2}\mathrm{e}^{-\|x\|^2/2}$. Due to (A.34) it holds for $y \stackrel{\text{def}}{=} x/\theta_k + \Sigma_{\boldsymbol{z}}^{-1/2}\overline{X}$

$$
\begin{aligned}
\mathbf{E}\varphi(W_k + \overline{X}) &= \mathbf{E}\varphi(\Sigma_{\boldsymbol{z}}^{1/2}Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} + \overline{X}) \\
&= \mathbf{E}\int_{\mathbb{R}^p}\varphi(\Sigma_{\boldsymbol{z}}^{1/2}x/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} + \overline{X})f_0(x)dx \\
&= \theta_k^p\mathbf{E}\int_{\mathbb{R}^p}\varphi(\Sigma_{\boldsymbol{z}}^{1/2}y + \overline{U}_{sum} + \overline{X}_{sum})f_0(\theta_k y - \theta_k\Sigma_{\boldsymbol{z}}^{-1/2}\overline{X})dy.
\end{aligned}
$$

Expand the density function $f_0(\theta_k y - \theta_k\Sigma_{\boldsymbol{z}}^{-1/2}\overline{X})$, conditioned on $\overline{X}$, by Taylor formula (A.7):

$$
\begin{aligned}
&f_0(\theta_k y - \theta_k\Sigma_{\boldsymbol{z}}^{-1/2}\overline{X}) \\
&= f_0(\theta_k y) - \theta_k\Sigma_{\boldsymbol{z}}^{-1/2}\overline{X}\mathbf{E}f_0'(\theta_k y - \tau\theta_k\Sigma_{\boldsymbol{z}}^{-1/2}\overline{X}) \\
&= f_0(\theta_k y) + \theta_k^2\mathbf{E}(y - \tau\Sigma_{\boldsymbol{z}}^{-1/2}\overline{X})^\top\Sigma_{\boldsymbol{z}}^{-1/2}\overline{X}f_0(\theta_k y - \tau\theta_k\Sigma_{\boldsymbol{z}}^{-1/2}\overline{X}).
\end{aligned}
$$

Consider

$$
\begin{aligned}
I_0 &\stackrel{\text{def}}{=} \theta_k^p \mathbf{E} \int_{\mathbb{R}^p} \varphi(\Sigma_{\boldsymbol{z}}^{1/2} y + \overline{U}_{sum} + \overline{X}_{sum}) f_0(\theta_k y) dy \\
&= \mathbf{E}\varphi(\Sigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum}),
\end{aligned} \tag{A.35}
$$

$$
\begin{aligned}
I_1 &\stackrel{\text{def}}{=} \theta_k^p \mathbf{E} \int_{\mathbb{R}^p} \varphi(\Sigma_{\boldsymbol{z}}^{1/2} y + \overline{U}_{sum} + \overline{X}_{sum}) \theta_k^2 (y - \tau \Sigma_{\boldsymbol{z}}^{-1/2} \overline{X})^\top \Sigma_{\boldsymbol{z}}^{-1/2} \overline{X} \\
&\times f_0(\theta_k y - \tau\theta_k \Sigma_{\boldsymbol{z}}^{-1/2} \overline{X}) dy \\
&= \theta_k \mathbf{E} \left\{ \varphi(\Sigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} + \tau\overline{X}) Z_0^\top \Sigma_{\boldsymbol{z}}^{-1/2} \overline{X} \right\} \\
&= \theta_k \mathbf{E} \Bigg\{ (Z_0^\top \Sigma_{\boldsymbol{z}}^{-1/2} \overline{X}) \sum_{j=1}^{N-1} \frac{1}{j!} \varphi^{(j)}(\Sigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum}) (\tau \overline{X})^j \\
&+ \frac{(1-\tau_1)^{N-1}}{(N-1)!} (Z_0^\top \Sigma_{\boldsymbol{z}}^{-1/2} \overline{X}) \varphi^{(N)}(\Sigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} + \tau\tau_1 \overline{X}) (\tau \overline{X})^N \Bigg\},
\end{aligned} \tag{A.36}
$$

where $N \stackrel{\text{def}}{=} K-2$; in the last equation we again used Taylor formula (A.7) (τ_1 is an i.i.d. copy of τ) and also the property $\mathbf{E}\overline{X} = 0$ together with independence of $\overline{X}$ from $Z_0, \overline{U}_{sum}, Z$. It holds

$$
\mathbf{E}\varphi(W_k + \overline{X}) = I_0 + I_1.
$$

Analogous relations hold for the second term in γ_k :

$$
\mathbf{E}\varphi(W_k + \overline{Y}) = I_0 + J_1,
$$

for

$$
\begin{aligned}
J_1 &\stackrel{\text{def}}{=} \theta_k \mathbf{E} \Bigg\{ (Z_0^\top \Sigma_{\boldsymbol{z}}^{-1/2} \overline{Y}) \sum_{j=1}^{N-1} \frac{1}{j!} \varphi^{(j)}(\Sigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum}) (\tau \overline{Y})^j \\
&+ \frac{(1-\tau_1)^{N-1}}{(N-1)!} (Z_0^\top \Sigma_{\boldsymbol{z}}^{-1/2} \overline{Y}) \varphi^{(N)}(\Sigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} + \tau\tau_1 \overline{Y}) (\tau \overline{Y})^N \Bigg\},
\end{aligned}
$$

here we substituted $\overline{X}$ with $\overline{Y}$ in expressions (A.35), (A.36) for I_0, I_1 , and I_0 remained the same.

Introduce

$$
\begin{aligned}
I_2 = \theta_k \mathbf{E}\Biggl\{ & (Z_0^\top \varSigma_{\boldsymbol{z}}^{-1/2}\overline{X}) \sum_{j=1}^{N-1} \frac{1}{j!}\varphi^{(j)}(\varSigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum})(\tau\overline{X})^j \\
& + \frac{(1-\tau_1)^{N-1}}{(N-1)!}(Z_0^\top \varSigma_{\boldsymbol{z}}^{-1/2}\overline{X})\varphi^{(N)}(\varSigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum})(\tau\overline{X})^N \Biggr\}, \\
J_2 = \theta_k \mathbf{E}\Biggl\{ & (Z_0^\top \varSigma_{\boldsymbol{z}}^{-1/2}\overline{Y}) \sum_{j=1}^{N-1} \frac{1}{j!}\varphi^{(j)}(\varSigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum})(\tau\overline{Y})^j \\
& + \frac{(1-\tau_1)^{N-1}}{(N-1)!}(Z_0^\top \varSigma_{\boldsymbol{z}}^{-1/2}\overline{Y})\varphi^{(N)}(\varSigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum})(\tau\overline{Y})^N \Biggr\}.
\end{aligned}
$$

$I_2 = J_2$ since $\overline{X}, \overline{Y}$ are independent of $Z_0, \overline{U}_{sum}, \overline{X}_{sum}, \tau$ and $\mathbf{E}(\overline{X}^k) = \mathbf{E}(\overline{Y}^k)$ $\forall\, k = 0, \dots, K-1$. Therefore, we can write:

$$
\gamma_k = |I_0 + I_1 - I_0 - J_1| = |I_1 - I_2 + J_2 - J_1|\,.
$$

Due to property (A.10) of function $\varphi(x)$, it holds (since $N = K-2$):

$$
\begin{aligned}
& |I_1 - I_2| \\
& \quad \le \frac{\widetilde{C}_{\phi,1}\theta_k}{(\varepsilon/2)^{N+1}} \mathbf{E}\Bigl[\frac{(1-\tau_1)^{N-1}}{(N-1)!} |Z_0^\top \varSigma_{\boldsymbol{z}}^{-1/2}\overline{X}| \tau^{N+1}\tau_1 \|\overline{X}\|^{N+1} \\
& \qquad \times \Bigl\{ \mathbf{P}\left(\varSigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} + \tau\tau_1\overline{X} \in B^{\varepsilon} \setminus B \,\middle|\, Z_0, \overline{U}_{sum}, \overline{X}, \tau, \tau_1 \right) \\
& \qquad\quad + \mathbf{P}\left(\varSigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} \in B^{\varepsilon} \setminus B \,\middle|\, Z_0, \overline{U}_{sum} \right) \Bigr\} \Bigr] \\
& \quad \le \frac{2\widetilde{C}_{\phi,1}\theta_k}{(\varepsilon/2)^{N+1}(N+2)!} \mathbf{E}\Bigl[|Z_0^\top \varSigma_{\boldsymbol{z}}^{-1/2}\overline{X}| \|\overline{X}\|^{N+1} \sup_{z\in\mathbb{R}^p} \mathbf{P}\left(z + S_{n-k} \in B^{\varepsilon\alpha_k} \setminus B \right) \Bigr] \\
& \quad \le \frac{2\widetilde{C}_{\phi,1}\theta_k \|\varSigma_{\boldsymbol{z}}^{-1/2}\|}{(\varepsilon/2)^{N+1}(N+2)!} \frac{\mathbf{E}\left(\|X\|^{N+2}\right)}{n^{(N+2)/2}} \left(2\varDelta_{n-k} + \varepsilon\alpha_k C_{\mathscr{B}} C_{\boldsymbol{z}}\right),
\end{aligned}
\tag{A.37}
$$

where

$$
\alpha_k \stackrel{\text{def}}{=} \sqrt{n/(n-k)}. \tag{A.38}
$$

Inequality (A.37) follows from bound (A.16) in Lemma A.5, from Lemma A.2 and the following bounds (cf. (A.31)):

$$
\begin{aligned}
& \mathbf{P}\left(S_{n-k} \in B^{\varepsilon\alpha_k}\right) - \mathbf{P}\left(S_{n-k} \in B\right) \\
& \le 2\varDelta_{n-k} + \mathbf{P}\left(\widetilde{S}_{n-k} \in B^{\varepsilon\alpha_k}\right) - \mathbf{P}\left(\widetilde{S}_{n-k} \in B\right) \\
& \le 2\varDelta_{n-k} + \varepsilon\alpha_k C_{\mathscr{B}} C_{\boldsymbol{z}}.
\end{aligned}
$$

The analogous bound holds for $|J_1 - J_2|$:

$$|J_1 - J_2| \le \frac{2\widetilde{C}_{\phi,1}\theta_k \|\Sigma_{\boldsymbol{z}}^{-1/2}\|}{(\varepsilon/2)^{N+1}(N+2)!} \frac{\mathbf{E}\left(\|Y\|^{N+2}\right)}{n^{(N+2)/2}} \left(2\Delta_{n-k} + \varepsilon\alpha_k C_{\mathscr{B}} C_{\boldsymbol{z}}\right).$$

Using that $N = K-2$, we have for any $k = 2, \ldots, n-1$

$$\begin{aligned}
\gamma_k &\le |I_1 - I_2| + |J_2 - J_1| \\
&\le \frac{2\widetilde{C}_{\phi,1}\theta_k \|\Sigma_{\boldsymbol{z}}^{-1/2}\|}{(\varepsilon/2)^{K-1}K!} \frac{\mathbf{E}\left(\|X\|^{K} + \|Y\|^{K}\right)}{n^{K/2}} \left(2\Delta_{n-k} + \varepsilon\alpha_k C_{\mathscr{B}} C_{\boldsymbol{z}}\right) \\
&\le \frac{2\alpha_k\widetilde{C}_{\phi,1} \|\Sigma_{\boldsymbol{z}}^{-1/2}\|}{(\varepsilon/2)^{K-1}K!} \frac{\beta}{n^{(K-1)/2}\sqrt{k-1}} \left(2MC_{\mathscr{B}}C_{\phi}\frac{\left\{C_{\boldsymbol{z}}^K\beta\right\}^{1/(K-2)}}{n^{1/2}} + \varepsilon C_{\mathscr{B}} C_{\boldsymbol{z}}\right),
\end{aligned}$$

where the last inequality follows from (A.33), (A.38), and induction assumption (A.19). For $2 \le m$

$$\sum\nolimits_{j=2}^{m} \frac{1}{\sqrt{k-1}} \le 1 + \int_2^m \frac{dt}{\sqrt{t-1}} = 2\sqrt{m-1} - 2.$$

Moreover, $\alpha_k = \sqrt{n/(n-k)} \le \sqrt{2}$ for $k \le n/2$. These properties and the last bound on γ_k imply the resulting inequlaity (A.32).

Induction step, part 3

Below it is shown that

$$\gamma_{m+1} + \cdots + \gamma_{n_0} \le \frac{C_{\boldsymbol{z}}^K\beta}{\sqrt{K!}} \left(\frac{MC_{\mathscr{B}}C_{\phi}\left\{C_{\boldsymbol{z}}^K\beta\right\}^{1/(K-2)}2\sqrt{2}}{n^{1/2}(m-1)^{(K-2)/2}} + \frac{2^{(K-2)/2}}{n^{(K-2)/2}}\right). \tag{A.39}$$

Let us fix some integer k s.t. $2 \le k \le n$. Due to Lemma A.3

$$\begin{aligned}
\gamma_k &= \left|\mathbf{E}\varphi(W_k + \overline{X}) - \mathbf{E}\varphi(W_k + \overline{Y})\right| \\
&= \left|\int_{\mathbb{R}} \phi(t/\widetilde{\varepsilon}) d\left[\mathbf{P}\left\{\widetilde{\rho}(W_k + \overline{X}) \le t\right\} - \mathbf{P}\left\{\widetilde{\rho}(W_k + \overline{Y}) \le t\right\}\right]\right| \\
&\le \int_{\mathbb{R}} |\phi'(t/\widetilde{\varepsilon})| \left|\mathbf{P}\left\{\widetilde{\rho}(W_k + \overline{X}) \le t\right\} - \mathbf{P}\left\{\widetilde{\rho}(W_k + \overline{Y}) \le t\right\}\right| \frac{dt}{\widetilde{\varepsilon}} \\
&\le \sup\nolimits_{B\in\mathscr{B}} \left|\mathbf{P}\left\{W_k + \overline{X} \in B\right\} - \mathbf{P}\left\{W_k + \overline{Y} \in B\right\}\right|.
\end{aligned}$$

Representations (A.34) imply

$$\begin{aligned}
\mathbf{P}\left\{W_k + \overline{X} \in B\right\} &= \mathbf{P}\left\{\Sigma_{\boldsymbol{z}}^{1/2} Z_0/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} + \overline{X} \in B\right\} \\
&= \mathbf{E}\int_{\mathbb{R}^p} \mathbb{1}\left\{\Sigma_{\boldsymbol{z}}^{1/2} x/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} + \overline{X} \in B\right\} f_0(x)dx \\
&= \theta_k^p \mathbf{E}\int_{\mathbb{R}^p} \mathbb{1}\left\{\Sigma_{\boldsymbol{z}}^{1/2} y + \overline{U}_{sum} + \overline{X}_{sum} \in B\right\} f_0(\theta_k y - \theta_k \Sigma_{\boldsymbol{z}}^{-1/2}\overline{X})dy.
\end{aligned}$$

Similarly

$$\mathbf{P}\left\{W_k+\overline{Y}\in B\right\}=\theta_k^p\mathbf{E}\int_{\mathbb{R}^p}1\!\!1\left\{\varSigma_{\boldsymbol{z}}^{1/2}y+\overline{U}_{sum}+\overline{X}_{sum}\in B\right\}f_0(\theta_k y-\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{Y})dy.$$

Applying Taylor formula (A.7) with $s=K-1$ to the density function $f_0(\theta_k y-\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{X})$, we have

$$\begin{aligned}
f_0(\theta_k y-\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{X})&=f_0(\theta_k y)+\sum_{j=1}^{K-1}\frac{(-1)^j}{j!}f_0^{(j)}(\theta_k y)\left\{\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{X}\right\}^j\\
&+\frac{(-1)^K}{(K-1)!}\mathbf{E}(1-\tau)^{K-1}f_0^{(K)}(\theta_k y-\tau\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{X})\left\{\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{X}\right\}^K.
\end{aligned}$$

Similar expansion holds for $f_0(\theta_k y-\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{Y})$, with $\overline{X}$ replaced by $\overline{Y}$. Since $\overline{X},\overline{Y}$ are independent of $\overline{U}_{sum}$ and $\overline{X}_{sum}$, and $\mathbf{E}(\overline{X}^j)=\mathbf{E}(\overline{Y}^j)$ $\forall j=0,\ldots,K-1$, it holds

$$\begin{aligned}
&\left|\mathbf{P}\left\{W_k+\overline{X}\in B\right\}-\mathbf{P}\left\{W_k+\overline{Y}\in B\right\}\right|\\
&=\frac{\theta_k^p}{(K-1)!}\Bigg|\mathbf{E}(1-\tau)^{K-1}\int_{\mathbb{R}^p}1\!\!1\left\{\varSigma_{\boldsymbol{z}}^{1/2}y+\overline{U}_{sum}+\overline{X}_{sum}\in B\right\}\\
&\quad\times\Bigg[f_0^{(K)}(\theta_k y-\tau\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{X})\left\{\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{X}\right\}^K\\
&\qquad-f_0^{(K)}(\theta_k y-\tau\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{Y})\left\{\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{Y}\right\}^K\Bigg]dy\Bigg|\\
&\le I_X+I_Y,
\end{aligned}\tag{A.40}$$

where

$$\begin{aligned}
I_X&\stackrel{\text{def}}{=}\Bigg|\mathbf{E}\int_{\mathbb{R}^p}1\!\!1\left\{\varSigma_{\boldsymbol{z}}^{1/2}x/\theta_k+\overline{U}_{sum}+\overline{X}_{sum}+\tau\overline{X}\in B\right\}f_0^{(K)}(x)dx\\
&\quad\times\left[\frac{(1-\tau)^{K-1}}{(K-1)!}\left\{\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{X}\right\}^K\right]\Bigg|,\\
I_Y&\stackrel{\text{def}}{=}\Bigg|\mathbf{E}\int_{\mathbb{R}^p}1\!\!1\left\{\varSigma_{\boldsymbol{z}}^{1/2}x/\theta_k+\overline{U}_{sum}+\overline{X}_{sum}+\tau\overline{Y}\in B\right\}f_0^{(K)}(x)dx\\
&\quad\times\left[\frac{(1-\tau)^{K-1}}{(K-1)!}\left\{\theta_k\varSigma_{\boldsymbol{z}}^{-1/2}\overline{Y}\right\}^K\right]\Bigg|.
\end{aligned}$$

Consider the term I_X, it holds

$$\begin{aligned}
&\mathbf{P}\left(\varSigma_{\boldsymbol{z}}^{1/2}x/\theta_k+\overline{U}_{sum}+\overline{X}_{sum}+\tau\overline{X}\in B\,\middle|\,\overline{U}_{sum},\overline{X},\tau\right)\\
&\le\varDelta_{n-k}+\mathbf{P}\left(\varSigma_{\boldsymbol{z}}^{1/2}x/\theta_k+\overline{U}_{sum}+\widetilde{S}_{n-k}/\alpha_k+\tau\overline{X}\in B\,\middle|\,\overline{U}_{sum},\overline{X},\tau\right).
\end{aligned}$$

Denote

$$I_{X,2} \overset{\text{def}}{=} \left| \mathbf{E} \int_{\mathbb{R}^p} \mathbb{1}\left\{ \varSigma_{\boldsymbol{z}}^{1/2} x/\theta_k + \overline{U}_{sum} + \widetilde{S}_{n-k}/\alpha_k + \tau \overline{X} \in B \right\} f_0^{(K)}(x) dx \right. \\ \left. \times \left[\frac{(1-\tau)^{K-1}}{(K-1)!} \left\{ \theta_k \varSigma_{\boldsymbol{z}}^{-1/2} \overline{X} \right\}^K \right] \right|;$$

we have

$$I_X \le \varDelta_{n-k} \mathbf{E} \int_{\mathbb{R}^p} \left| f_0^{(K)}(x) \frac{(1-\tau)^{K-1}}{(K-1)!} \left\{ \theta_k \varSigma_{\boldsymbol{z}}^{-1/2} \overline{X} \right\}^K \right| dx + I_{X,2}. \tag{A.41}$$

By (A.17) in Lemma A.5 and (A.33)

$$\mathbf{E} \int_{\mathbb{R}^p} \left| f_0^{(K)}(x) \frac{(1-\tau)^{K-1}}{(K-1)!} \left\{ \theta_k \varSigma_{\boldsymbol{z}}^{-1/2} \overline{X} \right\}^K \right| dx \le \frac{\| \varSigma_{\boldsymbol{z}}^{-1/2} \|^K \mathbf{E}(\|X\|^K)}{\sqrt{K!} (k-1)^{K/2}}.$$

Let us consider the term $I_{X,2}$. By definition (A.38)

$$\widetilde{S}_{n-k}/\alpha_k \overset{w}{=} \varSigma_{\boldsymbol{z}}^{1/2} Z_0/\overline{\theta}_k + \overline{\overline{U}}_{sum},$$

where $Z_0 \sim \mathcal{N}(0, \mathbf{I}_p)$ is independent of all other r.v., and

$$1/\overline{\theta}_k \overset{\text{def}}{=} \sqrt{(n-k)/n}, \quad \overline{\overline{U}}_{sum} \overset{\text{def}}{=} \sum\nolimits_{i=k+1}^n \overline{U}_i.$$

In this way

$$\begin{aligned} I_{X,2} = & \frac{\theta_k^K}{(K-1)!} \left| \mathbf{E} (1-\tau)^{K-1} \int_{\mathbb{R}^p} \mathbb{1}\Big\{ \varSigma_{\boldsymbol{z}}^{1/2} x/\theta_k + \overline{U}_{sum} + \overline{\overline{U}}_{sum} \right. \\ & \left. \qquad + \varSigma_{\boldsymbol{z}}^{1/2} Z_0/\overline{\theta}_k + \tau \overline{X} \in C \Big\} f_0^{(K)}(x) dx \left\{ \varSigma_{\boldsymbol{z}}^{-1/2} \overline{X} \right\}^K \right| \\ = & \frac{\theta_k^{K+p} \overline{\theta}_k^p}{(K-1)!} \left| \mathbf{E} (1-\tau)^{K-1} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathbb{1}\Big\{ \varSigma_{\boldsymbol{z}}^{1/2} x + \overline{U}_{sum} + \overline{\overline{U}}_{sum} \right. \\ & \left. \qquad + \varSigma_{\boldsymbol{z}}^{1/2} y + \tau \overline{X} \in C \Big\} f_0^{(K)}(\theta_k x) f_0(y \overline{\theta}_k) dx dy \left\{ \varSigma_{\boldsymbol{z}}^{-1/2} \overline{X} \right\}^K \right| \\ = & \frac{\theta_k^p \overline{\theta}_k^{K+p}}{(K-1)!} \left| \mathbf{E} (1-\tau)^{K-1} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathbb{1}\Big\{ \varSigma_{\boldsymbol{z}}^{1/2} x + \overline{U}_{sum} + \overline{\overline{U}}_{sum} \right. \qquad (A.42) \\ & \left. \qquad + \varSigma_{\boldsymbol{z}}^{1/2} y + \tau \overline{X} \in C \Big\} f_0(\theta_k x) f_0^{(K)}(y \overline{\theta}_k) dx dy \left\{ \varSigma_{\boldsymbol{z}}^{-1/2} \overline{X} \right\}^K \right| \\ \le & \frac{\overline{\theta}_k^K}{K!} \mathbf{E} \int_{\mathbb{R}^p} \left| f_0^{(K)}(y) \left\{ \varSigma_{\boldsymbol{z}}^{-1/2} \overline{X} \right\}^K \right| dy \\ \le & \frac{\| \varSigma_{\boldsymbol{z}}^{-1/2} \|^K \mathbf{E}(\|X\|^K)}{\sqrt{K!} (n-k)^{K/2}}, \end{aligned}$$

here equality (A.42) is obtained by integrating by parts K times, and the last inequality follows from (A.17) in Lemma A.5.

Therefore, bound (A.41) reads as

$$I_X \le \frac{\|\Sigma_{\boldsymbol{z}}^{-1/2}\|^K \mathbf{E}(\|X\|^K)}{\sqrt{K!}} \left(\frac{\Delta_{n-k}}{(k-1)^{K/2}} + \frac{1}{(n-k)^{K/2}} \right).$$

The analogous inequality holds for the term I_Y, which implies

$$\begin{aligned}
\gamma_k &\le \frac{\|\Sigma_{\boldsymbol{z}}^{-1/2}\|^K \beta}{\sqrt{K!}} \left(\frac{\Delta_{n-k}}{(k-1)^{K/2}} + \frac{1}{(n-k)^{K/2}} \right) \\
&\le \frac{C_{\boldsymbol{z}}^k \beta}{\sqrt{K!}} \left(\frac{M C_{\mathscr{B}} C_\phi \left\{ C_{\boldsymbol{z}}^K \beta \right\}^{1/(K-2)} \sqrt{2}}{n^{1/2}(k-1)^{K/2}} + \frac{1}{\{n/2\}^{K/2}} \right),
\end{aligned}$$

here we used induction assumption (A.19) and $k \le n/2$. Using the bound

$$\sum_{k=m+1}^{[n/2]} \frac{1}{(k-1)^{K/2}} \le \int_{m-1}^{+\infty} \frac{dt}{t^{K/2}} = \frac{1}{(m-1)^{K/2-1}} \frac{1}{K/2-1} \le \frac{2}{(m-1)^{(K-2)/2}},$$

we infer the resulting inequality (A.39):

$$\gamma_{m+1} + \dots + \gamma_{n_0} \le \frac{C_{\boldsymbol{z}}^k \beta}{\sqrt{K!}} \left(\frac{M C_{\mathscr{B}} C_\phi \left\{ C_{\boldsymbol{z}}^K \beta \right\}^{1/(K-2)} 2\sqrt{2}}{n^{1/2}(m-1)^{(K-2)/2}} + \frac{1}{\{n/2\}^{(K-2)/2}} \right).$$

Induction step, part 4

For the last part of the sum (A.24) it holds

$$\gamma_{n_0+1} + \dots + \gamma_n \le \frac{2 C_{\boldsymbol{z}}^K \beta}{n^{(K-2)/2}}. \tag{A.43}$$

Indeed, consider the bound (A.40), due to (A.17) from Lemma A.5 and definition (A.33)

$$\begin{aligned}
I_X &= \left| \mathbf{E} \frac{(1-\tau)^{K-1}}{(K-1)!} \int_{\mathbb{R}^p} \mathbb{1} \left\{ \Sigma_{\boldsymbol{z}}^{1/2} x/\theta_k + \overline{U}_{sum} + \overline{X}_{sum} + \tau \overline{X} \in C \right\} f_0^{(K)}(x) dx \right. \\
&\quad \left. \times \left\{ \theta_k \Sigma_{\boldsymbol{z}}^{-1/2} \overline{X} \right\}^K \right| \\
&\le \frac{\theta_k^K}{K!} \mathbf{E} \int_{\mathbb{R}^p} \left| f_0^{(K)}(x) \left\{ \Sigma_{\boldsymbol{z}}^{-1/2} \overline{X} \right\}^K \right| dx \le \frac{\|\Sigma_{\boldsymbol{z}}^{-1/2}\|^K \mathbf{E}(\|X\|^K)}{\sqrt{K!}(k-1)^{K/2}}.
\end{aligned}$$

Similar inequality for I_Y. Consider also

$$\begin{aligned}
\sum\nolimits_{k=n_0+1}^{n} \frac{1}{(k-1)^{K/2}} &\le \int_{n_0-1}^{+\infty} \frac{1}{t^{K/2}} dt = \frac{2}{K-2} \frac{1}{(n_0-1)^{(K-2)/2}} \\
&\le \frac{2}{K-2} \frac{1}{(n/6)^{(K-2)/2}}.
\end{aligned}$$

Hence,

$$
\begin{aligned}
&\sum\nolimits_{k=n_0+1}^{n} \left|\mathbf{P}\left\{W_k+\overline{X}\in B\right\}-\mathbf{P}\left\{W_k+\overline{Y}\in B\right\}\right| \\
&\leq \frac{\|\Sigma_{\boldsymbol{z}}^{-1/2}\|^K\beta}{\sqrt{K!}}\frac{2}{K-2}\frac{6^{(K-2)/2}}{n^{(K-2)/2}} \\
&\leq 2\frac{\|\Sigma_{\boldsymbol{z}}^{-1/2}\|^K\beta}{n^{(K-2)/2}}=\frac{2C_{\boldsymbol{z}}^K\beta}{n^{(K-2)/2}}.
\end{aligned}
$$

Induction step, collection of all parts together

Here we sum the bounds (A.20), (A.25), (A.32), (A.39), (A.43) and finalize the proof. Denote $\delta \stackrel{\text{def}}{=} \frac{\beta}{n^{(K-2)/2}}$.

$$
\begin{aligned}
\Delta_n \leq{}& \varepsilon C_{\mathscr{B}}C_{\boldsymbol{z}}+\frac{1.5\widetilde{C}_\phi C_{\boldsymbol{z}}^2}{(K-2)!}\frac{\delta}{(\varepsilon/2)^{(K-2)}}\left\{2MC_{\mathscr{B}}C_\phi\left(C_{\boldsymbol{z}}^K\delta\right)^{1/(K-2)}+\varepsilon C_{\mathscr{B}}C_{\boldsymbol{z}}\right\} \\
&+\frac{4\sqrt{2}\left(\sqrt{m-1}-1\right)\widetilde{C}_{\phi,1}C_{\boldsymbol{z}}}{\varepsilon n^{1/2}K!2^{1-K}}\frac{\delta}{\varepsilon^{(K-2)}}\left\{2MC_{\mathscr{B}}C_\phi\left(C_{\boldsymbol{z}}^K\delta\right)^{1/(K-2)}+\varepsilon C_{\mathscr{B}}C_{\boldsymbol{z}}\right\} \\
&+\frac{2\sqrt{2}C_{\boldsymbol{z}}^K\beta}{\sqrt{K!}(m-1)^{(K-2)/2}}MC_{\mathscr{B}}C_\phi\left(C_{\boldsymbol{z}}^K\delta\right)^{1/(K-2)}+\frac{2^{(K-2)/2}}{\sqrt{K!}}C_{\boldsymbol{z}}^K\delta+2C_{\boldsymbol{z}}^K\delta.
\end{aligned}
$$

Taking $\varepsilon=(C_{\boldsymbol{z}}^2C_\phi\delta)^{1/(K-2)}\mathbf{a}$ and $\sqrt{m-1}=\mathbf{b}\left\{C_{\boldsymbol{z}}^K\beta\right\}^{1/(K-2)}$ for some $\mathbf{a},\mathbf{b}>0$, and using that $K\geq 3$, $C_{\mathscr{B}},C_\phi\geq 1$, $C_\phi\geq\widetilde{C}_{\phi,1},\widetilde{C}_\phi$, we obtain

$$
\begin{aligned}
\Delta_n \leq{}& C_\phi C_{\mathscr{B}}\left(C_{\boldsymbol{z}}^K\delta\right)^{1/(K-2)}M\Bigg[\frac{\mathbf{a}}{M}+\frac{1.5(\mathbf{a}/2)^{-(K-2)}}{(K-2)!}\frac{(2M+\mathbf{a})}{M} \\
&+\frac{4\sqrt{2}\left(\sqrt{m-1}-1\right)}{(C_{\boldsymbol{z}}^K\beta)^{1/(K-2)}}\frac{(\mathbf{a}/2)^{-(K-1)}}{K!}\frac{(2M+\mathbf{a})}{M}+\frac{2\sqrt{2}C_{\boldsymbol{z}}^K\beta}{\sqrt{K!}(m-1)^{(K-2)/2}} \\
&+2.6\left\{C_{\boldsymbol{z}}^K\delta\right\}^{(K-3)/(K-2)}\frac{1}{M}\Bigg] \\
\leq{}& C_\phi C_{\mathscr{B}}\left(C_{\boldsymbol{z}}^K\delta\right)^{1/(K-2)}M\Bigg[\frac{\mathbf{a}}{M}+\frac{1.5(\mathbf{a}/2)^{-(K-2)}}{(K-2)!}\frac{(2M+\mathbf{a})}{M} \\
&+\frac{4\sqrt{2}\mathbf{b}}{(\mathbf{a}/2)^{(K-1)}}\frac{(2M+\mathbf{a})}{MK!}+\frac{2\sqrt{2}}{\sqrt{K!}\mathbf{b}^{K-2}}+2^{(K-3)/(K-2)}\frac{2.6}{M^{K-2}}\Bigg] \\
\leq{}& C_\phi C_{\mathscr{B}}\left(C_{\boldsymbol{z}}^K\delta\right)^{1/(K-2)}M,
\end{aligned}
$$

here we also used the condition $n^{1/2}>M\left\{C_{\boldsymbol{z}}^K\beta/2\right\}^{1/(K-2)}$ from the induction basis; the last inequality holds for parameters $M\geq 72.5$, $\mathbf{a}=M/1.95$, $\mathbf{b}=9.25$. Moreover, for these values $1<m\leq n/2$, as it is assumed in the proof. $\square$

A.3. Proof of Theorem 2.2

Let us take a smoothing function $\widetilde{\varphi}:\mathbb{R}^p\mapsto\mathbb{R}$ such that it satisfies (cf. (A.8) and (A.9) in Lemma A.3)

$$0\le\widetilde{\varphi}(x)\le 1,\quad \widetilde{\varphi}(x)=\begin{cases}1, & x\in B;\\ 0, & x\notin B^{\varepsilon},\end{cases}\quad \left|\widetilde{\varphi}^{(K)}(x)h^{K}\right|\le\frac{C_{\phi,2}\|h\|^{K}}{\varepsilon^{K}},$$

for some positive constant $C_{\phi,2}\in\mathbb{R}$ and for all $x,h\in\mathbb{R}^p$. For example, one can take $\widetilde{\varphi}(x)=\phi(\widetilde{\rho}(x)/\widetilde{\varepsilon})$, as in (A.11) in Lemma A.3 with function $\phi(x)$ K times continuously differentiable

$$0\le\phi(x)\le 1,\quad \phi(x)=0\text{ for } x\ge 1,\quad \phi(x)=1\text{ for } x\le 0\,.$$

Applying the smoothing Lemma A.4 and Lemma A.2, we have

$$\begin{aligned}\varDelta_n &\stackrel{\text{def}}{=}\sup\nolimits_{B\in\mathscr{B}}\left|\mathbf{P}\left(S_n\in B\right)-\mathbf{P}(\widetilde{S}_n\in B)\right|\\ &\le\sup\nolimits_{B\in\mathscr{B}}\left|\mathbf{E}\widetilde{\varphi}\left(S_n\right)-\mathbf{E}\widetilde{\varphi}(\widetilde{S}_n)\right|+\sup_{B\in\mathscr{B}}\mathbf{P}\big(\widetilde{S}_n\in B^{\varepsilon}\setminus B\big) &\text{(A.44)}\\ &\le\sup\nolimits_{B\in\mathscr{B}}\left|\mathbf{E}\widetilde{\varphi}\left(S_n\right)-\mathbf{E}\widetilde{\varphi}(\widetilde{S}_n)\right|+\varepsilon\widetilde{C}_{\mathscr{B}}\overline{C}_{\boldsymbol{z}}. &\text{(A.45)}\end{aligned}$$

Denote

$$\overline{X}_i\stackrel{\text{def}}{=}\frac{X_i}{\sqrt{n}},\quad \overline{Y}_i\stackrel{\text{def}}{=}\frac{Y_i}{\sqrt{n}},$$

then it holds

$$S_n\stackrel{\text{def}}{=}\overline{X}_1+\dots+\overline{X}_n,\quad \widetilde{S}_n\stackrel{\text{def}}{=}\overline{Y}_1+\dots+\overline{Y}_n.$$

Below we employ the telescopic sum approach by [25]:

$$\begin{aligned}W_k&\stackrel{\text{def}}{=}\sum\nolimits_{i=1}^{k-1}\overline{Y}_i+\sum\nolimits_{i=k+1}^{n}\overline{X}_i\ \text{ for } k=2,\dots,n\,,\\ W_1&\stackrel{\text{def}}{=}\overline{X}_2+\dots+\overline{X}_n,\quad W_n\stackrel{\text{def}}{=}\overline{Y}_1+\dots+\overline{Y}_{n-1},\end{aligned}$$

thus

$$S_n=\overline{X}_1+W_1,\quad \widetilde{S}_n=W_n+\overline{Y}_n,$$

here W_k is independent of $\overline{X}_k$ and $\overline{Y}_k$. Introduce

$$\widetilde{\gamma}_k\stackrel{\text{def}}{=}\left|\mathbf{E}\widetilde{\varphi}\left(W_k+\overline{X}_k\right)-\mathbf{E}\widetilde{\varphi}\left(W_k+\overline{Y}_k\right)\right|,$$

it holds

$$\left|\mathbf{E}\widetilde{\varphi}\left(S_n\right)-\mathbf{E}\widetilde{\varphi}(\widetilde{S}_n)\right|\le\sum\nolimits_{k=1}^{n}\widetilde{\gamma}_k. \tag{A.46}$$

By Taylor's formula (A.7)

$$\widetilde{\varphi}\left(W_k+\overline{X}_k\right)=\widetilde{\varphi}\left(W_k\right)+\sum_{j=1}^{K-1}\frac{\widetilde{\varphi}^{(j)}\left(W_k\right)\overline{X}_k^j}{j!}+\mathbf{E}\frac{(1-\tau)^{K-1}}{(K-1)!}\widetilde{\varphi}^{(K)}\left(W_k+\tau\overline{X}_k\right)\overline{X}_k^K.$$

Similar expression for $\widetilde{\varphi}\left(W_k+\overline{Y}_k\right)$ together with independence of $W_k, \overline{X}_k, \overline{Y}_k, \tau$ and condition (2.1) imply

$$\widetilde{\gamma}_k \le C_{\phi,2}\frac{\mathbf{E}\left(\|X_k\|^K+\|Y_k\|^K\right)}{K!\varepsilon^K n^{K/2}}. \tag{A.47}$$

Collecting (A.45), (A.46) and (A.47) implies

$$\begin{aligned}\varDelta_n &\le \frac{C_{\phi,2}}{K!\varepsilon^K n^{K/2}}\sum\nolimits_{k=1}^n \mathbf{E}\left(\|X_k\|^K+\|Y_k\|^K\right)+\varepsilon\widetilde{C}_{\mathscr{B}}\overline{C}_{\boldsymbol{z}}\\ &\le 2\left\{\frac{C_{\phi,2}(\widetilde{C}_{\mathscr{B}}\overline{C}_{\boldsymbol{z}})^K}{n^{K/2}(K-1)!}\sum\nolimits_{k=1}^n \mathbf{E}\left(\|X_k\|^K+\|Y_k\|^K\right)\right\}^{1/(K+1)},\end{aligned}$$

which finishes the proof of Theorem 2.2.

A.4. Proof of Theorem 2.3

repeats the main steps of the proof of Theorem 2.2 except for the smoothing (A.44). Using the notation similar to the proof of Theorem 2.2: $\overline{X}_k' \stackrel{\text{def}}{=} X_k'/\sqrt{n}$,

$$\begin{aligned}W_k' &\stackrel{\text{def}}{=} \sum\nolimits_{i=1}^{k-1}\overline{X}_k'+\sum\nolimits_{i=k+1}^{n}\overline{X}_i \quad \text{for } k=2,\dots,n,\\ W_1' &\stackrel{\text{def}}{=} \overline{X}_2+\dots+\overline{X}_n, \quad W_n' \stackrel{\text{def}}{=} \overline{X}_1'+\dots+\overline{X}_{n-1}',\end{aligned}$$

we can write

$$\left|\mathbf{E}f\left(S_n\right)-\mathbf{E}f(S_n')\right| \le \sum\nolimits_{k=1}^n\left|\mathbf{E}f\left(W_k'+\overline{X}_k\right)-\mathbf{E}f\left(W_k'+\overline{X}_k'\right)\right|.$$

By Taylor's formula (A.7)

$$f\left(W_k'+\overline{X}_k\right)=f\left(W_k'\right)+\sum_{j=1}^{K-1}\frac{f^{(j)}\left(W_k'\right)\overline{X}_k^j}{j!}+\mathbf{E}\frac{(1-\tau)^{K-1}}{(K-1)!}f^{(K)}\left(W_k'+\tau\overline{X}_k\right)\overline{X}_k^K.$$

Similar expression for $f\left(W_k'+\overline{X}_k'\right)$, independence of $W_k', \overline{X}_k, \overline{X}_k', \tau$ and condition (2.6) imply

$$\left|\mathbf{E}f\left(W_k'+\overline{X}_k\right)-\mathbf{E}f\left(W_k'+\overline{X}_k'\right)\right| \le C_f\frac{\mathbf{E}\left(\|X_k\|^K+\|X_k'\|^K\right)}{K!n^{K/2}},$$

therefore

$$\left|\mathbf{E}f\left(S_n\right)-\mathbf{E}f(S_n')\right| \le \sum\nolimits_{k=1}^n C_f\frac{\mathbf{E}\left(\|X_k\|^K+\|X_k'\|^K\right)}{K!n^{K/2}}.$$

If $\{X_i\}$ are i.i.d., and $\{X_i'\}$ are also identically distributed, we have

$$\left|\mathbf{E}f\left(S_n\right) - \mathbf{E}f(S_n')\right| \le C_f \frac{\mathbf{E}\left(\|X_1\|^K + \|X_1'\|^K\right)}{K!n^{(K-2)/2}}.$$

This finishes the proof of Theorem 2.3.

Appendix B: Proofs of the properties of the weighted bootstrap

This section contains proofs of the results from Section 3.

Proof of Theorem 3.1. Consider the following approximating random vector

$$\widetilde{S}_n \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^{n} X_i \varepsilon_i.$$

It holds

$$\mathbf{E}(X_i^k) = \mathbf{E}\{(X_i\varepsilon_i)^k\} = \mathbf{E}(Y_i^k) \;\; \forall k = 1, 2, 3,$$

here the expectation is taken w.r.t. the joint probability distribution of X_i and ε_i. Hence, applying Theorem 2.1 and triangle ineqiality, we obtain the following approximating bound between distributions of $\|S_n\|$ and $\|\widetilde{S}_n\|$:

$$\begin{aligned}
&\sup_{t\ge 0}\left|\mathbf{P}\left(\|S_n\| \le t\right) - \mathbf{P}\big(\|\widetilde{S}_n\| \le t\big)\right| \\
&\le M C_{\mathscr{B}} C_\phi C_{\boldsymbol{z}}^2 \frac{\left\{\mathbf{E}\left(\|X_1\|^4\right) + \mathbf{E}\left(\|Y_1\|^4\right)\right\}^{1/2}}{n^{1/2}} \\
&+ M C_{\mathscr{B}} C_\phi C_{\boldsymbol{z}}^2 \frac{\left\{\mathbf{E}\left(\|X_1\|^4\right)\mathbf{E}(\varepsilon_1^4) + \mathbf{E}\left(\|Y_1\|^4\right)\right\}^{1/2}}{n^{1/2}} \\
&\le \varDelta_1 \stackrel{\text{def}}{=} 2 M C_{\mathscr{B}} C_\phi C_{\boldsymbol{z}}^2 \frac{\left\{\mathbf{E}\left(\|X_1\|^4\right)\mathbf{E}(\varepsilon_1^4) + \mathbf{E}\left(\|Y_1\|^4\right)\right\}^{1/2}}{n^{1/2}}.
\end{aligned} \tag{B.1}$$

Introduce the upper quantile function for the r.v. $\|\widetilde{S}_n\|$:

$$\widetilde{Q}(\alpha) \stackrel{\text{def}}{=} \inf\left\{t \in \mathbb{R} : \mathbf{P}\left(\|\widetilde{S}_n\| > t\right) \le \alpha\right\}. \tag{B.2}$$

Bound (B.1) and definitions (3.1), (B.10) imply

$$Q(\alpha + 2\varDelta_1) \le \widetilde{Q}(\alpha + \varDelta_1) \le Q(\alpha). \tag{B.3}$$

By definitions (1.3), (B.10), and by inequalities (B.3) we infer

$$Q(\alpha + \varDelta_1) \le Q^{\circ}(\alpha) \le Q(\alpha - \varDelta_1) + \varDelta_{Q^{\circ}},$$

where

$$\varDelta_{Q^{\circ}} \stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } \mathbf{P}^{\circ}\left(\|S_n^{\circ}\| > t\right) \text{ is continuous in } Q^{\circ}(\alpha); \\ \varDelta_1, & \text{otherwise.}\end{cases}$$

Let us define similarly

$$\Delta_2 \stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } \mathbf{P}\left(\|S_n\| > t\right) \text{ is continuous in } Q(\alpha - \Delta_1); \\ \Delta_1, & \text{otherwise.} \end{cases}$$

Collecting the derived bounds, we have

$$\begin{aligned} \alpha + \Delta_1 &\geq \mathbf{P}\left(\|S_n\| > Q(\alpha + \Delta_1)\right) \geq \mathbf{P}\left(\|S_n\| > Q^{\circ}(\alpha)\right), \\ \alpha - \Delta_1 &\leq \mathbf{P}\left(\|S_n\| > Q(\alpha - \Delta_1) - \Delta_2\right) \\ &\leq \mathbf{P}\left(\|S_n\| > Q(\alpha - \Delta_1) + \Delta_{Q^{\circ}}\right) + \Delta_1 + (\Delta_2 + \Delta_{Q^{\circ}})\widetilde{C}_{\mathscr{B}} C_{\boldsymbol{z}} \qquad \text{(B.4)} \\ &\leq \mathbf{P}\left(\|S_n\| > Q^{\circ}(\alpha)\right) + \Delta_1(1 + 2\widetilde{C}_{\mathscr{B}} C_{\boldsymbol{z}}), \end{aligned}$$

where inequality (B.4) follows from Theorem 2.1, Lemma A.2, and bound (B.1). Indeed, for arbitrary $\varepsilon > 0$:

$$\begin{aligned} \mathbf{P}\left(t < \|S_n\| \leq t + \varepsilon\right) &\leq \mathbf{P}\left(t < \|n^{-1/2}\textstyle\sum_{i=1}^n Y_i\| \leq t + \varepsilon\right) + 2\Delta_1/2 \\ &\leq \varepsilon \widetilde{C}_{\mathscr{B}} C_{\boldsymbol{z}} + \Delta_1. \end{aligned} \qquad \text{(B.5)}$$

□

Proof of Theorem 3.2. is similar to the proof of Theorem 3.1, except that the bound (B.1) is replaced with the following one:

$$\begin{aligned} &\sup_{t \geq 0} \left| \mathbf{P}\left(\|S_n\| \leq t\right) - \mathbf{P}\big(\|\widetilde{S}_n\| \leq t\big) \right| \\ &\leq \Delta_2 \stackrel{\text{def}}{=} 2.8 \left\{ C_{\phi,2} \widetilde{C}_{\mathscr{B}}^4 \right\}^{1/5} \left\{ \frac{\overline{C}_{\boldsymbol{z}}^4 \sum_{i=1}^n \left[\mathbf{E}\left(\|X_i\|^4\right) \mathbf{E}(\varepsilon_1^4) + \mathbf{E}\left(\|Y_i\|^4\right)\right]}{n^2} \right\}^{1/5}. \end{aligned}$$

This inequality follows from Theorem 2.2 for $K = 4$, and from triangle inequality. □

Proof of Theorem 3.3. The statement follows from Theorem 3.2 applied to $X_i \coloneqq (\Psi\Psi^\top)^{-1/2}\Psi_i\epsilon_i$, and $\|S_n\| \coloneqq T$, $\|S_n^{\circ}\| \coloneqq T^{\circ}$. □

B.1. Conditions and proof of Theorem 3.4

Recall that $\ell_i(\theta) \stackrel{\text{def}}{=} \log\left(\frac{d\mathbf{P}_\theta}{d\mu_0}(y_i)\right)$, $d_0^2 \stackrel{\text{def}}{=} -\mathbf{E}\ell_1''(\theta^*)$. By defintion of θ^*, vectors $\ell_i'(\theta^*)$ are i.i.d. with zero mean. Below are the conditions for Theorem 3.4, they are required for the Wilks type bounds for the log-likelihood ratios $L(\widetilde{\theta}) - L(\theta^*)$ and $L^{\circ}(\widetilde{\theta}^{\circ}) - L^{\circ}(\widetilde{\theta})$, obtained in [45, 46] and [47] correspondingly. We took the set of conditions below from Section B.3.1 of supplement of [47].

1. The covariance matrix $\operatorname{Var}\{\ell_i'(\theta^*)\}$ is positive definite;

2. There exist a positive-definite symmetric matrix v_0^2 and constants $\mathtt{g} > 0, \nu \geq 1$ such that $\operatorname{Var}\{\ell_i'(\theta^*)\} \leq v_0^2$ and for all $|\lambda| \leq \overline{\mathtt{g}}$
$$\sup_{\substack{\gamma\in\mathbb{R}^p\\ \|\gamma\|=1}} \log \mathbf{E} \exp\left\{\lambda\gamma^\top v_0^{-1}\ell_i'(\theta^*)\right\} \leq \nu^2\lambda^2/2;$$
3. There exist a constant $\varpi > 0$ and for each $r > 0$ a constant $\overline{\mathtt{g}}_2(r)$ such that it holds for all $\theta \in \Theta$ such that $\|d_0(\theta-\theta^*)\| \leq r/\sqrt{n}$, for $j = 1, 2$, and for all $|\lambda| \leq \overline{\mathtt{g}}_2(r)$
$$\sup_{\substack{\gamma_j\in\mathbb{R}^p\\ \|\gamma_j\|=1}} \log \mathbf{E} \exp\left\{\varpi^{-1}\lambda\gamma_1^\top d_0^{-1}(\ell_i''(\theta) - \mathbf{E}\ell_i''(\theta))d_0^{-1}\gamma_2\right\} \leq \nu^2\lambda^2/2;$$
4. There exists a constant $\mathfrak{a} > 0$ s.t. $\mathfrak{a}^2 d_0^2 \geq v_0^2$;
5. There exists a constant $\mathtt{C}_{3m} \geq 0$ such that for each $r \geq 0$, and for all $\theta \in \Theta$ such that $\|d_0(\theta-\theta^*)\| \leq r/\sqrt{n}$
$$\sup_{\substack{\gamma\in\mathbb{R}^p\\ \|\gamma\|=1}} \|d_0^{-1}\gamma^\top \mathbf{E}\ell_i'''(\theta)d_0^{-1}\| \leq \mathtt{C}_{3m};$$
6. For each $r \geq C\sqrt{p+x}$ and for all $\theta \in \Theta$ such that $\|d_0(\theta-\theta^*)\| \leq r/\sqrt{n}$, it holds for some value $\mathtt{C}_{\mathtt{b}}(r) > 0$ s.t. $r\mathtt{C}_{\mathtt{b}}(r) \to +\infty$ with $r \to +\infty$
$$\left\|d_0^{-1}\mathbf{E}\ell_i''(\theta)d_0^{-1}\right\| \geq \mathtt{C}_{\mathtt{b}}(r);$$
7. For the bootstrap weights $\{u_i\}_{i=1}^n$ it holds for all $|\lambda| \leq \overline{\mathtt{g}}$
$$\log \mathbf{E} \exp\{\lambda(u_i - 1)\} \leq \nu^2\lambda^2/2.$$

Proof of Theorem 3.4. Take $S_n := n^{-1/2}\sum_{i=1}^n d_0^{-1}\ell_i'(\theta^*)$, and $S_n^{\circ} := n^{-1/2}\sum_{i=1}^n d_0^{-1}\ell_i'(\theta^*)\varepsilon_i$. Let also $S_n^{\circ}(\widetilde{\theta}) := n^{-1/2}\sum_{i=1}^n d_0^{-1}\ell_i'(\widetilde{\theta})\varepsilon_i$.

[45, 46] showed that under the conditions given above the following non-asymptotic Wilks type inequality ([49]) holds with probability $\geq 1 - 5\mathrm{e}^{-x}$:
$$\left|\sqrt{2\{L(\widetilde{\theta}) - L(\theta^*)\}} - \|S_n\|\right| \leq \Delta_{\mathrm{W}}(x), \tag{B.6}$$
where $\Delta_{\mathrm{W}}(x) \leq 3\sqrt{p+x}n^{-1/2}\left[C\sqrt{p+x} + 6\nu\varpi(\overline{\mathtt{g}}^2 n)^{-1/2}(2\sqrt{p}+\sqrt{2x}+4p\{x(\overline{\mathtt{g}}^2 n)^{-1}+1\})\right] \leq C(p+x)/\sqrt{n}$.

[47] obtained the bootstrap version of (B.6). If the conditions above are fulfilled, then the following bounds hold with probability $\geq 1 - 5\mathrm{e}^{-x}$:
$$\left|\sqrt{2\{L^{\circ}(\widetilde{\theta}^{\circ}) - L^{\circ}(\widetilde{\theta})\}} - \|S_n^{\circ}(\widetilde{\theta})\|\right| \leq \Delta_{\mathrm{W}}^{\circ}(x),$$
$$\left|\|S_n^{\circ}(\widetilde{\theta})\| - \|S_n^{\circ}\|\right| \leq \Delta_{\xi}^{\circ}(x),$$

where $\Delta^{\circ}_{\mathrm{W}}(x) \le 2\Delta_{\mathrm{W}}(x) + C\nu(p+x)/\sqrt{n}$, $\Delta^{\circ}_{\xi}(x) \le C\nu(p+x)/\sqrt{n}$. Using these bounds, we infer

$$\begin{aligned}
&\mathbf{P}\left(\sqrt{2\{L(\widetilde{\theta}) - L(\theta^*)\}} > t\right) \\
&\ge \mathbf{P}\left(\|S_n\| > t + \Delta_{\mathrm{W}}(x)\right) \\
&\ge \mathbf{P}\left(\|S^{\circ}_n\| > t - \Delta^{\circ}_{\mathrm{W}}(x) - \Delta^{\circ}_{\xi}(x)\right) - \Delta_L \\
&\ge \mathbf{P}\left(\sqrt{2\{L^{\circ}(\widetilde{\theta}^{\circ}) - L^{\circ}(\widetilde{\theta})\}} > t\right) - \Delta_L,
\end{aligned} \tag{B.7}$$

where

$$\begin{aligned}
\Delta_L &\stackrel{\text{def}}{=} \widetilde{\Delta}_L + \widetilde{C}_{\mathscr{B}} C_{\boldsymbol{z},L}\{\Delta_{\mathrm{W}}(x) + \Delta^{\circ}_{\mathrm{W}}(x) + \Delta^{\circ}_{\xi}(x)\}, \\
\widetilde{\Delta}_L &\stackrel{\text{def}}{=} 2MC_{\mathscr{B}}C_{\phi}C^2_{\boldsymbol{z},L}\frac{\left\{\mathbf{E}\left(\|d_0^{-1}\ell'_i(\theta^*)\|^4\right)\mathbf{E}(\varepsilon_1^4) + \mathbf{E}\left(\|Y_1\|^4\right)\right\}^{1/2}}{n^{1/2}};
\end{aligned} \tag{B.8}$$

inequality (B.7) follows from Theorem 3.1 and Lemma A.2 (similarly to bounds (B.5)). Indeed, it holds for arbitrary $\varepsilon > 0$

$$\begin{aligned}
&\mathbf{P}\left(\|S_n\| > t + \varepsilon\right) \\
&\ge \mathbf{P}\left(\|n^{-1/2}\textstyle\sum_{i=1}^n Y_i\| > t + \varepsilon\right) - \widetilde{\Delta}_L/2 \\
&\ge \mathbf{P}\left(\|n^{-1/2}\textstyle\sum_{i=1}^n Y_i\| > t\right) - \widetilde{\Delta}_L/2 - \varepsilon\widetilde{C}_{\mathscr{B}}C_{\boldsymbol{z},L} \\
&\ge \mathbf{P}\left(\|S^{\circ}_n\| > t\right) - \widetilde{\Delta}_L - \varepsilon\widetilde{C}_{\mathscr{B}}C_{\boldsymbol{z},L}.
\end{aligned}$$

Similar inequalities in the inverse direction imply

$$\left|\mathbf{P}\left(\sqrt{2\{L(\widetilde{\theta}) - L(\theta^*)\}} > t\right) - \mathbf{P}\left(\sqrt{2\{L^{\circ}(\widetilde{\theta}^{\circ}) - L^{\circ}(\widetilde{\theta})\}} > t\right)\right| \le \Delta_L. \tag{B.9}$$

Denote

$$\widetilde{Q}_L(\alpha) \stackrel{\text{def}}{=} \inf\left\{t \ge 0 : \mathbf{P}\left(\sqrt{2\{L^{\circ}(\widetilde{\theta}^{\circ}) - L^{\circ}(\widetilde{\theta})\}} > t\right) \le \alpha\right\}. \tag{B.10}$$

Due to (B.9)

$$Q_L(\alpha + 2\Delta_L) \le \widetilde{Q}_L(\alpha + \Delta_L) \le Q_L(\alpha). \tag{B.11}$$

Furthermore, by definitions (3.5), (B.10) and by inequality (B.11), it holds

$$Q_L(\alpha + \Delta_L) \le Q^{\circ}_L(\alpha) \le Q_L(\alpha - \Delta_L) + \Delta_{Q^{\circ}_L},$$

where

$$\Delta_{Q^{\circ}_L} \stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } \mathbf{P}^{\circ}\left(\sqrt{2\{L^{\circ}(\widetilde{\theta}^{\circ}) - L^{\circ}(\widetilde{\theta})\}} > t\right) \text{ is continuous in } Q^{\circ}_L(\alpha); \\ \Delta_L, & \text{otherwise.} \end{cases}$$

Denote similarly

$$\varDelta_3 \stackrel{\mathrm{def}}{=} \begin{cases} 0, & \text{if } \mathbf{P}\left(\sqrt{2\{L(\widetilde{\theta})-L(\theta^*)\}} > t\right) \text{ is continuous in } Q_L(\alpha-\varDelta_L); \\ \varDelta_L, & \text{otherwise.} \end{cases}$$

Collecting the derived bounds, we obtain

$$\begin{aligned} \alpha+\varDelta_L &\geq \mathbf{P}\left(\sqrt{2\{L(\widetilde{\theta})-L(\theta^*)\}} > Q_L(\alpha+\varDelta_L)\right) \\ &\geq \mathbf{P}\left(\sqrt{2\{L(\widetilde{\theta})-L(\theta^*)\}} > Q_L^{\circ}(\alpha)\right), \\ \alpha-\varDelta_L &\leq \mathbf{P}\left(\sqrt{2\{L(\widetilde{\theta})-L(\theta^*)\}} > Q_L(\alpha-\varDelta_L)-\varDelta_3\right) \\ &\leq \mathbf{P}\left(\sqrt{2\{L(\widetilde{\theta})-L(\theta^*)\}} > Q_L^{\circ}(\alpha)\right) + 2\varDelta_{\mathrm{W}}(x) + 2\widetilde{\varDelta}_L\big(1+\widetilde{C}_{\mathscr{B}}C_{\boldsymbol{z},L}\big), \end{aligned}$$

the last inequality follows from (B.6), Theorem 2.1 and Lemma A.2, indeed, it holds for any $\varepsilon > 0$:

$$\begin{aligned} &\mathbf{P}\left(t < \sqrt{2\{L(\widetilde{\theta})-L(\theta^*)\}} \leq t+\varepsilon\right) \\ &\leq \mathbf{P}\left(t < \|S_n\| \leq t+\varepsilon\right) + 2\varDelta_{\mathrm{W}}(x) \\ &\leq \mathbf{P}\left(t < \|n^{-1/2}\textstyle\sum_{i=1}^n Y_i\| \leq t+\varepsilon\right) + 2\varDelta_{\mathrm{W}}(x) + \widetilde{\varDelta}_L \\ &\leq \varepsilon\widetilde{C}_{\mathscr{B}}C_{\boldsymbol{z},L} + 2\varDelta_{\mathrm{W}}(x) + 2\widetilde{\varDelta}_L. \end{aligned}$$

□

Acknowledgments

I am cordially grateful to Prof. Gilles Blanchard, Prof. Moritz Jirak, and Prof. Vladimir Ulyanov for helpful discussions about the topic of the paper. I would especially like to thank Prof. Vladimir Koltchinskii for valuable comments, which helped to improve the paper.

References

[1] 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.
[2] Ball, K. (1993). The reverse isoperimetric problem for Gaussian measure. *Discrete & Computational Geometry*, 10(1):411–420.
[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 Rao, R. R. (1986). *Normal approximation and asymptotic expansions*, volume 64. SIAM.

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

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

[9] Chen, L. H. and Fang, X. (2011). Multivariate normal approximation by Stein's method: The concentration inequality approach. *arXiv preprint arXiv:1111.4073*.

[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. (2014). Central limit theorems and bootstrap in high dimensions. *To appear in the Annals of Probability. arXiv preprint arXiv:1412.3661*.

[12] 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.

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

[14] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *The Annals of Probability*, pages 724–739.

[15] Götze, F. and Zaitsev, A. Y. (2014). Explicit rates of approximation in the CLT for quadratic forms. *The Annals of Probability*, 42(1):354–397.

[16] Grad, H. (1949). Note on n-dimensional Hermite polynomials. *Communications on Pure and Applied Mathematics*, 2(4):325–330.

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

[18] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, pages 1926–1947.

[19] Harville, D. A. (1971). On the distribution of linear combinations of non-central chi-squares. *The Annals of Mathematical Statistics*, 42(2):809–811.

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

[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? *Annals of statistics*, pages 768–806.

[23] Kane, D. M. (2011). The Gaussian surface area and noise sensitivity of degree-d polynomial threshold functions. *computational complexity*, 20(2):389–412.

[24] Klivans, A. R., O'Donnell, R., and Servedio, R. A. (2008). Learning geomet-

ric concepts via Gaussian surface area. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 541–550. IEEE.

[25] Lindeberg, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15(1):211–225.

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

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

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

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

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

[31] Mason, D. M. and Newton, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *The Annals of Statistics*, pages 1611–1624.

[32] Nagaev, S. (1976). An estimate of the remainder term in the multidimensional central limit theorem. In *Proceedings of the Third Japan-USSR Symposium on Probability Theory*, pages 419–438. Springer.

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

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

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

[36] Press, S. J. (1966). Linear combinations of non-central chi-square variates. *The Annals of Mathematical Statistics*, pages 480–487.

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

[38] Rubin, D. B. (1981). The bayesian bootstrap. *The Annals of Statistics*, pages 130–134.

[39] Sazonov, V. V. (1972). On a bound for the rate of convergence in the multidimensional central limit theorem. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, pages 563–581, Berkeley, Calif. University of California Press.

[40] Sazonov, V. V. (1981). *Normal approximation-some recent advances.* Springer.

[41] Senatov, V. (1981). Uniform estimates of the rate of convergence in the multi-dimensional central limit theorem. *Theory of Probability & Its Applications*, 25(4):745–759.

[42] Senatov, V. V. (1998). *Normal approximation: new results, methods and*

problems. Walter de Gruyter.

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

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

[45] Spokoiny, V. (2012). Parametric estimation. Finite sample theory. *The Annals of Statistics*, 40(6):2877–2909.

[46] Spokoiny, V. (2013). Bernstein-von Mises Theorem for growing parameter dimension. *arXiv preprint arXiv:1302.3430*.

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

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

[49] Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.*, 9:60–62.

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

[51] 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.

[52] Zolotarev, V. M. (1997). *Modern theory of summation of random variables*. Walter de Gruyter.