

Optimal-order bounds on the rate of convergence to normality in the multivariate delta method

Iosif Pinelis and Raymond Molzon

*Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
e-mail: ipinelis@mtu.edu; remolzon@mtu.edu*

Abstract: Uniform and nonuniform Berry–Esseen (BE) bounds of optimal orders on the rate of convergence to normality in the delta method for vector statistics are obtained. The results are applicable almost as widely as the delta method itself – except that, quite naturally, the order of the moments needed to be finite is generally $3/2$ times as large as that for the corresponding central limit theorems. Our BE bounds appear new even for the one-dimensional delta method, that is, for smooth functions of the sample mean of univariate random variables. Specific applications to Pearson’s, noncentral Student’s and Hotelling’s statistics, sphericity test statistics, a regularized canonical correlation, and maximum likelihood estimators (MLEs) are given; all these uniform and nonuniform BE bounds appear to be the first known results of these kinds, except for uniform BE bounds for MLEs. The new method allows one to obtain bounds with explicit and rather moderate-size constants. For instance, one has the uniform BE bound $3.61 \mathbb{E}(Y_1^6 + Z_1^6) (1 + \sigma^{-3})/\sqrt{n}$ for the Pearson sample correlation coefficient based on independent identically distributed random pairs $(Y_1, Z_1), \dots, (Y_n, Z_n)$ with $\mathbb{E} Y_1 = \mathbb{E} Z_1 = \mathbb{E} Y_1 Z_1 = 0$ and $\mathbb{E} Y_1^2 = \mathbb{E} Z_1^2 = 1$, where $\sigma := \sqrt{\mathbb{E} Y_1^2 Z_1^2}$.

MSC 2010 subject classifications: Primary 60F05, 60E15, 62F12; secondary 60E10, 62F03, 62F05, 62G10, 62G20.

Keywords and phrases: Berry–Esseen bound, canonical correlation, delta method, rates of convergence, Cramér’s tilt, exponential inequalities, maximum likelihood estimators, noncentral Hotelling’s statistic, noncentral Student’s statistic, nonlinear statistics, Pearson’s correlation coefficient, sphericity test.

Received August 2015.

Contents

1	Introduction	1002
1.1	Notation and conventions	1005
2	General bounds on the convergence rate for smooth nonlinear functions of sums of independent random vectors	1006
3	Applications	1012
3.1	Student’s T	1013

3.2	Pearson's R	1015
3.3	Noncentral Hotelling's T^2 statistic	1017
3.4	Covariance test statistics	1019
3.5	Principal component analysis (PCA)	1023
3.6	Maximum likelihood estimators (MLEs)	1025
4	Proofs	1030
4.1	Toolkit for proofs	1031
4.2	Proofs of bounds in general non-i.d. setting	1033
4.3	Proofs of bounds in i.i.d. setting	1036
4.4	Proof of Lemma 4.5	1038
A	An explicit nonuniform bound in the multivariate delta method	1042
B	Optimality of the restriction $z = O(\sqrt{n})$ for the nonuniform bound	1049
C	Proof of explicit Berry–Esseen bounds for the Pearson statistic	1051
D	Asymptotic behavior of the uniform and nonuniform Berry–Esseen bounds	1055
E	Compactness of the covariance operator	1057
F	On the spectral decomposition of a covariance operator of a random vector in an arbitrary separable Hilbert space	1057
	References	1058

1. Introduction

Initially, we were interested in studying certain properties of the Pitman asymptotic relative efficiency (ARE) between Pearson's, Kendall's, and Spearman's correlation coefficients. As is well known (see e.g. [48]), the standard expression for the Pitman ARE is applicable when the distributions of the corresponding test statistics are close to normality uniformly over a neighborhood of the null set of distributions. Such uniform closeness can usually be provided by Berry–Esseen (BE) type of bounds on the rate of convergence to normality.

BE bounds, especially in the special case of linear statistics, constitute a well-established area of research, which originated mainly in work by Scandinavian authors, who were to a large degree concerned with applications in insurance industry and published many of their results on the accuracy of the normal approximation in actuarial journals. For a small sample of recent uses of BE bounds in various areas of sciences and engineering (again for linear statistics), see e.g. [43, 41, 80, 31, 40].

Kendall's and Spearman's correlation coefficients are instances of U -statistics, for which BE bounds are well known; see e.g. [38]. As for the Pearson statistic (say R), we have not been able to find a BE bound in the literature.

This may not be very surprising, considering that an optimal BE bound for the somewhat similar (and, perhaps, somewhat simpler) Student's statistic was obtained only in 1996, by Bentkus and Götze [6] for independent identically distributed (i.i.d.) random variables (r.v.'s) and by Bentkus, Bloznelis and Götze [4] in the general, non-i.i.d. case. (A necessary and sufficient condition, in the i.i.d. case, for the Student statistic to be asymptotically standard normal was

established only in 1997 by Giné, Götze and Mason [22], and Hall and Wang [26] derive the leading term in the convergence rate in this general setting.)

Employing such simple and standard tools as a delta-method type linearization together with the Chebyshev and Rosenthal inequalities, we quickly obtained (in the i.i.d. case) a uniform bound of the form $O(n^{-1/3})$ for the Pearson statistic. Indeed, Pearson's R can be expressed as $f(\bar{V})$, a smooth nonlinear function of the sample mean $\bar{V} = \frac{1}{n} \sum_{i=1}^n V_i$, where the V_i 's are independent zero-mean random vectors constructed based on the observations of a random sample; cf. (3.7). A natural approximation to $f(\bar{V}) - f(0)$, obtained by the delta method, is the linear statistic $L(\bar{V}) = \sum_{i=1}^n L(\frac{1}{n}V_i)$, where L is the linear functional that is the first derivative of f at the origin. Since BE bounds for linear statistics is a well-studied subject, we are left with estimating the closeness between $f(\bar{V})$ and $L(\bar{V})$. Assuming f is smooth enough, one will have $|f(\bar{V}) - L(\bar{V})|$ on the order of $\|\bar{V}\|^2$, and so, demonstrating the smallness of this remainder term becomes the main problem.

The reader is referred to [75] for a rather detailed description of the delta method and its applications; see [39, 71] for a more modern treatment of the delta method applied to infinite-dimensional random vectors.

While the mentioned uniform bound of the form $O(n^{-1/3})$ (obtained under the assumption $\mathbb{E}\|V_1\|^3 < \infty$) would have sufficed as far as the ARE is concerned, we became interested in obtaining an optimal-rate BE bound for the Pearson statistic, of the form $O(n^{-1/2})$.

Such a bound is obtained in this paper. In fact, we present (in Section 2) general optimal-rate BE bounds (both uniform and nonuniform) for the rate of convergence to normality in the multivariate delta method. The results are applicable almost as widely as the delta method itself – except that, quite naturally, we generally need to require the condition $\mathbb{E}\|V_1\|^3 < \infty$ (or something close to that) to hold, whereas the condition $\mathbb{E}\|V_1\|^2 < \infty$ (or something close to that) is generally needed for the corresponding central limit theorems.

As applications of our general bounds, we provide (in Section 3) uniform and nonuniform BE-type bounds for the Pearson statistic, the noncentral Student and Hotelling statistics, various statistics commonly used in testing hypotheses about a population covariance matrix, the largest eigenvalue of a certain linear operator on an infinite-dimensional Hilbert space, and maximum likelihood estimators (MLEs). No such BE bounds appear to be previously known, except for uniform BE bounds for MLEs. In fact, our BE bounds appear new even for the one-dimensional delta method, that is, for smooth functions of the sample mean of univariate r.v.'s. Moreover, for the Pearson statistic we obtain bounds with explicit constant factors, which are also of moderate sizes.

Our general BE bounds in the multivariate delta method can of course be used in applications other than the ones considered here; we mention a number of other potential applications in Subsections 3.4 and 3.5. In fact, a result from an earlier arXiv version of this paper, similar to Theorem 2.9, was already used in [20]. Of course, our results cannot perfectly cover the entire variety of uses of the delta method; they may require modification or use of different ideas; see e.g. [77, pages 1198 and 1211].

As for the requirement that the observations be identically distributed, it may (and will) be dispensed in general; that is, \bar{V} will in general be replaced by a sum S of independent but not necessarily identically distributed random vectors.

The paper is structured as follows:

- Section 2 contains our main results. Theorems 2.2 and 2.5 present comparatively abstract versions of uniform and nonuniform BE-type bounds in the delta method for smooth enough functions of sums of independent (not necessarily identically distributed) random vectors. These bounds, which are quite explicit, constitute the basis of all the subsequent results in the paper. However, the generality of the conditions in which the bounds of Theorems 2.2 and 2.5 hold and their rather abstract nature may make it somewhat difficult to discern the behavior and quality of the bounds in particular situations of interest. By specializing Theorems 2.2 and 2.5 to the case when the summands are i.i.d. random vectors, we obtain Theorems 2.9 and 2.11, where it becomes clear that the resulting BE-type bounds are of the optimal order $O(1/\sqrt{n})$ in n , whereas the nonuniform versions of the bound also have the optimal order $O(1/z^3)$ of decrease in z assuming the finiteness of the third absolute moment, where z is the number of the asymptotic standard deviations of the nonlinear statistic from its mean. In Theorem 2.10, the i.i.d. setting is specialized even further – to the case when the summands are one-dimensional, that is, real-valued; even in this special case, our results appear to be new.
- In Section 3, Theorems 2.9 and 2.11 are applied in order to obtain optimal-order BE-type bounds for several specific statistics, including the Student, Pearson, and noncentral Hotelling statistics, test statistics for sphericity, certain statistics arising in principal component analysis, and maximum likelihood estimators.
- Section 4 contains the proofs of all results from Section 2, as well as associated lemmas and proofs.

More technical results and proofs are relegated to appendices found at the end of the paper.

- In Appendix A we state and prove Theorem A.2, which presents an explicit nonuniform bound on the rate of convergence in the delta method in the i.i.d. setting.
- The nonuniform bound (2.25) in Theorem 2.9 holds only for $z = O(\sqrt{n})$; in Appendix B we prove that this condition cannot generally be relaxed.
- In Appendix C we prove Corollary 3.8, which provides a simple uniform bound on the distance to normality of Pearson's R .
- In Appendix D we analyze the asymptotic behavior of our uniform and nonuniform bounds in the i.i.d. setting.
- In Appendix E we provide a short, self-contained proof of the compactness of the covariance operator for a random vector taking values in a separable Hilbert space and possessing a finite second moment; this is used in one of

our applications on the principal component of a certain linear operator.

- In Appendix F we outline the proof of the existence of the spectral decomposition for the covariance operator of a random vector taking values in an infinite-dimensional separable Hilbert space.

More details can be found in the online version [65] of this paper. One significant distinction of [65] from the present paper is that there a more general moment condition is considered: $\mathbb{E} \|V_1\|^p < \infty$ for some $p \in (2, 3]$, in place of $\mathbb{E} \|V_1\|^3 < \infty$; of course, then the optimal BE bound is of the form $O(n^{1-p/2})$, in place of $O(n^{-1/2})$. The case $p = 3$ is given the most attention in the existing literature. The other significant difference between the present paper and [65] is the presence there, in Subsection 4.2.1, of explicit uniform and nonuniform BE-type bounds for the self-normalized sum, closely related to the centered Student statistic; the corresponding proofs are rather technical and rely heavily on the use of a computer algebra system.

1.1. Notation and conventions

Before moving on to the main results of the paper, let us adopt some conventions and introduce notation that will be used throughout the paper. We use the abbreviations a.s. for “almost surely”, d.f. for “distribution function”, and i.i.d. for “independent and identically distributed”; r.v. shall always mean a real-valued “random variable”, whereas the phrase “random vector” will not be abbreviated.

The symbol \mathfrak{X} shall be used to denote a separable Banach space (equipped with norm $\|\cdot\|$) of type 2; recall that this means there exists $D \in (0, \infty)$ such that

$$\mathbb{E} \left(\sum_{i=1}^n \varepsilon_i x_i \right)^2 \leq D^2 \sum_{i=1}^n \|x_i\|^2$$

for all vectors x_1, \dots, x_n in \mathfrak{X} , where $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. r.v.’s with $\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = 1/2$. In particular, it is well known and easy to see that, if \mathfrak{X} is a Hilbert space, then \mathfrak{X} is of type 2, and at that one may choose $D = 1$. In particular, any finite-dimensional Euclidean space \mathbb{R}^d is a Hilbert space and thus of type 2, with $D = 1$.

For any \mathfrak{X} -valued random vector ζ , we use the norm notation

$$\|\zeta\|_p := (\mathbb{E} \|\zeta\|^p)^{1/p} \text{ for any real } p \geq 1;$$

the set \mathbb{R} of all real numbers will always be equipped with the canonical norm $|\cdot|$.

The symbols $x \wedge y$ and $x \vee y$ denote the minimum and maximum, respectively, of the real numbers x and y ; also, let $x_+ := 0 \vee x$ denote the positive part of a real number x .

The symbols Φ and φ will denote, respectively, the distribution and density functions of the standard normal distribution.

2. General bounds on the convergence rate for smooth nonlinear functions of sums of independent random vectors

In this section we provide general BE-type bounds on the rate of convergence to normality in the delta method. The more general and abstract results here are explicit bounds on

$$|\mathbb{P}(f(S) \leq \sigma z) - \mathbb{P}(L(S) \leq \sigma z)|,$$

where S is the sum of independent random vectors, f is a smooth enough real-valued function, L is an appropriate linear approximation of f , σ is the standard deviation of $L(S)$, and z is a real number. More specifically, Theorems 2.2 and 2.5 will present, respectively, uniform (in z) and nonuniform (that is, dependent on z) bounds on $|\mathbb{P}(f(S) \leq \sigma z) - \mathbb{P}(L(S) \leq \sigma z)|$. The nonuniform bounds are smaller, and thus better, than the uniform ones in tail zones, for large enough values of z .

Once appropriate bounds on $|\mathbb{P}(f(S) \leq \sigma z) - \mathbb{P}(L(S) \leq \sigma z)|$ are established, one can combine them with any number of well-known BE bounds on the distance $|\mathbb{P}(L(S) \leq \sigma z) - \Phi(z)|$ for the linear statistic $L(S)$ (see e.g. inequality (2.19) in this paper), to immediately obtain bounds on $|\mathbb{P}(f(S) \leq \sigma z) - \mathbb{P}(L(S) \leq \sigma z)|$, that is, on the closeness of the nonlinear statistic $f(S)$ to normality.

After Theorems 2.2 and 2.5 are stated and discussed, we shall let (in (2.20)) $X_i = V_i/n$, where the V_i 's are i.i.d. random vectors, and the distribution of V_1 should be thought of as fixed – that is, not depending on n . Then, accordingly, one will have $S = \sum_i X_i = \bar{V}$ and $f(S) = f(\bar{V})$. This scaling, $X_i = V_i/n$, is precisely what is needed in the delta method, where a smooth function of the sample mean is approximated in a neighborhood of the true mean by a linear function of the sample mean.

Theorem 2.9 and its special case for one dimensional random summands presented in Theorem 2.10 state the existence of what we call non-explicit bounds in the i.i.d. setting; roughly, the uniform bound will be of the form \mathfrak{C}/\sqrt{n} and the nonuniform bound of the form $\mathfrak{C}/(z^3\sqrt{n})$. An explicit expression for \mathfrak{C} can be obtained by careful analysis of the proofs in Section 4. An example of such an explicit uniform bound is found in Theorem 2.11; such an explicit bound can then be used to obtain relatively simple expressions for \mathfrak{C} that depend only on universal constants and a few moments of V_1 and $L(V_1)$ in specific applications (such as Pearson's R , as in Corollary 3.8).

The following smoothness condition is crucial. Take any Borel-measurable functional $f: \mathfrak{X} \rightarrow \mathbb{R}$ such that there exist $\epsilon \in (0, \infty)$, $M_\epsilon \in (0, \infty)$, and a continuous linear functional $L: \mathfrak{X} \rightarrow \mathbb{R}$ such that

$$|f(x) - L(x)| \leq \frac{M_\epsilon}{2} \|x\|^2 \text{ for all } x \in \mathfrak{X} \text{ with } \|x\| \leq \epsilon. \quad (2.1)$$

Thus, $f(0) = 0$ and L necessarily coincides with the first Fréchet derivative, $f'(0)$, of the function f at 0. Moreover, given $f(0) = 0$, for the smoothness

condition (2.1) to hold, it is enough that the second derivative $f''(x)$ exist and be bounded (in the operator norm) by M_ϵ over all $x \in \mathfrak{X}$ with $\|x\| \leq \epsilon$.

Throughout the remainder of the paper, the use of the symbols f , L , ϵ , and M_ϵ shall refer to the smoothness condition (2.1).

Remark 2.1. A fact useful in applications is that the smoothness condition (2.1) continues to hold over compositions of functions. Specifically, suppose that \mathfrak{X} , \mathfrak{Y} , and \mathfrak{Z} are separable Banach spaces with respective norms $\|\cdot\|_{\mathfrak{X}}$, $\|\cdot\|_{\mathfrak{Y}}$, and $\|\cdot\|_{\mathfrak{Z}}$, and let $h: \mathfrak{X} \rightarrow \mathfrak{Y}$ and $g: \mathfrak{Y} \rightarrow \mathfrak{Z}$ be functions such that

$$\|h(x) - L_h(x)\|_{\mathfrak{Y}} \leq \frac{M_h}{2} \|x\|_{\mathfrak{X}}^2 \text{ for all } x \in \mathfrak{X} \text{ with } \|x\|_{\mathfrak{X}} \leq \epsilon_h \quad (2.2)$$

and

$$\|g(y) - L_g(y)\|_{\mathfrak{Z}} \leq \frac{M_g}{2} \|y\|_{\mathfrak{Y}}^2 \text{ for all } y \in \mathfrak{Y} \text{ with } \|y\|_{\mathfrak{Y}} \leq \epsilon_g \quad (2.3)$$

for some continuous linear operators $L_h: \mathfrak{X} \rightarrow \mathfrak{Y}$, $L_g: \mathfrak{Y} \rightarrow \mathfrak{Z}$ and positive real numbers $M_h, \epsilon_h, M_g, \epsilon_g$. Then the composition $f := g \circ h: \mathfrak{X} \rightarrow \mathfrak{Z}$ satisfies (2.1) with $\mathfrak{3}$ in place of \mathbb{R} , $L = L_g \circ L_h$, $M_\epsilon = M_h \|L_g\| + M_g m_h^2$, $m_h := \|L_h\| + M_h \epsilon_h / 2$, and $\epsilon = \epsilon_h$, provided that ϵ_h is chosen small enough to ensure $m_h \epsilon_h \leq \epsilon_g$. Such a statement can of course be generalized to the composition of any finite number of functions. We shall prove this assertion in Section 4.

Next let X_1, \dots, X_n be independent random vectors in \mathfrak{X} with $\mathbb{E} X_i = 0$ for $i = 1, \dots, n$, and set

$$S := \sum_{i=1}^n X_i.$$

Assume that

$$\sigma := \|L(S)\|_2 = \sqrt{\sum_i \mathbb{E} L(X_i)^2} \quad (2.4)$$

is finite and nonzero, and let

$$T := \frac{f(S)}{\sigma} \quad \text{and} \quad W := \frac{L(S)}{\sigma} = \sum_{i=1}^n \xi_i, \quad (2.5)$$

where

$$\xi_i := L(X_i)/\sigma, \quad i = 1, \dots, n. \quad (2.6)$$

For any real $p \geq 1$, let

$$s_p := \left(\sum_i \|X_i\|_p^p\right)^{1/p} \quad \text{and} \quad \sigma_p := \left(\sum_i \|\xi_i\|_p^p\right)^{1/p}. \quad (2.7)$$

The assumption (made in Subsection 1.1) that \mathfrak{X} is of type 2 implies

$$\|S\|_2 \leq D s_2 \quad (2.8)$$

for some $D \in (0, \infty)$ (cf. [30, 66]); we shall assume that D is chosen to be minimal with respect to this property, so that $D = 1$ with the equality in (2.8) whenever \mathfrak{X} is a Hilbert space. Take also an arbitrary $c_* \in (0, 1)$ and let δ be any real number such that

$$\sum_i \mathbb{E} |\xi_i| (\delta \wedge |\xi_i|) \geq c_*; \quad (2.9)$$

note that such a number δ always exists (because the limit of the left-hand side of (2.9) as $\delta \uparrow \infty$ is 1). Necessarily, $\delta > 0$.

We now have enough notation to present the following bound on the Kolmogorov distance between T and its linear approximation W :

Theorem 2.2. *Let X_1, \dots, X_n be independent zero-mean random vectors in \mathfrak{X} , let $f: \mathfrak{X} \rightarrow \mathbb{R}$ satisfy (2.1), and assume that $\sigma > 0$; also, take any $c_* \in (0, 1)$. Then for all $z \in \mathbb{R}$*

$$|\mathbb{P}(T > z) - \mathbb{P}(W > z)| \leq \frac{1}{2c_*} \left(4\delta + (\sigma_3 + (8/\pi)^{1/6})\mathbf{u} + \sigma_3\mathbf{v} \right) + \mathbb{P}(\|S\| > \epsilon), \tag{2.10}$$

where

$$\mathbf{u} := \frac{M_\epsilon}{2\sigma} \left(\frac{1 + D^2}{2} \right)^{2/3} (s_3^2 + 2^{2/3}s_2^2), \quad \mathbf{v} := \frac{M_\epsilon}{2\sigma} (s_3^2 + 2Ds_{3/2}s_2), \tag{2.11}$$

and δ is any real number satisfying (2.9).

Remark 2.3. As shown in [59, Theorem 1], (2.9) will hold with

$$\delta = \begin{cases} c_*\sigma_3^3 & \text{if } 0 < c_* \leq \frac{1}{2}, \\ \frac{\sigma_3^3 - (2c_* - 1)^2/\sigma_1}{4(1 - c_*)} & \text{if } \frac{1}{2} \leq c_* < 1, \end{cases} \tag{2.12}$$

provided that $\sigma_3 < \infty$. For $c_* = \frac{1}{2}$, this result follows from [11, Remark 2.1]. If δ is chosen to be as in (2.12), then $\frac{1}{2c_*} 4\delta = 2\sigma_3^3$ is constant over all $c_* \in (0, \frac{1}{2}]$, and hence it may be assumed in Theorem 2.2 that $c_* \in [\frac{1}{2}, 1)$. However, in applications of Theorem 2.5 below, some choices of $c_* \in (0, \frac{1}{2})$ in (2.12) will turn out beneficial.

Remark 2.4. The term $\mathbb{P}(\|S\| > \epsilon)$ in (2.10) can be bounded in a variety of ways. For instance, using Chebyshev’s inequality, one can write

$$\mathbb{P}(\|S\| > \epsilon) \leq \frac{\|S\|_2^2}{\epsilon^2} \leq \frac{D^2s_2^2}{\epsilon^2}. \tag{2.13}$$

Alternatively, one can write

$$\mathbb{P}(\|S\| > \epsilon) \leq \frac{\|S\|_3^3}{\epsilon^3} \leq \frac{1 + D^2}{2} \frac{s_3^3 + 2s_2^3}{\epsilon^3},$$

using the Rosenthal-type inequality [67, (12)].

The hardest to obtain result of this section is the following nonuniform bound:

Theorem 2.5. *Assume that the conditions of Theorem 2.2 are satisfied, and take any positive real numbers $\theta, w, \delta_0, \pi_1, \pi_2, \pi_3$, and ω satisfying the additional restrictions*

$$\delta_0 \leq w, \quad \pi_1 + \pi_2 + \pi_3 = 1, \quad \text{and} \quad \omega \leq \frac{M_\epsilon \epsilon^2}{2\pi_1}. \tag{2.14}$$

Then for all

$$z \in [0, \omega/\sigma] \tag{2.15}$$

one has

$$|\mathbb{P}(T > z) - \mathbb{P}(W > z)| \leq \gamma_z + \tau e^{-(1-\pi_1)z/\theta}, \tag{2.16}$$

where

$$\gamma_z := \sum_{i=1}^n \mathbb{P}(\xi_i > \pi_2 z) + \sum_{i=1}^n \mathbb{P}(W - \xi_i \geq \pi_3 z) \mathbb{P}(\xi_i > w) + \mathbb{P}\left(\|S\| > \sqrt{\frac{2\pi_1 \sigma z}{M_\epsilon}}\right), \tag{2.17}$$

$$\tau := c_1 \sigma_3 \mathbf{v} + c_2 \mathbf{u} + c_3 \delta, \tag{2.18}$$

\mathbf{u} and \mathbf{v} are as in (2.11), and $c_1, c_2,$ and c_3 are finite positive expressions defined in (4.14)–(4.16), which depend only on $c_*, \theta, w, \delta_0, \max_i \|\xi_i\|_2,$ and σ_3 .

Remark 2.6. The bound (2.16) is stated only for $z \geq 0$, which allows for one-tail expressions such as $\mathbb{P}(\xi_i > \pi_2 z)$ to be used in (2.17), as opposed to larger two-tail expressions such as $\mathbb{P}(|\xi_i| > \pi_2 |z|)$. In order to obtain a corresponding bound for $z < 0$, all that is needed is to replace f with $-f$ (and hence also L with $-L$ and ξ_i with $-\xi_i$).

Remark 2.7. The restriction (2.15) is of essence. Indeed, if $z \gg \frac{1}{\sigma}$ (that is, if z is much greater than $\frac{1}{\sigma}$) and the event $\{W > z\}$ in (2.16) occurs, then, by (2.5), $L(S) \gg 1$ and hence $\|S\| \gg 1$, and in this latter zone, of large deviations of S from its zero mean, the linear approximation of $f(S)$ by $L(S)$ will usually break down. This heuristics will be implicitly used in Appendix B, which shows that the upper bound $\frac{w}{\sigma}$ on z in (2.15) is indeed the best possible up to a constant factor, even when the Hilbert space \mathfrak{X} is one-dimensional. Note also that the last inequality in (2.14) can be satisfied for any given $\omega \in (0, \infty)$ by (say) taking π_1 to be small enough.

Remark 2.8. If, in the conditions of Theorem 2.5, it is additionally assumed that the ξ_i 's are all symmetric(ally distributed), then, according to the main result of [58], the expression $e^{3w/\theta} - 1$ (found in the definition (4.15) of c_2) may be replaced by the smaller quantity $\sinh(3w/\theta)$. This sharpening of the inequality (2.16) allows for smaller universal constants to be obtained in applications of Theorem 2.5.

The bounds in (2.10) and (2.16) on the closeness of the distribution of the linear approximation W to that of the original statistic T are to be complemented by any number of well-known BE-type bounds on the closeness of the distribution of the linear statistic W to the standard normal distribution; the reader may be referred to Petrov's monograph [50, Chapter V] or the paper [55]. For instance, for the linear statistic W as in (2.5) with i.i.d. ξ_1, \dots, ξ_n , results due to Shevtsova [74] and Michel [44] imply

$$|\mathbb{P}(W \leq z) - \Phi(z)| \leq n \left(0.33554 (\|\xi_1\|_3^3 + 0.415 \|\xi_1\|_2^3) \wedge \frac{30.2211 \|\xi_1\|_3^3}{|z|^3 + 1} \right). \tag{2.19}$$

While the expressions for the upper bounds given in Theorems 2.2 and 2.5 are quite explicit, they may seem complicated (especially as compared with the

classical uniform and nonuniform BE bounds). This should hardly be surprising, however, when one reflects upon the generality of the results. Indeed, there are a whole host of players here: those associated with the function f and the space \mathfrak{X} (namely, L , ϵ , M_ϵ , and D), the parameters we are free to choose (c_* , θ , w , δ_0 , π_1 , π_2 , π_3 , and ω), and more traditional terms (σ_3 , s_2 , s_3 , and σ) – each with a significant and rather circumscribed role to play.

In applications to problems of the asymptotic relative efficiency of statistical tests, usually it is the closeness of the distribution of the test statistic to a normal distribution (on \mathbb{R}) that is needed or most convenient; in fact, obtaining uniform bounds on such closeness was our original motivation for this work. On the other hand, there have been a number of deep results on the closeness of the distribution of $f(S)$, not to a normal distribution, but to that of $f(N)$, where N is a normal random vector with the mean and covariance matching those of S . In particular, Götze [25] provided an upper bound of the order $O(1/\sqrt{n})$ on the uniform distance between the d.f.'s of the r.v.'s $f(S)$ and $f(N)$ under comparatively mild restrictions on the smoothness of f ; however, the bound increases to ∞ with the dimension k of the space \mathfrak{X} (which is \mathbb{R}^k therein). Chen and Fang [10, Theorem 3.5] recently showed the constant in this bound is $O(k^{1/2})$. One important feature of the bounds in Theorems 2.2 and 2.5 is that they do not explicitly depend on the dimension of the space \mathfrak{X} but only on the choice of the norm $\|\cdot\|$ on \mathfrak{X} .

One should also note here such results as the ones obtained by Götze [24] (uniform bounds) and Zalesskiĭ [78, 79] (nonuniform bounds), also on the closeness of the distribution of $f(S)$ to that of $f(N)$. There (in an i.i.d. case), \mathfrak{X} can be any type 2 Banach space, but f is required to be at least thrice differentiable, with certain conditions on the derivatives. Moreover, Bentkus and Götze [5] provide several examples showing that, in an infinite-dimensional space \mathfrak{X} , the existence of the first three derivatives (and the associated smoothness conditions on such derivatives) cannot be relaxed in general.

Another advantage of the bounds in (2.10) and (2.16) is that they do not explicitly depend on n . Indeed, n is irrelevant when the X_i 's are not identically distributed (because one could e.g. introduce any number of additional zero summands X_i). In fact, (2.10) and (2.16) remain valid when S is the sum of an infinite series of independent zero-mean random vectors, i.e. $S = \sum_{i=1}^{\infty} X_i$, provided that the series converges in an appropriate sense; see e.g. Jain and Marcus [33].

On the other hand, for i.i.d. random vectors X_i our bounds have the correct order of magnitude in n . Indeed, let

$$V, V_1, \dots, V_n \text{ be i.i.d. random vectors}$$

in \mathfrak{X} , with $\mathbb{E}V = 0$. Throughout the rest of this section and in Section 3, we shall use

$$V_i/n \text{ in place of } X_i \text{ and } \bar{V} := \frac{1}{n} \sum_{i=1}^n V_i \text{ in place of } S. \quad (2.20)$$

Further let

$$\tilde{\sigma} := \|L(V)\|_2, \quad v_p := \|V\|_p, \quad \text{and} \quad \varsigma_p := \frac{\|L(V)\|_p}{\tilde{\sigma}} \quad (2.21)$$

for any $p \geq 1$, so that (2.4), (2.6), and (2.7) yield

$$\sigma = \frac{\tilde{\sigma}}{\sqrt{n}}, \quad \xi_i = \frac{L(V_i)}{\tilde{\sigma}\sqrt{n}}, \quad s_p = \frac{v_p}{n^{1-1/p}}, \quad \text{and} \quad \sigma_p = \frac{\varsigma_p}{n^{1/2-1/p}}. \quad (2.22)$$

Theorem 2.9. *Suppose that (2.1) holds, and let V, V_1, V_2, \dots, V_n be i.i.d. zero-mean random vectors with $\tilde{\sigma} > 0$ and $v_3 < \infty$. Then for all $z \in \mathbb{R}$*

$$\left| \mathbb{P}\left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{\sqrt{n}}. \quad (2.23)$$

Moreover, for any $\omega \in (0, \infty)$ and for all

$$z \in (0, \omega\sqrt{n}] \quad (2.24)$$

one has

$$\left| \mathbb{P}\left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \mathfrak{C} \left(n \mathbb{P}(\|V\| > \mathfrak{C}z\sqrt{n}) + \frac{n \mathbb{P}(\|V\| > \mathfrak{C}\sqrt{n})}{z^3} + \frac{1}{z^3 n^{3/2}} + \frac{1}{e^z \sqrt{n}} \right) \quad (2.25)$$

$$\leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}. \quad (2.26)$$

Each instance of \mathfrak{C} above is a finite positive expression that depends only upon the space \mathfrak{X} (through the constant D in (2.8)), the function f (through (2.1)), and the moments $\tilde{\sigma}$, ς_3 , $v_{3/2}$, v_2 , and v_3 , with \mathfrak{C} in (2.25) and (2.26) also depending on ω . Moreover, (2.23) and (2.25) both hold when $\mathbb{P}(\sqrt{n}L(\bar{V})/\tilde{\sigma} \leq z)$ replaces $\Phi(z)$.

The restriction (2.24) concerning (2.25) cannot be relaxed in general; see Appendix B.

Theorem 2.10. *Suppose that a function $f: \mathbb{R} \rightarrow \mathbb{R}$ is twice continuously differentiable in a neighborhood of 0, with $f(0) = 0$ and $f'(0) \neq 0$. Let Y, Y_1, Y_2, \dots be a sequence of i.i.d. zero-mean unit-variance real-valued r.v.'s with $\|Y\|_3 < \infty$, and let $\bar{Y}_n := \frac{1}{n} \sum_{i=1}^n Y_i$. Then there exists a real number $\mathfrak{C} > 0$ such that for all $n \in \mathbb{N}$ and all $z \in \mathbb{R}$*

$$\left| \mathbb{P}\left(\frac{f(\bar{Y}_n)}{|f'(0)|/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{\sqrt{n}}. \quad (2.27)$$

Moreover, for any $\omega \in (0, \infty)$ there exists a real number $\mathfrak{C} > 0$ such that for all $n \in \mathbb{N}$ and all z as in (2.24)

$$\left| \mathbb{P}\left(\frac{f(\bar{Y}_n)}{|f'(0)|/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}. \quad (2.28)$$

Theorem 2.10, a straightforward consequence of Theorem 2.9, is stated here to provide an example of uniform and nonuniform BE bounds for the “classical”, “univariate” delta method; even this very simple case appears to be new to the literature. Just as with the BE bound for linear statistics, we see that the moment restriction $\|Y\|_3 < \infty$ is sufficient to obtain a bound on the order of $O(1/\sqrt{n})$. That bounds such as (2.27) are useful in applications was suggested to us by E. MolavianJazi [45], who needed such a result in his research in electrical engineering.

Theorem 2.11. *Let \mathfrak{X} be a Hilbert space, let f satisfy (2.1) for some real $\epsilon > 0$, and assume that $\mathbb{E}V = 0$, $\tilde{\sigma} > 0$, and $v_3 < \infty$. Take any real number $c_* \in [\frac{1}{2}, 1)$. Then*

$$\left| \mathbb{P}\left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{K}_0 + \mathfrak{K}_1 \varsigma_3^3 + (\mathfrak{K}_{20} + \mathfrak{K}_{21} \varsigma_3) v_2^2 + (\mathfrak{K}_{30} + \mathfrak{K}_{31} \varsigma_3) v_3^2 + \mathfrak{K}_\epsilon}{\sqrt{n}} \quad (2.29)$$

for all $z \in \mathbb{R}$ and $n \in \mathbb{N}$, where

$$\begin{aligned} \mathfrak{K}_0 &:= 0.13925 - \frac{(2c_* - 1)^2}{2c_*(1 - c_*)}, & \mathfrak{K}_1 &:= 0.33554 + \frac{1}{2c_*(1 - c_*)}, \\ (\mathfrak{K}_{20}, \mathfrak{K}_{21}, \mathfrak{K}_{30}, \mathfrak{K}_{31}) &:= \frac{M_\epsilon}{4c_*\tilde{\sigma}} \left(2\left(\frac{2}{\pi}\right)^{1/6}, 2 + \frac{2^{2/3}}{n^{1/6}}, \frac{(8/\pi)^{1/6}}{n^{1/3}}, \frac{2}{n^{1/2}} \right), & (2.30) \\ \text{and } \mathfrak{K}_\epsilon &:= \frac{v_2^2}{\epsilon^2 n^{1/2}} \wedge \frac{2v_2^3 + v_3^3/n^{1/2}}{\epsilon^3 n}. \end{aligned}$$

Thus, Theorem 2.11 provides an explicit expression of the constant \mathfrak{C} in (2.23), depending only on n , the moments $\tilde{\sigma}$, ς_3 , v_2 , and v_3 , and a choice of the parameters c_* , ϵ (and hence M_ϵ). However, in many applications the non-explicit bounds presented in Theorem 2.9 will be enough; see e.g. [20].

An explicit counterpart to the nonuniform bounds in (2.25)–(2.26) is presented in Theorem A.2. It takes nearly three pages of notation just to state that theorem (and many more to prove it). Therefore, we defer the proof of Theorem A.2 to Appendix A. However, the very long expressions of the bounds in Theorem A.2 can be greatly simplified in specific applications, as is illustrated e.g. in the application to the central T statistic in [65]. Moreover, the corresponding explicit constants in that application are rather moderate in size, and indeed are comparable with the constant factor in the nonuniform BE inequality (2.19) for linear statistics. Furthermore, the constant factor in the asymptotic bound for general statistics deduced from Theorem A.2 is just the same as that for sums of independent real-valued r.v.’s; see (D.3) in Appendix D.

3. Applications

Here we shall apply the results of Section 2 to present several novel bounds on the rate of convergence to normality in some common applications of the delta

method. Throughout this section, we shall use the i.i.d. notation as presented in (2.20)–(2.22). Moreover, in most of our applications V will take values in the Euclidean space $\mathfrak{X} = \mathbb{R}^k$ for some natural number k ; an exception to this assumption is taken in Subsection 3.5, where we allow \mathfrak{X} to be an infinite-dimensional Hilbert space.

The components of the zero-mean random vector V will typically be some functions of real-valued zero-mean r.v.'s denoted by Y and Z . E.g., in our application to the Pearson statistic in Subsection 3.2, V will stand for the 5-dimensional zero-mean random vector $(Y, Z, Y^2 - 1, Z^2 - 1, YZ - \mathbb{E}YZ)$, where Y and Z are zero-mean unit-variance r.v.'s.

The non-explicit bounds for the Student, Pearson, noncentral Hotelling, certain covariance statistics, principal component analysis, and maximum likelihood estimators to be given in Theorems 3.1, 3.4, 3.9, 3.12, 3.15 and 3.16, respectively, are more or less straightforward applications of Theorem 2.9.

In contrast, the explicit bounds, such as the one for the Pearson statistic given in Corollary 3.8, are based on Theorem 2.11, and they require significantly more work. As the proof of Corollary 3.8 is rather lengthy and technical, it is placed in Appendix C.

3.1. Student's T

Let Y, Y_1, \dots, Y_n be i.i.d. real-valued r.v.'s, with

$$\mu := \mathbb{E}Y \quad \text{and} \quad \text{Var}Y \in (0, \infty).$$

Consider the statistic commonly referred to as Student's T (or simply T):

$$T := \frac{\bar{Y}}{S_Y/\sqrt{n}} = \frac{\sqrt{n} \bar{Y}}{(\bar{Y}^2 - \bar{Y}^2)^{1/2}},$$

where

$$\begin{aligned} \bar{Y} &:= \frac{1}{n} \sum_i Y_i, & \bar{Y}^2 &:= \frac{1}{n} \sum_i Y_i^2, \\ \text{and } S_Y &:= \left(\frac{1}{n} \sum_i (Y_i - \bar{Y})^2 \right)^{1/2} = \left(\bar{Y}^2 - \bar{Y}^2 \right)^{1/2}; \end{aligned}$$

let $T := 0$ when $\bar{Y}^2 = \bar{Y}^2$. Note that S_Y is defined here as the empirical standard deviation of the sample $(Y_i)_{i=1}^n$, rather than the sample standard deviation $(\frac{n}{n-1}(\bar{Y}^2 - \bar{Y}^2))^{1/2}$. Let us call T “central” when $\mu = 0$ and “noncentral” when $\mu \neq 0$.

As T is invariant under the transformation $Y_i \mapsto aY_i$ for arbitrary $a > 0$, let us assume without loss of generality (w.l.o.g.) that

$$\text{Var}Y = 1.$$

Now let $\mathfrak{X} = \mathbb{R}^2$, and for $\mathbf{x} = (x_1, x_2) \in \mathfrak{X}$ such that $1 + x_2 - x_1^2 > 0$, let $f: \mathfrak{X} \rightarrow \mathbb{R}$ be defined by

$$f(\mathbf{x}) = f(x_1, x_2) = \frac{x_1 + \mu}{\sqrt{1 + x_2 - x_1^2}} - \mu;$$

let $f(\mathbf{x}) := -\mu$ for all other $\mathbf{x} \in \mathfrak{X}$. Since

$$\min_{x_1^2 + x_2^2 \leq \epsilon^2} (1 + x_2 - x_1^2) = \begin{cases} 1 - \epsilon & \text{if } 0 < \epsilon \leq \frac{1}{2}, \\ \frac{3}{4} - \epsilon^2 & \text{if } \epsilon \geq \frac{1}{2}, \end{cases} \quad (3.1)$$

it is easy to see that f'' is continuous (and hence uniformly bounded) on the closed ball $\{\mathbf{x} \in \mathfrak{X}: \|\mathbf{x}\| \leq \epsilon\}$ for any fixed $\epsilon \in (0, \sqrt{3}/2)$. So, the smoothness condition (2.1) is satisfied, with $L(\mathbf{x}) = x_1 - \mu x_2/2$ for $\mathbf{x} = (x_1, x_2) \in \mathfrak{X}$, and upon letting

$$V = (Y - \mu, (Y - \mu)^2 - 1) \quad (3.2)$$

we see that $\sqrt{n}f(\bar{V}) = T - \sqrt{n}\mu$. Then Theorem 2.9 immediately yields

Theorem 3.1. *Assume that $\tilde{\sigma} > 0$ and $v_3 < \infty$, for $\tilde{\sigma}$ and v_p defined in (2.21). Then for all $n \in \mathbb{N}$ and $z \in \mathbb{R}$*

$$\left| \mathbb{P}\left(\frac{T - \sqrt{n}\mu}{\tilde{\sigma}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{\sqrt{n}}, \quad (3.3)$$

where \mathfrak{C} is a finite expression depending only on the distribution of Y ; also, for any $\omega > 0$ and for all real $z > 0$ and $n \in \mathbb{N}$ satisfying (2.24),

$$\left| \mathbb{P}\left(\frac{T - \sqrt{n}\mu}{\tilde{\sigma}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}, \quad (3.4)$$

where \mathfrak{C} is a finite expression depending only on ω and the distribution of Y .

Remark 3.2. If $\mu = 0$ then $\tilde{\sigma} \neq 0$, and otherwise $\tilde{\sigma} = 0$ only if Y has a 2-point distribution, which depends only on μ . Indeed, if $\mu \neq 0$ then $\tilde{\sigma} = 0 \Leftrightarrow L(V) = 0$ almost surely (a.s.) $\Leftrightarrow Y - \mu = (1 \pm \sqrt{1 + \mu^2})/\mu$ a.s. That is, $\tilde{\sigma} = 0$ if and only if $Y = 2\sqrt{p(1-p)}/(1-2p) + B_p$ a.s., where B_p is a standardized Bernoulli(p) r.v. with $p \in (0, 1) \setminus \{\frac{1}{2}\}$.

Remark 3.3. The upper bound in (3.3) is optimal in its dependence on n for the noncentral T . Indeed, suppose that a function $f: \mathbb{R}^k \rightarrow \mathbb{R}$ is twice continuously differentiable in a neighborhood of the origin (so that f satisfies the smoothness condition (2.1)), and let L and H denote here the gradient vector and Hessian matrix of f at 0. Further assume, in addition to the assumptions $\tilde{\sigma} > 0$ and $v_3 < \infty$, that V satisfies the Cramér-type condition $\limsup_{\|t\| \rightarrow \infty} |\mathbb{E} e^{it^\top V}| < 1$. Then a calculation of the asymptotic distribution of $\sqrt{n}f(\bar{V})/\tilde{\sigma}$ using [8, Theorem 2] implies

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}\left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) - \frac{\Delta(z)}{\sqrt{n}} \right| = o\left(\frac{1}{\sqrt{n}}\right), \quad (3.5)$$

where

$$\Delta(z) := -\left[\left(\frac{\mathbb{E}[(L^\top V)^3]}{6\tilde{\sigma}^3} + a_3\right)(z^2 - 1) + a_1\right]\varphi(z), \tag{3.6}$$

$$a_1 := \frac{1}{2\tilde{\sigma}} \operatorname{tr}(H\Sigma), \quad a_3 := \frac{1}{4\tilde{\sigma}^3}(L^\top \Sigma L - \tilde{\sigma}^2) \operatorname{tr}(H\Sigma) + \frac{1}{2\tilde{\sigma}^3} L^\top \Sigma H \Sigma L,$$

Σ denotes the covariance matrix of V , and φ is the standard normal density.

In the conditions of Theorem 3.1, take the simple case where Y is unit-variance, symmetric about its non-zero mean μ , and has an absolutely continuous distribution; let $\nu_k := \mathbb{E}(Y - \mu)^k$ denote the k^{th} central moment of Y , so that $\nu_k = 0$ for odd natural k . Then, for $\Delta(z)$ as in (3.6), $\Delta(1) = -\frac{\mu(1+3\nu_4)}{8\tilde{\sigma}} \varphi(1)$ and $\tilde{\sigma} = 1 + \frac{\mu^2}{4}(\nu_4 - 1)$. That is, $\tilde{\sigma} > 0$ and $\Delta(1) \neq 0$, and we see that the dependence of the upper bound in (3.3) on n is optimal.

The bounds in (3.3) and (3.4) appear to be new for the noncentral T . It was shown in [7] that, if $\|Y\|_4 < \infty$, then (after some standardization) T has a limit distribution which is either the standard normal distribution or the χ^2 distribution with one degree of freedom; the latter will be the case if and only if Y has the two-point distribution described above in Remark 3.2 concerning the degeneracy condition $\tilde{\sigma} = 0$.

The bounds in Theorem 3.1 apply to the central T as well. In Subsection 4.2.1 of the arXiv version of this paper [65] we present several uniform and nonuniform BE-type bounds for the self-normalized sum – which is a simple one-to-one monotonic transformation of the central Student T , along with comparisons of those bounds to existing ones in the literature. The reader is referred to [73] for a survey of limit theorems concerning the self-normalized sums.

3.2. Pearson’s R

Let $(Y, Z), (Y_1, Z_1), \dots, (Y_n, Z_n)$ be a sequence of i.i.d. random points in \mathbb{R}^2 , with

$$\operatorname{Var} Y \in (0, \infty) \quad \text{and} \quad \operatorname{Var} Z \in (0, \infty).$$

Recall the definition of Pearson’s product-moment correlation coefficient:

$$R := \frac{\sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}} = \frac{\overline{YZ} - \bar{Y} \bar{Z}}{\sqrt{\overline{Y^2} - \bar{Y}^2} \sqrt{\overline{Z^2} - \bar{Z}^2}}, \tag{3.7}$$

where

$$\begin{aligned} \bar{Y} &:= \frac{1}{n} \sum_i Y_i, & \bar{Z} &:= \frac{1}{n} \sum_i Z_i, & \overline{Y^2} &:= \frac{1}{n} \sum_i Y_i^2, & \overline{Z^2} &:= \frac{1}{n} \sum_i Z_i^2, \\ & & & & & & \text{and } \overline{YZ} &:= \frac{1}{n} \sum_i Y_i Z_i; \end{aligned}$$

let $R := 0$ if the denominator in (3.7) is 0. Note that R is invariant under all affine transformations of the form $Y_i \mapsto a + bY_i$ and $Z_i \mapsto c + dZ_i$ with positive

b and d ; so, in what follows we may (and shall) assume that the r.v.'s Y and Z are standardized:

$$\mathbb{E}Y = \mathbb{E}Z = 0 \quad \text{and} \quad \mathbb{E}Y^2 = \mathbb{E}Z^2 = 1, \quad \text{and we let } \rho := \mathbb{E}YZ = \text{Corr}(Y, Z).$$

Let $\mathfrak{X} = \mathbb{R}^5$, and for $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5) \in \mathfrak{X}$ such that $1 + x_3 - x_1^2 > 0$ and $1 + x_4 - x_2^2 > 0$, let

$$f(\mathbf{x}) = f(x_1, x_2, x_3, x_4, x_5) = \frac{x_5 + \rho - x_1x_2}{\sqrt{1 + x_3 - x_1^2}\sqrt{1 + x_4 - x_2^2}} - \rho; \quad (3.8)$$

let $f(\mathbf{x}) := -\rho$ for all other $\mathbf{x} \in \mathfrak{X}$. Recall (3.1) to see that $f''(\mathbf{x})$ exists and is continuous on the closed ϵ -ball about the origin for any fixed $\epsilon \in (0, \sqrt{3}/2)$; then the smoothness condition (2.1) holds, with $L(\mathbf{x}) = f'(0)(x_1, x_2, x_3, x_4, x_5) = -\rho x_3/2 - \rho x_4/2 + x_5$. Letting $V = (Y, Z, Y^2 - 1, Z^2 - 1, YZ - \rho)$, so that $L(V) = YZ - \frac{\rho}{2}(Y^2 + Z^2)$, we see that $f(\bar{V}) = R - \rho$. Then Theorem 2.9 immediately yields

Theorem 3.4. *Assume that $\tilde{\sigma} > 0$ and $v_3 < \infty$, for $\tilde{\sigma}$ and v_p defined in (2.21). Then for all $n \in \mathbb{N}$ and $z \in \mathbb{R}$*

$$\left| \mathbb{P}\left(\frac{R - \rho}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{\sqrt{n}}, \quad (3.9)$$

where \mathfrak{C} is a finite expression depending only on the distribution of the random point (Y, Z) . Also, for any $\omega > 0$ and all real $z > 0$ and $n \in \mathbb{N}$ satisfying (2.24),

$$\left| \mathbb{P}\left(\frac{R - \rho}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}, \quad (3.10)$$

where \mathfrak{C} is a finite expression depending only on ω and the distribution of (Y, Z) .

Remark 3.5. Note that the degeneracy condition $\tilde{\sigma} = 0$ is equivalent to the following: there exists some $\kappa \in \mathbb{R}$ such that the random point (Y, Z) lies a.s. on the union of the two straight lines through the origin with slopes κ and $1/\kappa$ (for $\kappa = 0$, these two lines should be understood as the two coordinate axes in the plane \mathbb{R}^2). Indeed, if $\tilde{\sigma} = 0$, then $YZ - \frac{\rho}{2}(Y^2 + Z^2) = 0$ a.s.; solving this equation for the slope Z/Y , one obtains two roots, whose product is 1. Vice versa, if (Y, Z) lies a.s. on the union of the two lines through the origin with slopes κ and $1/\kappa$, then $YZ = \frac{r}{2}(Y^2 + Z^2)$ a.s. for $r := 2\kappa/(\kappa^2 + 1)$ and, moreover, $r = \mathbb{E} \frac{r}{2}(Y^2 + Z^2) = \mathbb{E}YZ = \rho$.

For example, let the random point (Y, Z) equal $(cx, \kappa cx)$, $(-cx, -\kappa cx)$, $(\kappa cy, cy)$, $(-\kappa cy, -cy)$ with probabilities $\frac{p}{2}$, $\frac{p}{2}$, $\frac{q}{2}$, $\frac{q}{2}$, respectively, where $x \neq 0$, $y \neq 0$, $\kappa \in \mathbb{R}$, $c := \sqrt{\frac{x^{-2} + y^{-2}}{\kappa^2 + 1}}$, $p := \frac{y^2}{x^2 + y^2}$, and $q := 1 - p$; then $\tilde{\sigma} = 0$ (and the r.v.'s Y and Z are standardized). In particular, one can take here $x = y = 1$, so that $p = q = \frac{1}{2}$.

Remark 3.6. The condition $v_3 < \infty$ in Theorem 3.4 is equivalent to $\|Y\|_6 + \|Z\|_6 < \infty$, which might seem overly restrictive, since only the finiteness of the third absolute moments is needed to have a BE-type bound of order $O(1/\sqrt{n})$ for linear statistics. However, the moments $\|Y\|_6$ and $\|Z\|_6$ do appear in an asymptotic expansion (up to an order $n^{-1/2}$) of the distribution of R when $\rho \neq 0$; for details, one can see [53]. When $\rho = 0$, the most restrictive moment assumption for the existence of the asymptotic expansion is that $\|YZ\|_3 < \infty$.

Remark 3.7. Recall the asymptotic distribution results of [8] as outlined in Remark 3.3. In the conditions of Theorem 3.4, take the very simple case when Y and Z are zero-mean, unit-variance, absolutely continuous r.v.'s independent of each other. Then a straightforward calculation shows that $a_1 = 0$, $a_3 = 0$, and hence $\Delta(z) = -\frac{1}{6} \mathbb{E} Y^3 \mathbb{E} Z^3 (z^2 - 1) \varphi(z)$. So, the bound in (3.9) has an optimal dependence on n whenever $\mathbb{E} Y^3 \neq 0$ and $\mathbb{E} Z^3 \neq 0$. Moreover, since $\Delta(z)$ is real-analytic in z , L , H , and moments of V , we see that generally $\Delta(z) \neq 0$ and hence the bound in (3.9) is generally of the optimal order in n .

The bounds in (3.9) and (3.10) appear to be new. In fact, we have not been able to find in the literature any uniform (or nonuniform) bound on the closeness of the distribution of R to normality. Note that such bounds are important in considerations of the asymptotic relative efficiency of statistical tests; see e.g. Noether [48].

A simple and explicit bound for the Pearson statistic is given in the following corollary.

Corollary 3.8 (to Theorem 2.11). *Assume that $\mathbb{E} YZ = 0$ and $\tilde{\sigma} = \|YZ\|_2 > 0$. Then for all $n \in \mathbb{N}$ and $z \in \mathbb{R}$*

$$\left| \mathbb{P} \left(\frac{R}{\tilde{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \leq \frac{B_0 + B_3/\tilde{\sigma}^3}{\sqrt{n}} (\|Y\|_6^6 + \|Z\|_6^6), \tag{3.11}$$

where (B_0, B_3) is any ordered pair in the set

$$\{(3.61, 3.61), (1.12, 8.94), (13.33, 1.69), (0.56, 14.97), (36.32, 1.37)\}. \tag{3.12}$$

The proof of Corollary 3.8 (given in Appendix C) provides a method by which one may obtain a variety of values for the pair (B_0, B_3) . The specific pairs listed in (3.12) are obtained by trying to minimize $B_0 \vee B_3 / \tilde{\sigma}^3$ for $\tilde{\sigma} \in \{1, 2, 1/2, 3, 1/3\}$.

3.3. Noncentral Hotelling's T^2 statistic

Let $k \geq 2$ be an integer, and let Y, Y_1, \dots, Y_n be i.i.d. random vectors in \mathbb{R}^k , with finite

$$\mu := \mathbb{E} Y \quad \text{and} \quad \text{Cov } Y = \mathbb{E} Y Y^\top - \mu \mu^\top \text{ strictly positive definite.}$$

Consider Hotelling's T^2 statistic

$$T^2 := \bar{Y}^\top (S_Y^2/n)^{-1} \bar{Y} = n \bar{Y}^\top (\overline{Y Y^\top} - \bar{Y} \bar{Y}^\top)^{-1} \bar{Y}, \tag{3.13}$$

where

$$\begin{aligned} \bar{Y} &:= \frac{1}{n} \sum_i Y_i, & \overline{Y Y^\top} &:= \frac{1}{n} \sum_i Y_i Y_i^\top, \\ & & \text{and } S_Y^2 &:= \frac{1}{n} \sum_i (Y_i - \bar{Y})(Y_i - \bar{Y})^\top = \overline{Y Y^\top} - \bar{Y} \bar{Y}^\top; \end{aligned}$$

the generalized inverse is often used in place of the inverse in (3.13), though here we may just let $T^2 := 0$ whenever S_Y^2 is singular. Also note that S_Y^2 is defined as the empirical covariance matrix of the sample $(Y_i)_{i=1}^n$, rather than the sample covariance matrix $\frac{n}{n-1} S_Y^2$. Call T^2 “central” when $\mu = 0$ and “noncentral” otherwise.

For any nonsingular matrix B , T^2 is invariant under the invertible transformation $Y_i \mapsto B Y_i$, so let us assume w.l.o.g. that

$$\text{Cov } Y = I,$$

the $k \times k$ identity matrix.

Now let $\mathfrak{X} = \{(x_1, x_2) : x_1 \in \mathbb{R}^k, x_2 \in \mathbb{R}^{k \times k}\}$ be equipped with the norm

$$\|(x_1, x_2)\| := \sqrt{\|x_1\|^2 + \|x_2\|_F^2}, \quad (3.14)$$

where $\|x_2\|_F := \sqrt{\text{tr}(x_2 x_2^\top)}$ is the Frobenius norm. For $\mathbf{x} = (x_1, x_2) \in \mathfrak{X}$ such that $I + x_2 - x_1 x_1^\top$ is nonsingular, let

$$f(\mathbf{x}) = (x_1 + \mu)^\top (I + x_2 - x_1 x_1^\top)^{-1} (x_1 + \mu) - \mu^\top \mu,$$

and let $f(\mathbf{x}) := -\mu^\top \mu$ for all other $\mathbf{x} \in \mathfrak{X}$. The Fréchet derivative of f at the origin is the linear functional defined by $L(\mathbf{x}) = f'(0)(x_1, x_2) = 2x_1^\top \mu - \mu^\top x_2 \mu$. Let us recall a couple of other useful facts (found in, say, the monograph [32]): the spectral norm $\|B\|$ of any $k \times k$ matrix B does not exceed $\|B\|_F$, and $\|B\| < 1$ implies $I - B$ is nonsingular and $\|(I - B)^{-1}\| \leq 1/(1 - \|B\|)$. In particular,

$$\|x_1 x_1^\top - x_2\| \leq \|x_1 x_1^\top - x_2\|_F \leq \|x_1 x_1^\top\|_F + \|x_2\|_F = \|x_1\|^2 + \|x_2\|_F < 1$$

for any \mathbf{x} in the closed ϵ -ball about the origin and any fixed $\epsilon \in (0, \sqrt{3}/2)$ (which again follows from (3.1)), so that the smoothness condition (2.1) holds. Upon letting

$$V = (Y - \mu, (Y - \mu)(Y - \mu)^\top - I),$$

we see that $n f(\bar{V}) = T^2 - n \mu^\top \mu$. Then Theorem 2.9 immediately yields

Theorem 3.9. *Assume that $\tilde{\sigma} > 0$ and $v_3 < \infty$, for $\tilde{\sigma}$ and v_p defined in (2.21). Then for all $n \in \mathbb{N}$ and $z \in \mathbb{R}$*

$$\left| \mathbb{P}\left(\frac{T^2 - n \mu^\top \mu}{\tilde{\sigma} \sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{\sqrt{n}}, \quad (3.15)$$

where \mathfrak{C} is a finite expression depending only on the distribution of Y ; also, for any $\omega > 0$ and all real $z > 0$ and $n \in \mathbb{N}$ satisfying (2.24),

$$\left| \mathbb{P}\left(\frac{T^2 - n \mu^\top \mu}{\tilde{\sigma} \sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}, \quad (3.16)$$

where \mathfrak{C} is a finite expression depending only on ω and the distribution of Y .

Remark 3.10. The non-degeneracy condition $\tilde{\sigma} > 0$ immediately implies that $\mu \neq 0$, so that Theorem 3.9 is applicable only to the noncentral T^2 . If $\mu \neq 0$, then $\tilde{\sigma} = 0$ if and only if $(Y - \mu)^\top \mu = 1 \pm \sqrt{1 + \|\mu\|^2}$ a.s., that is, if and only if $\mathbb{P}(Y^\top \mu = x_1) = 1 - \mathbb{P}(Y^\top \mu = x_2) = p$, where

$$x_1 = 1 + \|\mu\|^2 + \sqrt{1 + \|\mu\|^2}, \quad x_2 = 1 + \|\mu\|^2 - \sqrt{1 + \|\mu\|^2},$$

$$p = \frac{1}{2} \left(1 - \frac{1}{\sqrt{1 + \|\mu\|^2}} \right);$$

in other words, $\tilde{\sigma} = 0$ if and only if Y lies a.s. in the two hyperplanes defined by $Y^\top \mu = x_1$ or $Y^\top \mu = x_2$. Note the similarity to the degeneracy condition of Student's T statistic described in Remark 3.2. Recalling the conditions $\mathbb{E} Y = \mu$ and $\text{Cov} Y = I$, we have $\tilde{\sigma} = 0$ if and only if

$$Y = \xi \frac{\mu}{\|\mu\|} + \tilde{Y} \quad \text{a.s.},$$

where

$$\xi = \frac{2\sqrt{p(1-p)}}{1-2p} + B_p \quad \text{for some } p \in (0, \frac{1}{2}),$$

and \tilde{Y} is a random vector in \mathbb{R}^k such that $\mathbb{E} \tilde{Y} = 0$, $\mathbb{E} \xi \tilde{Y} = 0$, $\tilde{Y}^\top \mu = 0$ a.s., and $\text{Cov} \tilde{Y}$ is the orthoprojector onto the hyperplane $\{\mu\}^\perp := \{\mathbf{x} \in \mathbb{R}^k : \mathbf{x}^\top \mu = 0\}$.

Remark 3.11. Using again [8, Theorem 2] (cf. Remark 3.3), we can show that generally the upper bound in (3.15) has an optimal dependence on n as well. For instance, consider the simple case when $Y = (Y_1, Y_2)$, where Y_1 and Y_2 have absolutely continuous distributions and are independent of one another; further suppose that $\mathbb{E} Y_2 = 0$, $\mathbb{E} Y_1 = \mu_1 \neq 0$, and that Y_1 is symmetric, so that $\mathbb{E}(Y_1 - \mu_1)^m = 0$ for odd natural m . Then, for $\Delta(z)$ as in (3.6),

$$\Delta(1) = -\frac{\mu_1^2(\nu_4 + 1) + 2}{\tilde{\sigma}} \varphi(1) \quad \text{and} \quad \tilde{\sigma} = |\mu_1| \sqrt{(\nu_4 - 1)\mu_1^2 + 4},$$

with $\nu_4 := \mathbb{E}(Y_1 - \mu_1)^4 \geq \mathbb{E}^2(Y_1 - \mu_1)^2 = 1$. So, $\tilde{\sigma} > 0$ and $\Delta(1) \neq 0$. Thus, the dependence of the upper bound in (3.15) on n is optimal.

Again, the bounds in (3.15) and (3.16) appear to be new; we have found no mention of BE bounds for T^2 in the literature. A potential for application of our result for the noncentral Hotelling statistic to certain problems in electrical engineering was indicated in [35].

3.4. Covariance test statistics

For any natural $k \geq 2$, let Y, Y_1, \dots, Y_n be i.i.d. random vectors in \mathbb{R}^k with

$$\mathbb{E} Y = 0 \quad \text{and} \quad \Sigma := \text{Cov} Y = \mathbb{E} Y Y^\top > 0.$$

Further let

$$\alpha := \text{tr}(\Sigma)/k, \quad \beta := \det(\Sigma)^{1/k}, \quad \delta := \sqrt{\text{tr}[(\Sigma - \alpha I)^2]/k} \quad (3.17)$$

be the arithmetic mean, geometric mean, and standard deviation, respectively, of the eigenvalues of Σ ; the assumption that $\Sigma > 0$ implies $\alpha > 0$ and $\beta > 0$.

We consider here a few statistics used to test either the null hypothesis of sphericity ($H_{0,1}: \Sigma = \sigma^2 I$ for some unknown $\sigma^2 > 0$) or the null hypothesis of the identity covariance ($H_{0,2}: \Sigma = I$). Each of these statistics is a smooth function of the sample covariance matrix

$$S := \overline{Y Y^T} - \bar{Y} \bar{Y}^T, \quad \text{where } \bar{Y} := \frac{1}{n} \sum_i Y_i, \quad \text{and } \overline{Y Y^T} := \frac{1}{n} \sum_i Y_i Y_i^T.$$

In turn, S is a smooth function of the zero-mean random vector

$$\bar{V} := \frac{1}{n} \sum_i V_i = (\bar{Y}, \overline{Y Y^T} - \Sigma), \quad \text{where } V_i := (Y_i, Y_i Y_i^T - \Sigma).$$

Let $\mathfrak{X} = \mathbb{R}^k \times \mathbb{R}^{k \times k}$, $\mathfrak{Y} = \mathbb{R}^{k \times k}$, and $\mathfrak{Z} = \mathbb{R}$, where \mathfrak{X} has the norm defined by (3.14) and \mathfrak{Y} is equipped with the spectral norm. Then the function $h: \mathfrak{X} \rightarrow \mathfrak{Y}$ defined by the formula $h(x_1, x_2) = x_2 - x_1 x_1^T$ satisfies the smoothness condition (2.2) with $L_h(x_1, x_2) = x_2$, $M_h = 2$, and any $\epsilon_h \in (0, \infty)$. Moreover, $h(\bar{V}) = S - \Sigma$.

The likelihood-ratio tests of $H_{0,1}$ and $H_{0,2}$ against their negations, based on a normal population, reject for small values of the statistics

$$\Lambda_1 = \frac{\det(S)}{(\text{tr}(S)/k)^k} \quad \text{and} \quad \Lambda_2 = \frac{\det(S)}{e^{\text{tr}(S)}},$$

respectively; see e.g. Muirhead [46, Theorems 8.3.2 and 8.4.2]; one can also find in [46] asymptotic properties of these tests, including expansions of their distributions under both null and nonnull distributions. Associate with Λ_1 the functions $g: \mathfrak{Y} \rightarrow \mathfrak{Z}$ and $L_g: \mathfrak{Y} \rightarrow \mathfrak{Z}$ defined by

$$g(x) = \frac{\det(x + \Sigma)}{[\text{tr}(x + \Sigma)/k]^k} \mathbb{I}\{\text{tr}(x) > -k\alpha\} - \left(\frac{\beta}{\alpha}\right)^k$$

$$\text{and } L_g(x) = \left(\frac{\beta}{\alpha}\right)^k \text{tr}[(\Sigma^{-1} - \alpha^{-1}I)x].$$

Similarly, with the statistic Λ_2 associate the functions g and L_g defined by

$$g(x) = \frac{\det(x + \Sigma)}{e^{\text{tr}(x + \Sigma)}} - \left(\frac{\beta}{e^\alpha}\right)^k \quad \text{and} \quad L_g(x) = \left(\frac{\beta}{e^\alpha}\right)^k \text{tr}[(\Sigma^{-1} - I)x].$$

It is clear that, for either of the two functions g defined above, $L_g = g'(0)$ and g satisfies (2.3) for small enough ϵ_g . Hence $f := g \circ h$ satisfies (2.1), for both versions of the function g , and so, Theorem 2.9 may be applied to $f(\bar{V}) = \Lambda_1 - (\beta/\alpha)^k$ and $f(\bar{V}) = \Lambda_2 - (\beta/e^\alpha)^k$.

For the case when the dimension k is large, Nagao [47] proposes the test statistics

$$U := \frac{1}{k} \text{tr} \left[\left(\frac{S}{\text{tr}(S)/k} - I \right)^2 \right] \quad \text{and} \quad \tilde{V} := \frac{1}{k} \text{tr}[(S - I)^2]$$

in place of the statistics Λ_1 and Λ_2 , respectively. John [36] shows that the test of $H_{0,1}$ based on U is locally most powerful (assuming a normal population). Associate with U the functions

$$g(x) = \frac{1}{k} \operatorname{tr} \left[\left(\frac{x + \Sigma}{\operatorname{tr}(x + \Sigma)/k} - I \right)^2 \right] - \frac{\delta^2}{\alpha^2}$$

and $L_g(x) = \frac{2}{k^2 \alpha^3} \operatorname{tr} [(\Sigma - \alpha I)(k\alpha I - \Sigma)x]$

and with \tilde{V} the functions

$$g(x) = \frac{1}{k} \operatorname{tr} [(x + \Sigma - I)^2] - \delta^2 - (1 - \alpha)^2 \quad \text{and} \quad L_g(x) = \frac{2}{k} \operatorname{tr} [(\Sigma - I)x].$$

It is straightforward to verify that either of the above functions g satisfies the smoothness condition (2.3), and hence Theorem 2.9 may be applied to either of the functions $f(\bar{V}) = U - \delta^2/\alpha^2$ or $f(\bar{V}) = \tilde{V} - \delta^2 - (1 - \alpha)^2$.

Yet one more variation on these tests we consider is the “large-dimensional” case. Ledoit and Wolf [42] investigate the asymptotic behavior of both U and \tilde{V} when $k/n \rightarrow c \in (0, \infty)$ as $n \rightarrow \infty$, as opposed to the “fixed-dimensional” case (where $n \rightarrow \infty$ while k is assumed a constant). They show that the test of $H_{0,1}$ based on U remains consistent in the large-dimensional setting, whereas the test of $H_{0,2}$ based on \tilde{V} is not necessarily consistent. By not dropping terms like k/n in investigations of the asymptotics of \tilde{V} , the authors propose the statistic

$$W := \tilde{V} - \frac{k}{n} \left(\left[\frac{\operatorname{tr}(S)}{k} \right]^2 - 1 \right) = \frac{1}{k} \operatorname{tr} [(S - I)^2] - \frac{k}{n} \left(\left[\frac{\operatorname{tr}(S)}{k} \right]^2 - 1 \right) \quad (3.18)$$

as an alternative to \tilde{V} in the test of $H_{0,2}$. It is shown that W has the same limiting distribution as \tilde{V} in the fixed-dimensional setting while also being consistent in a large-dimensional framework. We see that $f(\bar{V}) = W - \delta^2 - (1 - \alpha)^2 + \frac{k}{n}(\alpha^2 - 1)$ when $f = g \circ h$ and g is defined by

$$g(x) = \frac{1}{k} \operatorname{tr} [(x + \Sigma - I)^2] - \frac{k}{n} \left(\left(\frac{\operatorname{tr}(x + \Sigma)}{k} \right)^2 - 1 \right) - \delta^2 - (1 - \alpha)^2 + \frac{k}{n}(\alpha^2 - 1);$$

moreover, g satisfies (2.3) with $L_g(x) = \frac{2}{k} \operatorname{tr}[(\Sigma - I - \frac{k}{n} \alpha I)x]$.

Theorem 3.12. *Take any $t \in \{\Lambda_1, \Lambda_2, U, \tilde{V}, W\}$, and let $f = g \circ h$ and $L = L_g \circ L_h$ for the functions g and L_g paired with the statistic t as described above. Assume that $\tilde{\sigma} > 0$ and $v_3 < \infty$, for $\tilde{\sigma}$ and v_p defined in (2.21). Then for all $n \in \mathbb{N}$ and $z \in \mathbb{R}$,*

$$\left| \mathbb{P} \left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{\sqrt{n}}, \quad (3.19)$$

where \mathfrak{C} is a finite expression depending only on the distribution of Y ; also, for any $\omega > 0$ and all real $z > 0$ and $n \in \mathbb{N}$ satisfying (2.24),

$$\left| \mathbb{P} \left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}, \quad (3.20)$$

where \mathfrak{C} is a finite expression depending only on ω and the distribution of Y .

Remark 3.13. The non-degeneracy condition $\tilde{\sigma} > 0$ immediately implies that Theorem 3.12 – and the delta method itself – are applicable only to non-null distributions of the statistics Λ_1 , Λ_2 , U , and \tilde{V} , since $L_g = 0$ for any of these statistics under the assumption of their respective null hypotheses. This should hardly be surprising, as it is known that these statistics (or some normalizing function of them) all have a limiting χ^2 distribution under the null hypothesis. However, one can fix the null-degeneracy of the statistics Λ_1 , Λ_2 , U , or \tilde{V} and thus make the delta method and our BE bounds applicable even to the null distributions by using essentially the same trick as in the definition of the statistic W in (3.18), that is, by adding a term of the form $\alpha \left(\left[\frac{\text{tr}(S)}{k} \right]^2 - 1 \right)$ for some nonzero real α .

By diagonalization of Σ , we can simply characterize the degeneracy condition $\tilde{\sigma} = 0$ for any of the above statistics in this subsection. Indeed, by the spectral decomposition, $\Sigma = Q^T D Q$, where D is the diagonal matrix with the eigenvalues $\lambda_1, \dots, \lambda_k$ of Σ on its diagonal and Q is an orthogonal matrix whose columns are corresponding orthonormal eigenvectors of Σ . Let $Z = (Z_1, \dots, Z_k)^T := QY$. Then, for the statistic Λ_1 ,

$$\begin{aligned} \left(\frac{\alpha}{\beta}\right)^k L(V) &= \text{tr}[(\Sigma^{-1} - \alpha^{-1}I)(Y Y^T - \Sigma)] \\ &= \text{tr}[Q^T(D^{-1} - \alpha^{-1}I)QQ^T(Z Z^T - D)Q] \\ &= \text{tr}[(D^{-1} - \alpha^{-1}I)(Z Z^T - D)] \\ &= \text{tr}[(D^{-1} - \alpha^{-1}I)Z Z^T] - \text{tr}[I - \alpha^{-1}D] \\ &= \sum_{j=1}^k \left(\frac{1}{\lambda_j} - \frac{1}{\alpha}\right) Z_j^2. \end{aligned}$$

Since $\tilde{\sigma} = 0$ means precisely that $L(V) = 0$ a.s., it follows that for any non-null alternative, $\tilde{\sigma} = 0$ for the statistic Λ_1 if and only if the support of the distribution of the random vector Y degenerates so as to lie entirely on a certain quadratic conical surface in \mathbb{R}^k . Similar work shows that for one of the statistics Λ_2 , U , \tilde{V} , and W we have $\tilde{\sigma} = 0$ if and only if the respective one of the random (homogeneous or not) quadratic forms

$$\begin{aligned} \sum_{j=1}^k \left(\frac{1}{\lambda_j} - 1\right) Z_j^2, \quad \sum_{j=1}^k (\lambda_j - \alpha)(k\alpha - \lambda_j)(Z_j^2 - \lambda_j), \\ \sum_{j=1}^k (\lambda_j - 1)(Z_j^2 - \lambda_j), \quad \sum_{j=1}^k \left(\lambda_j - 1 - \frac{k\alpha}{n}\right)(Z_j^2 - \lambda_j) \end{aligned}$$

equals 0 a.s. In particular, whenever the random vector Y is absolutely continuous, one has $\tilde{\sigma} > 0$ for all these statistics in the non-null case, and then $\tilde{\sigma} > 0$ for the statistic W even in the null case provided that $(1 - \frac{k}{n})\Sigma \neq I$.

Remark 3.14. Let Σ_0 be any given positive definite symmetric matrix. Then the hypotheses $\Sigma = \sigma^2 \Sigma_0$ (with an unknown $\sigma^2 > 0$) and $\Sigma = \Sigma_0$ on the common covariance matrix Σ of i.i.d. random vectors Y_i are obviously equivalent to the

respective hypotheses $\tilde{\Sigma} = \sigma^2 I$ (with an unknown $\sigma^2 > 0$) and $\tilde{\Sigma} = I$ on the common covariance matrix $\tilde{\Sigma}$ of the i.i.d. random vectors $\tilde{Y}_i := \Sigma_0^{-1/2} Y_i$. So, the results in this subsection can be obviously extended to the more general case of the null hypotheses $\Sigma = \sigma^2 \Sigma_0$ and $\Sigma = \Sigma_0$.

It appears certain that the bounds in Theorem 3.12 are all new to the literature; indeed, any of the results concerning these statistics that we have found investigates their asymptotic properties under the assumption of a normal population, whereas our bounds have only mild moment restrictions on Y . We mention here that Theorem 2.9 could be applied to several other popular statistics which are smooth functions of the sample covariance matrix S . For instance, our results can easily yield BE bounds for statistics proposed by Srivastava [76] or Fisher et al. [18]; Chen et al. [12] propose a statistic for the sphericity test which is a function of a U -statistic, for which the methods of this paper and [11] could presumably be adapted. The reader is referred to [46] for other statistics used in testing for the equality of population covariances or independence between certain projections applied to Y .

3.5. Principal component analysis (PCA)

It is well known that any simple eigenvalue of a (say, symmetric real matrix) and the orthoprojector onto the corresponding eigenspace are smooth functions of the matrix. Therefore, the delta method is almost universally applicable to PCA, and hence so are our results such as Theorem 2.9. The actual verification of the smoothness condition (2.1) in PCA may involve operator perturbation theory and related tools, based on a representation of analytic functions of a linear operator as certain integrals of the resolvent. This representation largely reduces the problem of the smoothness of a general analytic function of an operator to the obvious smoothness of the map $A \mapsto A^{-1}$ on the set of all bounded invertible linear operators A (cf. (F.5) and (F.6)). Whereas this idea is rather transparent, its execution may in some cases be rather nontrivial, and it may result in complicated expressions for ϵ and M_ϵ in (2.1).

As an illustration of these general theses, let us consider here a statistic rather recently introduced by Cupidon et al. [14, 13]. Let Y, Y_1, \dots, Y_n be i.i.d. random vectors taking values in a separable real Hilbert space H with inner product $\langle \cdot, \cdot \rangle$ and the corresponding norm $\| \cdot \|$. Assume at this point that $\mathbb{E} \|Y\|^2 < \infty$, $\mathbb{E} Y = 0$, and the covariance operator

$$R := \mathbb{Cov} Y = \mathbb{E}(Y \otimes Y)$$

of Y is (strictly) positive definite. Here, as usual, \otimes denotes the tensor product on H , so that $Rx = \mathbb{E} \langle x, Y \rangle Y$ for all $x \in H$. Given the condition $\mathbb{E} \|Y\|^2 < \infty$, the covariance operator R is known to be compact, which allows its spectral decomposition – see e.g. [37, Theorem 2.10, page 260]; a short proof of the compactness of R is presented in Appendix E for the readers' convenience.

Next suppose that $H = H_1 \oplus H_2$, where H_1, H_2 are closed orthogonal subspaces of H ; for $j, k \in \{1, 2\}$, let Π_j denote the orthoprojector onto H_j , $R_{jk} :=$

$\Pi_j R \Pi_k$, and also let I_j denote the identity operator on H_j . Then, for any fixed $\alpha > 0$, the *regularized squared principal canonical correlation*, RSPCC or ρ^2 , is defined by the formula

$$\rho^2 := \rho^2(\alpha) := \max_{\substack{x \in H_1 \setminus \{0\} \\ y \in H_2 \setminus \{0\}}} \frac{\langle x, R_{12}y \rangle^2}{\langle x, (\alpha I_1 + R_{11})x \rangle \langle y, (\alpha I_2 + R_{22})y \rangle}; \quad (3.21)$$

that this is a well-defined quantity is proved in [14]. Define the sample RSPCC, $\hat{\rho}^2$, by replacing R_{jk} in (3.21) with S_{jk} , where

$$S_{jk} = \Pi_j S \Pi_k, \quad S := \overline{Y \otimes Y} - \overline{Y} \otimes \overline{Y}, \quad \overline{Y} := \frac{1}{n} \sum_i Y_i, \\ \text{and} \quad \overline{Y \otimes Y} := \frac{1}{n} \sum_i Y_i \otimes Y_i;$$

thus, S is the sample covariance operator of the random vector Y . See e.g. [27, 17] for discussion and results on the use of canonical correlations in functional data.

Next define the (bounded self-adjoint nonnegative-definite linear) operators

$$R_1 := (\alpha I_1 + R_{11})^{-1/2} R_{12} (\alpha I_2 + R_{22})^{-1} R_{21} (\alpha I_1 + R_{11})^{-1/2}, \\ R_2 := (\alpha I_2 + R_{22})^{-1/2} R_{21} (\alpha I_1 + R_{11})^{-1} R_{12} (\alpha I_2 + R_{22})^{-1/2}, \quad (3.22)$$

and similarly let \hat{R}_j denote the sample analogues of R_j (obtained by replacing R_{jk} with S_{jk}); under the assumption that $\mathbb{E} \|Y\|^2 < \infty$ (which implies that R is compact), we see that R_1 and R_2 are also compact. Moreover, by [13, Theorem 2.4], $\|R_1\| = \|R_2\| = \rho^2$ and $\|\hat{R}_1\| = \|\hat{R}_2\| = \hat{\rho}^2$, where ρ^2 is as in (3.21) and $\|\cdot\|$ denotes the operator norm, so that $\|\hat{R}_j\|$ is the largest eigenvalue of \hat{R}_j .

Fix any $j \in \{1, 2\}$ and assume that ρ^2 is a simple nonzero eigenvalue of R_j , and then let P denote the orthoprojector onto the corresponding (one-dimensional) eigenspace of R_j . Let $B(H)$ and $B(H_j)$ denote the Hilbert spaces of all bounded linear operators on H and H_j , respectively, equipped with the corresponding operator norms.

Let $g(x) := \|x + R_j\| - \|R_j\|$ for any $x \in B(H_j)$, so that $g(\hat{R}_j - R_j) = \hat{\rho}^2 - \rho^2$. By formulas (3.6)–(3.8) on page 89, (2.32) on page 79, and (3.4) on page 88 in [37] (with $n = 1$, $\varkappa = 1$, $\hat{\lambda}(\varkappa) = \|x + \hat{R}_j\|$, $\lambda = \|R_j\| = \rho^2$, $\hat{\lambda}^{(1)} = \text{tr}(xP)$, $T^{(1)} = x$, $T^{(2)} = T^{(3)} = \dots = 0$, $a = \|x\|$, $c = 0$, and $0^0 := 1$), the smoothness condition (2.3) will be satisfied with $\epsilon_g = \beta/\mathfrak{m}$, $\beta \in (0, 1)$, $L(x) = \hat{\lambda}^{(1)} = \text{tr}(xP)$, and $M_g = 2\frac{\varrho \mathfrak{m}^2}{1-\beta}$, where $\varrho := \max_{z \in \Gamma} |z - \lambda|$, $\mathfrak{m} := \max_{z \in \Gamma} \|\mathcal{R}_j(z)\|$, $\mathcal{R}_j(z) := (R_j - zI)^{-1}$ is the resolvent of R_j , and Γ is the boundary of any open disc D in \mathbb{C} such that $\lambda \in D$ but the closure of D does not contain 0 or any eigenvalue of R_j other than λ .

(The results from [37] referred to in the above paragraph were stated there for the case when the Hilbert space H is finite-dimensional. All those results carry verbatim to the “infinite-dimensional” case. Such information can be extracted from other chapters in [37]. However, for readers’ convenience, in Appendix F we provide the few necessary stepping stones to make the transition to the infinite dimension.)

By [21, Theorem 2.1], condition (2.2) holds for the function $y \mapsto h_j(y) := (\alpha I_j + R_{jj} + y)^{-1/2} - (\alpha I_j + R_{jj})^{-1/2}$ in place of h for some real $\epsilon_{h_j} > 0$ and all $y \in H_j$ with $\|y\| \leq \epsilon_{h_j}$. So, in view of definitions (3.22) of R_j , their counterparts for \hat{R}_j , and Remark 2.1, one can set up a function $h: H \times B(H) \rightarrow H_j$ in a straightforward manner so that condition (2.2) holds and $h(\bar{V}) = \hat{R}_j - R_j$, with the zero-mean vector $V = (Y, Y \otimes Y - R)$. Using Remark 2.1 once again, one sees that the function $f = g \circ h$ satisfies the smoothness condition (2.1), and at that $f(\bar{V}) = \hat{\rho}^2 - \rho^2$. Thus, Theorem 2.9 yields

Theorem 3.15. *Assume that $\tilde{\sigma} > 0$ and $v_3 < \infty$, for $\tilde{\sigma}$ and v_p defined in (2.21). Then for all $n \in \mathbb{N}$ and $z \in \mathbb{R}$,*

$$\left| \mathbb{P}\left(\frac{\hat{\rho}^2 - \rho^2}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{\sqrt{n}}, \tag{3.23}$$

where \mathfrak{C} is a finite expression depending only on the distribution of Y ; also, for any $\omega > 0$ and all real $z > 0$ and $n \in \mathbb{N}$ satisfying (2.24),

$$\left| \mathbb{P}\left(\frac{\hat{\rho}^2 - \rho^2}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}, \tag{3.24}$$

where \mathfrak{C} is a finite expression depending only on ω and the distribution of Y .

Expressions for $\tilde{\sigma}$ can be obtained from [14, (4.20), (5.1)]. We see the recurring theme that $\|Y\|_4 < \infty$ is used to establish asymptotic normality of $\hat{\rho}^2$ (cf. [14, (2.1), Theorem 4.2]), while the moment restriction $\|Y\|_6 < \infty$ (equivalent to $v_3 < \infty$ in Theorem 3.15) is needed here to bound the rate of convergence on the order $O(1/\sqrt{n})$. Again, it appears that the bounds in Theorem 3.15 are entirely new to the literature.

In Subsection 3.4, we considered various smooth functions of the determinant and trace of the sample covariance matrix for finite-dimensional random vectors Y , and in the present subsection we have a function of the largest eigenvalue of some smooth function of a sample covariance operator. Other statistics which are functions of eigenvalues from a sample covariance operator (be it constructed from a finite-dimensional or infinite-dimensional population) may of course lie in the class of statistics to which Theorem 2.9 could be applied; the primary problem to the practitioner is the demonstration of the smoothness condition (2.1). The use of perturbation theory, as was done above, appears to be valuable for many such potential applications; we mention here statistics proposed in [34, 19], concerning the testing of equality of two covariance operators, as further examples. Yet another potential application of our results would be to the empirical Wasserstein distance, for which central limit theorems were recently given in [70]; cf. [49, 16, 23] (as noted by Dudley in his review MR0752258 on MathSciNet, the normality assumption is not actually needed there).

3.6. Maximum likelihood estimators (MLEs)

Bounds on the closeness of the distribution of the MLE to normality in the so-called bounded Wasserstein distance, d_{bW} , were recently obtained in [2] under

certain regularity conditions. In [1], these bounds were improved in the rather common case when the MLE $\hat{\theta}$ satisfies the condition

$$q(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n g(Y_i), \quad (3.25)$$

where $q: \Theta \rightarrow \mathbb{R}$ is a twice continuously differentiable one-to-one mapping, $g: \mathbb{R} \rightarrow \mathbb{R}$ is a Borel-measurable function, and the Y_i 's are i.i.d. real-valued r.v.'s.

It is noted in [2, Proposition 2.1] that for any r.v. Y and $Z \sim N(0, 1)$ one has $d_{\text{Ko}}(Y, Z) \leq 2\sqrt{d_{\text{bW}}(Y, Z)}$, where d_{Ko} denotes the Kolmogorov distance. This bound on d_{Ko} in terms of d_{bW} is the best possible one, up a constant factor. Indeed, for each real $\varepsilon > 0$, define a r.v. Y_ε as follows: $Y_\varepsilon = \varepsilon$ if $0 < Z < \varepsilon$ and $Y_\varepsilon = Z$ otherwise. Then for any Lip(1) function $h: \mathbb{R} \rightarrow \mathbb{R}$ one has $|\mathbb{E} h(Y_\varepsilon) - \mathbb{E} h(Z)| \leq \mathbb{E} |h(Y_\varepsilon) - h(Z)| \leq \mathbb{E} |Y_\varepsilon - Z| = \int_0^\varepsilon (\varepsilon - z)\varphi(z) dz \leq \varphi(0)\varepsilon^2/2$. So, $d_{\text{bW}}(Y_\varepsilon, Z) \leq d_{\text{W}}(Y_\varepsilon, Z) \leq \varphi(0)\varepsilon^2/2$, where d_{W} is the Wasserstein distance: $d_{\text{W}}(X, Y) := \sup\{|\mathbb{E} h(X) - \mathbb{E} h(Y)| : h \in \text{Lip}(1), h \text{ bounded}\}$ for any r.v.'s X and Y . On the other hand, $d_{\text{Ko}}(Y_\varepsilon, Z) \geq \mathbb{P}(Z < \varepsilon) - \mathbb{P}(Y_\varepsilon < \varepsilon) = \Phi(\varepsilon) - 1/2 \sim \varphi(0)\varepsilon$, so that $d_{\text{Ko}}(Y_\varepsilon, Z) \geq \sqrt{2\varphi(0) - o(1)}\sqrt{d_{\text{bW}}(Y_\varepsilon, Z)}$ as $\varepsilon \downarrow 0$.

Therefore, even though the bounds on d_{bW} obtained in [2, 1] are of the optimal order $O(1/\sqrt{n})$, the resulting bounds on the Kolmogorov distance are only of the order $O(1/n^{1/4})$.

In this subsection, as an application of our general results, we shall obtain bounds of the optimal order $O(1/\sqrt{n})$ on the closeness of the distribution of the MLE to normality in the Kolmogorov distance assuming a somewhat relaxed version of the condition (3.25). In addition, we shall present a corresponding nonuniform bound. At that, our regularity conditions appear simpler than those in [2, 1].

Indeed, let here Y, Y_1, Y_2, \dots be r.v.'s mapping a measurable space (Ω, \mathcal{A}) to another measurable space $(\mathcal{X}, \mathcal{B})$ and let $(\mathbb{P}_\theta)_{\theta \in \Theta}$ be a parametric family of probability measures on (Ω, \mathcal{A}) such that the r.v.'s Y_1, Y_2, \dots are i.i.d. with respect to each of the probability measures \mathbb{P}_θ with $\theta \in \Theta$; here the parameter space Θ is assumed to be a subset of \mathbb{R} . As usual, let \mathbb{E}_θ denote the expectation with respect to the probability measure \mathbb{P}_θ . Suppose that for each $\theta \in \Theta$ the distribution $\mathbb{P}_\theta Y^{-1}$ of the r.v. Y with respect to the probability measure \mathbb{P}_θ has a density p_θ with respect to a measure μ on \mathcal{B} . For each point $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$ such that the likelihood function $\Theta \ni \theta \mapsto L_{\mathbf{x}}(\theta) := \prod_{i=1}^n p_\theta(x_i)$ has a unique maximizer, denote this maximizer by $\hat{\theta}_n(\mathbf{x})$; otherwise, assign to $\hat{\theta}_n(\mathbf{x})$ any value in Θ . Let us then refer to $\hat{\theta}_n(\mathbf{Y})$ as the MLE of θ , where $\mathbf{Y} := (Y_1, Y_2, \dots)$. Clearly, this is a more general definition of the MLE than usual, and we can even allow the function $\hat{\theta}_n$ to be non-measurable. So, the MLE $\hat{\theta}_n(\mathbf{Y})$ does not have to be a r.v. Let $\theta_0 \in \Theta$ be the "true" value of the unknown parameter θ , such that $\Theta_0 := (\theta_0 - \varepsilon, \theta_0 + \varepsilon) \subseteq \Theta$ for some real $\varepsilon > 0$.

We assume the following relaxed version of the condition (3.25): for some real constant $C > 0$ and each natural n there exists a set $E_n \in \mathcal{B}^{\otimes n}$ such that

$$\mathbb{P}_{\theta_0}(\mathbf{Y} \notin E_n) \leq C/\sqrt{n} \quad (3.26)$$

and for each point $\mathbf{x} = (x_1, \dots, x_n) \in E_n$ the value $\hat{\theta}_n(\mathbf{x})$ of the MLE belongs to the neighborhood Θ_0 of the point θ_0 and satisfies the condition

$$q(\hat{\theta}_n(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n g(x_i), \tag{3.27}$$

for some measurable function $g: \mathfrak{X} \rightarrow \mathbb{R}$ and some twice continuously differentiable mapping $q: \Theta_0 \rightarrow \mathbb{R}$ with $q'(\theta) \neq 0$ for all $\theta \in \Theta_0$, so that the mapping q is one-to-one. Suppose also that the MLE $\hat{\theta}_n(\mathbf{Y})$ is consistent at the point θ_0 , that is, $\hat{\theta}_n(\mathbf{Y}) \xrightarrow[n \rightarrow \infty]{\text{in probability}} \theta_0$ in probability with respect to the probability measure \mathbb{P}_{θ_0} ; since the MLE $\hat{\theta}_n(\mathbf{Y})$ does not have to be a r.v., the precise meaning of this consistency is that $(\mathbb{P}_{\theta_0})^*(|\hat{\theta}_n(\mathbf{Y}) - \theta_0| > \delta) \xrightarrow[n \rightarrow \infty]{} 0$ for each real $\delta > 0$, where $(\mathbb{P}_{\theta_0})^*$ denotes the outer measure induced by the probability measure \mathbb{P}_{θ_0} . Then, under the condition $\mathbb{E}_{\theta_0} |g(Y_1)| < \infty$, it follows from (3.27) by the law of large numbers that $q(\theta_0) = \mu_g := \mathbb{E}_{\theta_0} g(Y_1)$ or, equivalently, $\theta_0 = q^{-1}(\mu_g)$, where q^{-1} stands for the inverse of the function q .

Assuming further that $\sigma_g := \sqrt{\text{Var}_{\theta_0} g(Y_1)} \in (0, \infty)$, let us introduce

$$V_i := \frac{g(Y_i) - \mu_g}{\sigma_g}$$

for $i = 1, \dots, n$ and

$$f(v) := q^{-1}(\mu_g + \sigma_g v) - q^{-1}(\mu_g) = q^{-1}(\mu_g + \sigma_g v) - \theta_0$$

for real v such that $\mu_g + \sigma_g v \in q(\Theta_0)$ and $f(v) = 0$ (say) for the other real values of v . Then, in view of (3.27), on the event $\{\mathbf{Y} \notin E_n\}$ one has $f(\bar{V}) = \hat{\theta}_n(\mathbf{Y}) - \theta_0$, and at that $f(0) = 0$, $f'(0) = \sigma_g (q^{-1})'(\mu_g) = \sigma_g / q'(q^{-1}(\mu_g)) = \sigma_g / q'(\theta_0)$, and f is twice continuously differentiable in a neighborhood of 0. So, Theorem 2.10 immediately yields

Theorem 3.16. *In addition to the conditions specified above, assume that $\mathbb{E}_{\theta_0} |g(Y_1)|^3 < \infty$. Then for all $n \in \mathbb{N}$ and $z \in \mathbb{R}$*

$$\left| \mathbb{P}_{\theta_0} \left(\frac{\hat{\theta}_n(\mathbf{Y}) - \theta_0}{\sigma_g / \sqrt{n}} \leq \frac{z}{|q'(\theta_0)|} \right) - \Phi(z) \right| \leq \frac{C + \mathfrak{C}}{\sqrt{n}}, \tag{3.28}$$

where C is as in (3.26) and \mathfrak{C} is a finite expression depending only on the \mathbb{P}_θ -distributions of Y_1 for θ in a neighborhood of θ_0 . Also, if in (3.26) one can replace \sqrt{n} by n^2 , then for any $\omega > 0$ and for all real $z > 0$ and $n \in \mathbb{N}$ satisfying (2.24),

$$\left| \mathbb{P}_{\theta_0} \left(\frac{\hat{\theta}_n(\mathbf{Y}) - \theta_0}{\sigma_g / \sqrt{n}} \leq \frac{z}{|q'(\theta_0)|} \right) - \Phi(z) \right| \leq \frac{C + \mathfrak{C}}{z^3 \sqrt{n}}, \tag{3.29}$$

where \mathfrak{C} is a finite expression depending only on ω and the \mathbb{P}_θ -distributions of Y_1 for θ in a neighborhood of θ_0 .

As was noted, the MLE $\hat{\theta}_n(\mathbf{Y})$ does not have to be a r.v., and so, the \mathbb{P}_{θ_0} -probability in (3.28) and (3.29) does not have to be defined. Thus, strictly speaking, one should understand this probability as the corresponding outer or inner probability, $(\mathbb{P}_{\theta_0})^*$ or $(\mathbb{P}_{\theta_0})_*$ – each one of the two versions will do in each of the two inequalities, (3.28) and (3.29).

Let us show that, under certain mild and natural conditions, (3.27) is fulfilled if the densities p_θ form an exponential family with a natural parameter (cf. [1]), so that

$$p_\theta(x) = e^{\theta g(x) - c(\theta)} \quad (3.30)$$

for some function $c: \Theta \rightarrow \mathbb{R}$ and all $\theta \in \Theta$ and $x \in \mathcal{X}$. Here, as before, $g: \mathcal{X} \rightarrow \mathbb{R}$ is a measurable function. The natural choice of the parameter space here is $\Theta := \{\theta \in \mathbb{R}: \mathcal{E}(\theta) := \int_{\mathcal{X}} e^{\theta g(x)} \mu(dx) < \infty\}$, and then of course $c(\theta) = \ln \mathcal{E}(\theta)$ for all $\theta \in \Theta$. As before, assume that $\Theta_0 := (\theta_0 - \varepsilon, \theta_0 + \varepsilon) \subseteq \Theta$ for some real $\varepsilon > 0$. In fact, by decreasing ε if necessary, we may and shall assume that $[\theta_0 - \varepsilon, \theta_0 + \varepsilon] \subseteq \Theta$. If $\mu(\{x \in \mathcal{X}: g(x) \neq a\}) = 0$ for some real a , then for all $\theta \in \Theta$ one has $\mathcal{E}(\theta) = e^{\theta a} \mu(\mathcal{X}) < \infty$, whence $p_\theta(x) = 1/\mu(\mathcal{X})$ for $x \in \mathcal{X}$, so that the densities p_θ are the same for all $\theta \in \Theta$, and therefore parameter θ is not identifiable. Let us exclude this trivial case. Note that the function c is infinitely many times differentiable (and even real-analytic) on $\Theta_0 = (\theta_0 - \varepsilon, \theta_0 + \varepsilon)$. Moreover, its derivative c' is (strictly) increasing and hence c is strictly convex on Θ_0 , because $c''(\theta) = (\ln \mathcal{E})''(\theta) = \text{Var}_\theta g(Y_1) > 0$ for $\theta \in \Theta_0$, since the trivial case of the non-identifiability of θ has just been excluded. In particular, it follows that the condition $\sigma_g := \sqrt{\text{Var}_{\theta_0} g(Y_1)} \in (0, \infty)$ holds. At that, $\mu_g := \mathbb{E}_{\theta_0} g(Y_1) = c'(\theta_0)$.

Let now

$$E_n := \left\{ \mathbf{x} \in \mathcal{X}^n: c'(\theta_0 - \varepsilon) < \frac{1}{n} \sum_{i=1}^n g(x_i) < c'(\theta_0 + \varepsilon) \right\}. \quad (3.31)$$

By Markov's inequality,

$$\begin{aligned} \mathbb{P}_{\theta_0} \left(\frac{1}{n} \sum_{i=1}^n g(Y_i) \leq c'(\theta_0 - \varepsilon) \right) &= \mathbb{P}_{\theta_0} \left(\exp \left\{ -\varepsilon \sum_{i=1}^n g(Y_i) \right\} \geq \exp \left\{ -n\varepsilon c'(\theta_0 - \varepsilon) \right\} \right) \\ &\leq \exp \left\{ n\varepsilon c'(\theta_0 - \varepsilon) \right\} \mathbb{E}_{\theta_0} \exp \left\{ -\varepsilon \sum_{i=1}^n g(Y_i) \right\} \\ &= \exp \left\{ n\varepsilon c'(\theta_0 - \varepsilon) + nc(\theta_0 - \varepsilon) - nc(\theta_0) \right\} = e^{-n\delta(\varepsilon)}, \end{aligned}$$

where $\delta(\varepsilon) := c(\theta_0) - c(\theta_0 - \varepsilon) - c'(\theta_0 - \varepsilon)\varepsilon > 0$; the latter inequality holds because (i) the function c is strictly convex and (ii) one has $h(u+v) > h(u) + h'(u)v$ for any strictly convex differentiable function h , any u , and any nonzero v . Quite similarly, $\mathbb{P}_{\theta_0} \left(\frac{1}{n} \sum_{i=1}^n g(Y_i) \geq c'(\theta_0 + \varepsilon) \right) \leq e^{-n\delta(-\varepsilon)}$, with $\delta(-\varepsilon) > 0$. So,

$$\mathbb{P}_{\theta_0}(\mathbf{Y} \notin E_n) \leq e^{-n\delta(\varepsilon)} + e^{-n\delta(-\varepsilon)}, \quad (3.32)$$

so that condition (3.26) holds, even with n^2 in place of \sqrt{n} . On the other hand, in view of (3.31) and because c' is continuous and increasing on Θ_0 , we see that (3.27) holds for all $\mathbf{x} \in E_n$, with $q(\theta) = c'(\theta)$ for all $\theta \in \Theta_0$. Now the consistency of the MLE at point θ_0 follows because (i) by (3.32), $\mathbb{P}_{\theta_0}(\mathbf{Y} \notin E_n) \xrightarrow[n \rightarrow \infty]{} 0$ and (ii) by the law of large numbers, $\frac{1}{n} \sum_{i=1}^n g(Y_i) \xrightarrow[n \rightarrow \infty]{} \mathbb{E}_{\theta_0} g(Y_1) = \mu_g = c'(\theta_0)$ in \mathbb{P}_{θ_0} -probability.

Note finally that the condition $\mathbb{E}_{\theta_0} |g(Y_1)|^3 < \infty$ in Theorem 3.16 holds as well, since

$$\begin{aligned} \mathbb{E}_{\theta_0} \exp\{\varepsilon |g(Y_1)|\} &< \mathbb{E}_{\theta_0} \exp\{\varepsilon g(Y_1)\} + \mathbb{E}_{\theta_0} \exp\{-\varepsilon g(Y_1)\} \\ &= c(\theta_0 + \varepsilon) + c(\theta_0 - \varepsilon) < \infty. \end{aligned}$$

We have verified all the conditions needed in order to apply Theorem 3.16. In addition to this, note that in the present context of exponential families, $q'(\theta) = c''(\theta) = -\frac{\partial^2}{\partial \theta^2} \ln p_\theta(x)$ does not depend on x , whence for each $\theta \in \Theta_0$ one has $q'(\theta) = -\mathbb{E}_\theta \frac{\partial^2}{\partial \theta^2} \ln p_\theta(Y_1) = I(\theta)$, the Fisher information contained in Y_1 . Also, recall that $\sigma_g = \sqrt{\text{Var}_{\theta_0} g(Y_1)} = \sqrt{c''(\theta_0)} = \sqrt{I(\theta_0)}$. Thus, we have

Corollary 3.17. *Suppose that the conditions introduced above starting with the exponential family condition (3.30) hold. Then for all $n \in \mathbb{N}$ and $z \in \mathbb{R}$*

$$\left| \mathbb{P}_{\theta_0} \left(\hat{\theta}_n(\mathbf{Y}) - \theta_0 \leq \frac{z}{\sqrt{nI(\theta_0)}} \right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{\sqrt{n}}, \tag{3.33}$$

where \mathfrak{C} is a finite expression depending only on the \mathbb{P}_θ -distributions of Y_1 for θ in a neighborhood of θ_0 . Also, for any $\omega > 0$ and for all real $z > 0$ and $n \in \mathbb{N}$ satisfying (2.24),

$$\left| \mathbb{P}_{\theta_0} \left(\hat{\theta}_n(\mathbf{Y}) - \theta_0 \leq \frac{z}{\sqrt{nI(\theta_0)}} \right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}, \tag{3.34}$$

where \mathfrak{C} is a finite expression depending only on ω and the \mathbb{P}_θ -distributions of Y_1 for θ in a neighborhood of θ_0 .

Example 3.18. Let here $\mathcal{X} = \mathbb{R}$ and let \mathcal{B} be the Borel σ -algebra over \mathbb{R} . Let the measure μ on \mathcal{B} be defined by the formula $\mu(dx) = (x + 1)^{-3} \mathbb{I}\{x \geq 0\} dx$, and let $g(x) = x$ for all real x . Let then p_θ be as in (3.30), with $\Theta = (-\infty, 0]$. It follows that c' increases on Θ , with $c'(0-) = \int_0^\infty x(x + 1)^{-3} dx / \int_0^\infty (x + 1)^{-3} dx = 1 < \infty$. On the other hand, for each natural n , with nonzero \mathbb{P}_θ -probability for each $\theta \in \Theta$, the r.v. $\frac{1}{n} \sum_{i=1}^n g(Y_i) = \frac{1}{n} \sum_{i=1}^n Y_i$ may take arbitrarily large values, in particular values exceeding $1 = c'(0-) = \sup_{\theta \in \Theta} c'(\theta)$. So, the equality (3.25) will be violated with nonzero \mathbb{P}_θ -probability for each $\theta \in \Theta$ and for each natural n . However, Theorem 3.16 and Corollary 3.17 will hold in this situation. This shows the usefulness of the relaxed version (3.26)–(3.27) of the condition (3.25).

As shown in [64], with more effort one can utilize the “multivariate” Theorem 2.9 (rather than the “univariate” Theorem 2.10, used in this subsection)

to obtain bounds of optimal order $O(1/\sqrt{n})$ on the Kolmogorov distance for MLEs in general, without assuming (3.25) or (3.26)–(3.27). It is also shown in [64] that, again without assuming (3.25) or (3.26)–(3.27), one can obtain the corresponding nonuniform bounds of the optimal orders in n and z . All these results can be extended to the more general case of M -estimators or, even more generally, to the estimators that are zeros of estimating functions; see e.g. [28]. Indeed, the condition that p_θ is a pdf for $\theta \neq \theta_0$ is used in our proofs only in order to state that $\mathbb{E}_\theta \ell'_X(\theta) = 0$ and $\mathbb{E}_\theta \ell'_X(\theta)^2 = -\mathbb{E}_\theta \ell''_X(\theta) = I(\theta) \in (0, \infty)$. In the case of M -estimators or zeros of estimating functions, the corresponding conditions will have to be just assumed, with some other expressions in place of the Fisher information $I(\theta)$, as it is done e.g. in [51, 52], where uniform (but not nonuniform) bounds of optimal order $O(1/\sqrt{n})$ for M -estimators were obtained (via different, specialized methods): in [51] for a one-dimensional parameter space Θ and in [52] in the multidimensional case.

4. Proofs

In the proofs of Theorems 2.2 and 2.5, we start with a linearization argument mentioned in the Introduction, as is done in the delta method (which is allowed here by the smoothness condition (2.1)), in a combination with the Stein-type concentration method developed by Chen and Shao [11]. The idea of linearization of nonlinear statistics has been used in large variety of papers. However, few of them resulted in Berry–Esseen-type bounds. A notable exception in this regard is work by Bolthausen and Götze; see e.g. [9]. However, our context and methods are rather different from those in [9]. Also, our goals here are different in that we want to be able to pursue BE-type bounds with explicit and moderately sized constants.

In order to obtain bounds of the optimal order and dependence on moments, as well as possessing sufficient flexibility regarding a variety of specific applications, we first make a generalization, presented in Lemma 4.5, of Chen and Shao's uniform and nonuniform bounds. In doing so, we use a number of known (but so far hardly ever used) and new probabilistic and analytical tools, several of which were developed specifically for the needs of this paper. In particular, to obtain small constants in the uniform bound in Theorem 2.2, we also use an exact Rosenthal-type inequality from [63] and an exact bound on quantiles [59].

The proof of the nonuniform bound in Theorem 2.5 is understandably a significantly more difficult task. Aside from the introduction of various parameters (such as in (2.14)), we use there bounds on exponential moments, first found in [69], which will generally be much smaller than the more classical bounds of Hoeffding [29] and Bennett [3]; on the other hand, effective use of these bounds requires a much more delicate approach. We also use Cramér's tilt transform to bound a certain expectation; without this modification, our bounds would depend on moments higher than the third. Again, in order to obtain relatively small constant factors, we use several new exact bounds developed in [67, 56, 57, 58]. The last term in (2.17) is the nonuniform counterpart to

the term $\mathbb{P}(\|S\| > \epsilon)$ of the uniform bound, and is the source of the restriction (2.15); to ensure that this term can be bounded with a correct order and with small constants, we employ results from [54, 55].

As mentioned before, Theorem 2.9 is applicable whenever the smoothness condition (2.1) holds, together with only a few conditions on moments. In many applications, the verification of the smoothness condition (2.1) is straightforward. However, sometimes this task is rather involved, as e.g. in Theorem 3.15, whose proof relies in part on perturbation theory for linear operators [37]. In the proof of Theorem 3.12 in the infinite-dimensional case we use various results from functional analysis. The explicit bounds found in the application to Pearson's R (Corollary 3.8) make use of computer algebra to search for pseudo-optimal parameter choices and of a grid method to bound M_ϵ for arbitrary values of ϵ .

This section is organized in the following manner:

- In Subsection 4.1, we introduce several inequalities that will be used at various points in the remainder of the section; these results are presented in Lemmas 4.1–4.4.
- In Subsection 4.2, we state the aforementioned Lemma 4.5. We then link these bounds to the smoothness condition (2.1), after which Theorems 2.2 and 2.5 are fairly quickly proved. We also prove Remark 2.1, concerning the smoothness condition holding over compositions of functions.
- In Subsection 4.3, we prove Theorems 2.9 (non-explicit uniform and non-uniform bounds in the i.i.d. setting) and 2.11 (explicit uniform bound in the i.i.d. setting); these proofs use Theorems 2.2 and 2.5, along with a few of the tools presented in Subsection 4.1.
- In Subsection 4.4, we prove Lemma 4.5, employing such tools as a Stein-type concentration inequality from [11], Rosenthal-type inequalities (cf. Lemma 4.1), bounds on an exponential moment of a sum (cf. Lemma 4.2), a Cramér-type tilt transform, and exact lower bounds on the exponential moment of a Winsorized r.v. (cf. Lemma 4.4).

4.1. Toolkit for proofs

Before moving on to the proofs of results from Section 2, we collect here for ease of reference several inequalities which will be employed in the proofs. In this subsection alone, we shall assume that

$$S = \sum_i \zeta_i, \text{ where } \zeta_1, \dots, \zeta_n \text{ are independent real-valued r.v.'s.} \quad (4.1)$$

We first state bounds on the third absolute moment of S .

Lemma 4.1. *If the ζ_i in (4.1) are all zero-mean, then*

$$\|S\|_3^3 \leq \sum_i \|\zeta_i\|_3^3 + \sqrt{\frac{8}{\pi}} \left(\sum_i \|\zeta_i\|_2^2 \right)^{3/2}. \quad (4.2)$$

Otherwise, if $\mathbb{E} S$ is finite, one has

$$\|S - \mathbb{E} S\|_3^3 \leq 1.316 \sum_i \|\zeta_i\|_3^3 + 2(\sum_i \|\zeta_i\|_2^2)^{3/2}. \quad (4.3)$$

A result such as (4.2) is known as a Rosenthal-type inequality, since it was first obtained by Rosenthal in [72, Theorem 3]; however, the constants there were too large. The inequality in (4.2) follows from [63, Theorem 1.5] (take $X = 0$ a.s. in the notation there), and (4.3) follows from [67, Corollary 2] (take $g(x_1, \dots, x_n) = x_1 + \dots + x_n$ there, and cf. [67, Corollary 3] concerning the constants 1.316 and 2).

Next, we shall need upper bounds on the exponential moment $\mathbb{E} e^{\lambda S}$ and the tail probability $\mathbb{P}(S > x)$ when the ζ_i 's are a.s. uniformly bounded from above.

Lemma 4.2. *Take any real numbers $y > 0$, $m, B > 0$, and $\varepsilon \in [0, 1]$, and assume that*

$$\sum_i \mathbb{P}(\zeta_i > y) = 0, \quad \mathbb{E} S \leq m, \quad \sum_i \mathbb{E} \zeta_i^2 \leq B^2, \quad \text{and} \quad \frac{\sum_i \mathbb{E}(\zeta_i)_+^3}{B^2 y} \leq \varepsilon. \quad (4.4)$$

Then for any $\lambda \geq 0$

$$\mathbb{E} \exp\{\lambda(S - m)\} \leq \text{PU}_{\text{exp}}(\lambda, y, B, \varepsilon) := \exp\left\{\frac{\lambda^2}{2} B^2(1 - \varepsilon) + \frac{e^{\lambda y} - 1 - \lambda y}{y^2} B^2 \varepsilon\right\}. \quad (4.5)$$

Moreover,

$$\text{PU}_{\text{exp}} \text{ is nondecreasing in } B \text{ and in } \varepsilon. \quad (4.6)$$

Lemma 4.2 is implied by [69, Theorem 2] for $\varepsilon \in [0, 1]$; that we may use the inequalities in (4.4) (in lieu of the equalities used in [69]) follows from (4.6), which in turn follows because $e^t - 1 - t - t^2/2 \geq 0$ for all $t \geq 0$. That (4.5) is true when $\varepsilon = 1$ is a result by Bennett [3] and Hoeffding [29], and we let $\text{BH}_{\text{exp}}(\lambda, y, B) := \text{PU}_{\text{exp}}(\lambda, y, B, 1)$. The bound PU_{exp} can be much less than BH_{exp} when ε is significantly less than 1.

Lemma 4.3. *Under the assumptions of Lemma 4.2, let*

$$u := \frac{(x - m)_+ y}{B^2} \quad \text{and} \quad \kappa := \frac{(x - m)_+}{y}.$$

Then

$$\mathbb{P}(S > x) \leq \text{BH}_{\text{tail}}(u, \kappa) := \exp\left[\kappa\left(1 - \left(1 + \frac{1}{u}\right)\ln(1 + u)\right)\right] \quad (4.7)$$

for $u > 0$, with $\text{BH}_{\text{tail}}(0, \kappa) = 1$ for any $\kappa \geq 0$. Moreover, if the ζ_i 's are \mathfrak{X} -valued random vectors, (4.7) holds under the assumptions of (4.4) when S and ζ_i are replaced with $\|S\|$ and $\|\zeta_i\|$, respectively.

The notation BH_{tail} is used for the classical bounds of Bennett [3, (8b)] and Hoeffding [29, (2.9)]. That Lemma 4.3 remains true for \mathfrak{X} -valued random vectors is implied by [69, p. 343] (see also [68, Theorem 1]). In Lemma A.1, we present

a smaller bound PU_{tail} on the tail probability $\mathbb{P}(S > x)$, derived from (4.5) and thus also depending on ε in (4.4).

Lastly, we shall also have cause to find a lower bound on the exponential moment of a Winsorized r.v.

Lemma 4.4. *Let ζ be a zero-mean real-valued r.v. with $\sqrt{\mathbb{E}\zeta^2} \leq B$ for some $B \in (0, \infty)$. Then for any $c > 0$,*

$$\mathbb{E} \exp\{c(1 \wedge \zeta)\} \geq L_W(c, B) := \frac{a_{c,B}^2 e^c + B^2 e^{-ca_{c,B}}}{a_{c,B}^2 + B^2}, \tag{4.8}$$

where $a_{c,B}$ is the unique positive root of the function $a \mapsto \frac{a}{c} (2(e^{c+ac} - 1) - ac) - B^2$.

Lemma 4.4 is proved in [56, Theorem 2.1]; in fact, as shown there, $L_W(c, B)$ is the exact lower bound on $\mathbb{E} \exp\{c(1 \wedge \xi)\}$ over all zero-mean r.v.'s ξ with $\sqrt{\mathbb{E}\xi^2} \leq B$, and hence $L_W(c, B)$ is nonincreasing in $B \in (0, \infty)$.

4.2. Proofs of bounds in general non-i.i.d. setting

The proofs of Theorems 2.2 and 2.5 were inspired by a pair of uniform and nonuniform bounds first developed by Chen and Shao in [11], which bound $|\mathbb{P}(\tilde{T} > z) - \mathbb{P}(W > z)|$ for some nonlinear statistic \tilde{T} and a standardized linear statistic W in terms of certain moments and tail probabilities associated with W and $\Delta := |\tilde{T} - W|$. However, the results in [11] were not general enough to quite fit our needs in obtaining BE-type bounds with the correct dependence on moments; we present modifications, suitable for our purposes, in the following lemma.

Lemma 4.5. *Let \tilde{T} and $W = \sum_{i=1}^n \xi_i$ be real-valued r.v.'s defined on a common probability space, where ξ_1, \dots, ξ_n are independent zero-mean r.v.'s such that $\|W\|_2 = 1$. Take any r.v. Δ such that $|\Delta| \geq |\tilde{T} - W|$ a.s., and for each $i = 1, \dots, n$ let Δ_i be any r.v. such that the r.v. ξ_i and family $(\Delta_i, (\xi_j : j \neq i))$ of r.v.'s are independent. Then for all $c_* \in (0, 1)$ and $z \in \mathbb{R}$*

$$|\mathbb{P}(\tilde{T} \leq z) - \mathbb{P}(W \leq z)| \leq \frac{1}{2c_*} (4\delta + \mathbb{E}|W\Delta| + \sum_i \mathbb{E}|\xi_i(\Delta - \Delta_i)|), \tag{4.9}$$

where δ satisfies (2.9). Also, for any positive numbers $\theta, w, \delta_0, \pi_1, \pi_2$, and π_3 such that $\delta_0 \leq w$ and $\pi_1 + \pi_2 + \pi_3 = 1$,

$$|\hat{\mathbb{P}}(\tilde{T} \leq z) - \hat{\mathbb{P}}(W \leq z)| \leq \tilde{\gamma}_z + \tilde{\tau} e^{-(1-\pi_1)z/\theta} \tag{4.10}$$

for all $z \geq 0$, where

$$\hat{\mathbb{P}}(E) := \mathbb{P}(E \cap \{|\Delta| \leq \pi_1 z\}) \text{ for any } \mathbb{P}\text{-measurable event } E, \tag{4.11}$$

$$\tilde{\gamma}_z := \sum_i \mathbb{P}(\xi_i > \pi_2 z) + \sum_i \mathbb{P}(W - \xi_i \geq \pi_3 z) \mathbb{P}(\xi_i > w), \tag{4.12}$$

$$\tilde{\tau} := c_1 \sum_i \|\xi_i\|_3 \|\Delta - \Delta_i\|_{3/2} + c_2 \|\Delta\|_{3/2} + c_3 \delta, \quad (4.13)$$

and

$$c_1 := \frac{1}{c_*} \text{PU}_{\text{exp}}\left(\frac{3}{\theta}, w, \frac{1}{\sqrt{3}}, \varepsilon_1\right) e^{\delta_0/\theta}, \quad (4.14)$$

$$c_2 := c_1 \left((1.316a_1 e^{3w/\theta})^{1/3} \sigma_3 + (2^{2/3} a_1 e^{3w/\theta})^{1/2} + (e^{3w/\theta} - 1)/w \right), \quad (4.15)$$

$$c_3 := \left(2c_2 + \frac{1}{c_*} \sqrt{2} \text{PU}_{\text{exp}}\left(\frac{2}{\theta}, w, \frac{1}{\sqrt{2}}, \varepsilon_1\right) \right) \vee \left(\frac{1}{\delta_0} \text{PU}_{\text{exp}}\left(\frac{1}{\theta}, w, 1, \varepsilon_1\right) \right), \quad (4.16)$$

$$\varepsilon_1 := \frac{\sigma_3^3}{w} \wedge 1, \quad (4.17)$$

$$a_1 := 1/L_W(3w/\theta, \max_i \|\xi_i\|_2/w). \quad (4.18)$$

The proof of (4.9) is nearly identical to that of [11, (2.3)] (as suggested by the similarity of those two inequalities). In contrast, the proof of (4.10) requires significant changes to the proof of [11, (2.6)]; as the proof of Lemma 4.5 is relatively lengthy, we defer it to Subsection 4.4.

Assume that the conditions of Theorem 2.2 and 2.5 are satisfied; particularly, we have a nonlinear functional f satisfying the smoothness condition (2.1), along with the statistics T and W defined in (2.5). Then let

$$\tilde{T} := T \mathbf{I}\{\|S\| \leq \epsilon\} + W \mathbf{I}\{\|S\| > \epsilon\} \quad (4.19)$$

and

$$\Delta := \frac{M_\epsilon}{2\sigma} \|S\|^2 \quad \text{and} \quad \Delta_i = \frac{M_\epsilon}{2\sigma} \|S - X_i\|^2 \quad \text{for each } i = 1, \dots, n. \quad (4.20)$$

By (2.1),

$$|\tilde{T} - W| = \sigma^{-1} |f(S) - L(S)| \mathbf{I}\{\|S\| \leq \epsilon\} \leq \frac{M_\epsilon}{2\sigma} \|S\|^2 = \Delta; \quad (4.21)$$

moreover, ξ_i and $(\Delta_i, (\xi_j : j \neq i))$ are independent for each $i = 1, \dots, n$, and thus the conditions of Lemma 4.5 are met. Theorems 2.2 and 2.5 are then fairly easily proved by substituting the above \tilde{T} , Δ , and Δ_i into the bounds of Lemma 4.5; the following lemma (which will be proved at the end of this subsection) provides some bounds which bridge the Δ and Δ_i defined above to the expressions \mathbf{u} and \mathbf{v} found in (2.11).

Lemma 4.6. *Under the conditions of Theorem 2.2,*

$$\|\Delta\|_{3/2} \leq \mathbf{u} \quad \text{and} \quad \sum_i \|\xi_i\|_3 \|\Delta - \Delta_i\|_{3/2} \leq \sigma_3 \mathbf{v}. \quad (4.22)$$

Proof of Theorem 2.2. By (4.19), (4.9), and Hölder's inequality,

$$\begin{aligned} |\mathbb{P}(T \leq z) - \mathbb{P}(W \leq z)| &\leq \mathbb{P}(\|S\| > \epsilon) + |\mathbb{P}(\tilde{T} \leq z) - \mathbb{P}(W \leq z)| \\ &\leq \frac{1}{2c_*} (4\delta + \|W\|_3 \|\Delta\|_{3/2} + \sum_i \|\xi_i\|_3 \|\Delta - \Delta_i\|_{3/2}) + \mathbb{P}(\|S\| > \epsilon) \end{aligned}$$

for all $z \in \mathbb{R}$. Next, $\|W\|_3 \leq \sigma_3 + (8/\pi)^{1/6}$, by the Rosenthal-type inequality (4.2); recall here also the definition of σ_3 in (2.7). So, (2.10) follows by Lemma 4.6, which completes the proof. \square

Proof of Theorem 2.5. First note that, by (4.20), (2.15), and the last inequality in (2.14),

$$\begin{aligned} \{|\Delta| \leq \pi_1 z\} &= \{\|S\| \leq (2\pi_1 \sigma z / M_\epsilon)^{1/2}\} \subseteq \{\|S\| \leq (2\pi_1 \omega / M_\epsilon)^{1/2}\} \\ &\subseteq \{\|S\| \leq \epsilon\}. \end{aligned}$$

Recalling now the respective definitions (4.19) and (4.11) of \tilde{T} and $\hat{\mathbb{P}}$, we have

$$\begin{aligned} |\mathbb{P}(T \leq z) - \mathbb{P}(W \leq z)| &\leq |\hat{\mathbb{P}}(\tilde{T} \leq z) - \hat{\mathbb{P}}(W \leq z)| + \mathbb{P}(|\Delta| > \pi_1 z) \\ &\leq \tilde{\gamma}_z + \tilde{\tau} e^{-(1-\pi_1)z/\theta} + \mathbb{P}(|\Delta| > \pi_1 z) \\ &= \gamma_z + \tilde{\tau} e^{-(1-\pi_1)z/\theta} \end{aligned}$$

for all z as in (2.15), with the last inequality following from (4.10) of Lemma 4.5 and the equality following from the definitions (2.17), (4.12), and (4.20) of γ_z , $\tilde{\gamma}_z$, and Δ , respectively. Since $\tilde{\tau} \leq \tau$ by Lemma 4.6 (cf. the definitions (2.18) and (4.13) of τ and $\tilde{\tau}$), the inequality in (2.16) follows and the proof is complete. \square

Proof of Remark 2.1. In view of (2.2), there exists $m_h \in (0, \infty)$ such that

$$\|h(x)\|_{\mathfrak{Y}} \leq m_h \|x\|_{\mathfrak{X}} \quad \text{for all } x \in \mathfrak{X} \text{ with } \|x\|_{\mathfrak{X}} \leq \epsilon_h; \quad (4.23)$$

indeed, we may let $m_h := \|L_h\| + M_h \epsilon_h / 2$. Assume that ϵ_h is chosen small enough to ensure $m_h \epsilon_h \leq \epsilon_g$.

Take any $x \in \mathfrak{X}$ with $\|x\|_{\mathfrak{X}} \leq \epsilon_h$. Then, by (2.2), there is some $y_x \in \mathfrak{Y}$ such that $\|y_x\|_{\mathfrak{Y}} \leq 1$ and $h(x) = L_h(x) + \frac{1}{2} M_h \|x\|_{\mathfrak{X}}^2 y_x$. By (4.23), $\|h(x)\|_{\mathfrak{Y}} \leq m_h \epsilon_h \leq \epsilon_g$, and so, by (2.3), there is some $z_x \in \mathfrak{Z}$ such that $\|z_x\|_{\mathfrak{Z}} \leq 1$ and

$$\begin{aligned} g(h(x)) &= L_g(h(x)) + \frac{1}{2} M_g \|h(x)\|_{\mathfrak{Y}}^2 z_x \\ &= L_g(L_h(x)) + \frac{1}{2} M_h \|x\|_{\mathfrak{X}}^2 L_g(y_x) + \frac{1}{2} M_g \|h(x)\|_{\mathfrak{Y}}^2 z_x. \end{aligned}$$

Thus, by (4.23) (recall also $\|y_x\|_{\mathfrak{Y}} \leq 1$ and $\|z_x\|_{\mathfrak{Z}} \leq 1$),

$$\begin{aligned} \|(g \circ h)(x) - (L_g \circ L_h)(x)\|_{\mathfrak{Z}} &\leq \frac{1}{2} (M_h \|L_g\| + M_g m_h^2) \|x\|_{\mathfrak{X}}^2 \\ &\quad \text{for all } x \in \mathfrak{X} \text{ with } \|x\|_{\mathfrak{X}} \leq \epsilon_h; \end{aligned}$$

that is, (2.1) with \mathfrak{Z} in place of \mathbb{R} holds for $f = g \circ h$ with $L = L_g \circ L_h$, $M_\epsilon = M_h \|L_g\| + M_g m_h^2$, and $\epsilon = \epsilon_h$. \square

Proof of Lemma 4.6. By (4.20) and [67, (12)], we have

$$\|\Delta\|_{3/2} = \frac{M_\epsilon}{2\sigma} \|S\|_3^2 \leq \frac{M_\epsilon}{2\sigma} \left(\frac{1+D^2}{2} (s_3^3 + 2s_2^3) \right)^{2/3} \leq u,$$

where u is as in (2.11).

For each $i = 1, \dots, n$, (4.20) implies

$$\begin{aligned} \frac{2\sigma}{M_\epsilon} |\Delta - \Delta_i| &= \left| \|S\|^2 - \|S - X_i\|^2 \right| = \left| \|S\| - \|S - X_i\| \right| (\|S\| + \|S - X_i\|) \\ &\leq \|X_i\| (\|X_i\| + 2\|S - X_i\|) = \|X_i\|^2 + 2\|X_i\| \|S - X_i\|. \end{aligned}$$

Also, by (2.8), $\|S - X_i\|_{3/2} \leq \|S - X_i\|_2 \leq Ds_2$. It follows that

$$\|\Delta - \Delta_i\|_{3/2} \leq \frac{M_\epsilon}{2\sigma} (\|X_i\|_3^2 + 2\|X_i\|_{3/2} \|S - X_i\|_{3/2}) \leq \frac{M_\epsilon}{2\sigma} (\|X_i\|_3^2 + 2Ds_2 \|X_i\|_{3/2}).$$

Then Hölder’s inequality yields

$$\begin{aligned} \sum_{i=1}^n \|\xi_i\|_3 \|\Delta - \Delta_i\|_{3/2} &\leq \frac{M_\epsilon}{2\sigma} \sum_{i=1}^n \|\xi_i\|_3 (\|X_i\|_3^2 + 2Ds_2 \|X_i\|_{3/2}) \\ &\leq \frac{M_\epsilon}{2\sigma} \sigma_3 (s_3^2 + 2Ds_2 s_{3/2}) = \sigma_3 \mathbf{v}, \end{aligned}$$

by the definition (2.11) of \mathbf{v} . □

4.3. Proofs of bounds in i.i.d. setting

In this subsection we prove Theorems 2.11 and 2.9, concerning our BE-type bounds in the i.i.d. setting. Recall the i.i.d. notation as introduced in (2.20) and (2.21), along with the subsequent substitutions for σ , ξ_i , s_p , and σ_p that can be made (cf. (2.22)).

Theorem 2.11 follows nearly directly from Theorem 2.2:

Proof of Theorem 2.11. Concerning the bound in Theorem 2.2,

- (i) let $D = 1$ since \mathfrak{X} is assumed to be a Hilbert space;
- (ii) use Remark 2.3, along with the inequality $\sigma_1 = \sqrt{n}\varsigma_1 \leq \sqrt{n}\varsigma_2 = \sqrt{n}$, to set

$$\delta = \frac{\varsigma_3^3 - (2c_* - 1)^2}{4(1 - c_*)\sqrt{n}}; \tag{4.24}$$

- (iii) use the inequality $v_{3/2} \leq v_2$ (recall also the definitions (2.11)) to assert

$$\mathbf{u} = \frac{1}{\sqrt{n}} \frac{M_\epsilon}{2\tilde{\sigma}} \left(\frac{v_3^2}{n^{1/3}} + 2^{2/3} v_2^2 \right) \quad \text{and} \quad \sigma_3 \mathbf{v} \leq \frac{1}{\sqrt{n}} \frac{M_\epsilon}{2\tilde{\sigma}} \varsigma_3 \left(\frac{v_3^2}{n^{1/2}} + 2v_2^2 \right); \tag{4.25}$$

- (iv) and use Remark 2.4 to see that $\mathbb{P}(\|S\| > \epsilon) \leq \mathfrak{R}_\epsilon$ (cf. the definition (2.30)).

Note also that in (2.19), we may make the substitutions $n\|\xi_1\|_3^3 = \varsigma_3^3/\sqrt{n}$ and $n\|\xi_1\|_2^3 = 1/\sqrt{n}$. Upon using the abovementioned substitutions and inequalities in (2.10) and (2.19), (2.29) follows. □

The proof of Theorem 2.9 is also a fairly straightforward application of Theorems 2.2 and 2.5. Most of the work involves ensuring the last term of γ_z in (2.17), the presence of which stems in part from the use of the smoothness condition (2.1) (cf. Remark 2.7), can be bounded on the order of $1/(z^3\sqrt{n})$ with the moment assumption $v_3 < \infty$. For this we use truncation of the V_i ’s in tandem with Lemma 4.3. In Appendix A, we use a stronger bound on the tail probability of $\|\bar{V}\|$, provided by Lemma A.1.

Proof of Theorem 2.9. The proof of (2.23) is virtually the same as that of (2.29) in Theorem 2.11, only we do not assume that \mathfrak{X} is a Hilbert space, hence nor $D = 1$, when obtaining explicit expressions for \mathfrak{C} . It should be clear that (2.26) follows from (2.25) after an application of the Markov inequality.

Throughout the rest of the proof, let us write $a \leq b$ if $|a| \leq \mathfrak{C}b$ for some \mathfrak{C} as in Theorem 2.9.

To prove (2.25), set $c_* = \frac{1}{2}$, $w = \delta_0 = 1$, $\pi_1 = ((M_\epsilon \epsilon^2 / (2\tilde{\sigma}\omega)) \wedge \frac{1}{3})$, $\pi_2 = \pi_3 = \frac{1}{2}(1 - \pi_1)$, and $\theta = 1 - \pi_1$. Then the conditions of Theorem 2.5 are met, with $\tilde{\sigma}\omega$ replacing ω there.

Choose δ as in (4.24), and refer to (4.25) to see that

$$\tau e^{-(1-\pi_1)z/\theta} \leq \frac{e^{-z}}{\sqrt{n}}. \tag{4.26}$$

Recall (2.6) and (2.22) to see that $\mathbb{P}(\xi_1 > x) \leq \mathbb{P}(\|V\| > x\tilde{\sigma}\sqrt{n}/\|L\|)$ for any $x > 0$; also apply the Markov inequality and the Rosenthal-type inequality (4.2) of Lemma 4.1 to see that $\mathbb{P}(W - \xi_1 \geq \pi_3 z) \leq 1/z^3$. Then the definition (2.17) of γ_z implies

$$\gamma_z \leq n \mathbb{P}(\|V\| > \mathfrak{C}z\sqrt{n}) + \mathfrak{C} \frac{n \mathbb{P}(\|V\| > \mathfrak{C}\sqrt{n})}{z^3} + \mathbb{P}\left(\|\bar{V}\| > \left(\frac{2\pi_1 \tilde{\sigma} z}{M_\epsilon \sqrt{n}}\right)^{1/2}\right). \tag{4.27}$$

Next, choose any $\kappa_2 > 0$, and let

$$c_x := \left(\frac{2\pi_1}{M_\epsilon}\right)^{1/2}, \quad x_2 := c_x \left(\frac{\tilde{\sigma} z}{\sqrt{n}}\right)^{1/2}, \quad y_2 := \frac{x_2}{\kappa_2}, \tag{4.28}$$

$$\bar{V}_{y_2} := \frac{1}{n} \sum_{i=1}^n V_i \mathbb{I}\{\|V_i\| \leq ny_2\}.$$

Then

$$\begin{aligned} \mathbb{P}\left(\|\bar{V}\| > \left(\frac{2\pi_1 \tilde{\sigma} z}{M_\epsilon \sqrt{n}}\right)^{1/2}\right) &= \mathbb{P}(\|\bar{V}\| > x_2) \\ &\leq \mathbb{P}(\max_i \|V_i\| > ny_2) + \mathbb{P}(\|\bar{V}_{y_2}\| > x_2) \\ &\leq n \mathbb{P}(\|V\| > ny_2) + \mathbb{P}(\|\bar{V}_{y_2}\| > x_2) \end{aligned} \tag{4.29}$$

$$\leq n \mathbb{P}(\|V\| > \mathfrak{C}z\sqrt{n}) + \mathbb{P}(\|\bar{V}_{y_2}\| > x_2), \tag{4.30}$$

with (4.30) following from (2.24) and $ny_2 \geq ny_2 \sqrt{z/(\omega\sqrt{n})} = \mathfrak{C}z\sqrt{n}$.

We assert that

$$\mathbb{P}(\|\bar{V}_{y_2}\| > x_2) \leq \left(\frac{6eM_\epsilon}{\pi_1 \tilde{\sigma}} \frac{v_2^2}{z\sqrt{n}}\right)^3. \tag{4.31}$$

Assume w.l.o.g. that the upper bound in (4.31) is no greater than 1, and choose $\kappa_2 = 6$. Then

$$\mathbb{E}\|\bar{V}_{y_2}\| \leq \mathbb{E}\|\bar{V}\| + \mathbb{E}\|\bar{V} - \bar{V}_{y_2}\|$$

$$\begin{aligned}
&\leq \|\bar{V}\|_2 + \frac{1}{n} \mathbb{E} \left\| \sum_i V_i \mathbb{I} \{ \|V_i\| > ny_2 \} \right\| \\
&\leq \frac{v_2}{\sqrt{n}} + \frac{v_2^2}{ny_2} \\
&= \frac{x_2}{4} \left(\left(\frac{8M_\epsilon}{\pi_1 \tilde{\sigma}} \frac{v_2^2}{z\sqrt{n}} \right)^{1/2} + \frac{12M_\epsilon}{\pi_1 \tilde{\sigma}} \frac{v_2^2}{z\sqrt{n}} \right) \\
&< \frac{x_2}{4} \left(\left(\frac{6eM_\epsilon}{\pi_1 \tilde{\sigma}} \frac{v_2^2}{z\sqrt{n}} \right)^{1/2} + \frac{6eM_\epsilon}{\pi_1 \tilde{\sigma}} \frac{v_2^2}{z\sqrt{n}} \right) \leq \frac{x_2}{2},
\end{aligned}$$

where the equality above follows from the definitions of x_2 and y_2 in (4.28).

Then Lemma 4.3 implies (use $x = x_2$, $y = y_2$, $B = v_2/\sqrt{n}$, and $m = x_2/2$ in (4.4))

$$\begin{aligned}
\mathbb{P}(\|\bar{V}_{y_2}\| > x_2) &\leq \text{BH}_{\text{tail}}\left(\frac{x_2 y_2 / 2}{v_2^2 / n}, \frac{x_2 / 2}{y_2}\right) = \text{BH}_{\text{tail}}\left(\frac{x_2^2}{2\kappa_2 v_2^2 / n}, \frac{\kappa_2}{2}\right) \\
&\leq \left(\frac{2e\kappa_2}{x_2^2} \frac{v_2^2}{n}\right)^{\kappa_2 / 2},
\end{aligned}$$

with the last inequality following from $\text{BH}_{\text{tail}}(u, \kappa) \leq (e/u)^\kappa$ whenever $u > 0$ (which in turn is easily verified by referring to (4.7)); (4.31) follows after recalling $\kappa_2 = 6$ and the definition (4.28) of x_2 .

Referring to (4.26), (4.27), (4.30), and (4.31), (2.25) follows when $\Phi(z)$ is replaced by $\mathbb{P}(\sqrt{n}|L(\bar{V})| > \tilde{\sigma}z)$ there. To obtain (2.25) as stated, note that

$$\left| \mathbb{P}\left(\frac{L(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\sigma_3^3}{e^z} + n \mathbb{P}(\xi_1 > \frac{2}{5}z) + \frac{n \mathbb{P}(\xi_1 > 1)}{z^3} \quad (4.32)$$

for all $z > 0$; this follows by [55, Corollary 1.3] with $v = w = 1$, $p = 3$, $c = 0$, and $\lambda = 1$ (in notation therein), using at that the inequalities $\beta_v \leq \mu_p/v^p$ (displayed right after [55, (1.2)]) and $P_1 \wedge \dots \wedge P_5 \leq P_4$. \square

4.4. Proof of Lemma 4.5

The inequality in (4.9) is very similar to one proved by Chen and Shao in [11]. From the condition that $|\Delta| \geq |\tilde{T} - W|$ follows the concentration inequality

$$-\mathbb{P}(z - |\Delta| \leq W \leq z) \leq \mathbb{P}(\tilde{T} \leq z) - \mathbb{P}(W \leq z) \leq \mathbb{P}(z \leq W \leq z + |\Delta|) \quad (4.33)$$

for all $z \in \mathbb{R}$. Upon replacing [11, (2.2)] by the inequality (2.9) in the present paper, the proof of (4.9) is then, *mutatis mutandis*, identical to the proof of [11, (2.3)].

The differences between the proof of (4.10) and the analogous inequality in [11, Theorem 2.2] are significant enough to warrant more details here; particularly, aside from the use of the measure \mathbb{P} and the introduction of various parameters (e.g. c_* , θ , w , δ_0 , and π_i), we also employ a Cramér-tilt absolutely continuous transformation of measure along with the previously mentioned Rosenthal-type

and exponential bounds. For the ease of comparison between the two proofs, we shall use notation similar to that in [11].

Introduce the Winsorized r.v.'s

$$\bar{\xi}_i := \xi_i \wedge w \quad \text{and their sum,} \quad \bar{W} := \sum_{i=1}^n \bar{\xi}_i. \tag{4.34}$$

In view of the definition (4.11) of the measure $\hat{\mathbb{P}}$,

$$\begin{aligned} & \hat{\mathbb{P}}(z - |\Delta| \leq W \leq z) \\ & \leq \sum_{i=1}^n \mathbb{P}(W \geq (1 - \pi_1)z, \xi_i > w) + \hat{\mathbb{P}}(z - |\Delta| \leq W \leq z, \max_i \xi_i \leq w) \\ & \leq \sum_{i=1}^n \mathbb{P}(\xi_i > \pi_2 z) + \sum_{i=1}^n \mathbb{P}(W - \xi_i \geq \pi_3 z) \mathbb{P}(\xi_i > w) \\ & \qquad \qquad \qquad + \hat{\mathbb{P}}(z - |\Delta| \leq \bar{W} \leq z) \\ & = \tilde{\gamma}_z + \hat{\mathbb{P}}(z - |\Delta| \leq \bar{W} \leq z). \end{aligned} \tag{4.35}$$

The second inequality above follows from the independence of $W - \xi_i$ and ξ_i , the assumption that $\pi_3 = 1 - \pi_1 - \pi_2$ (recall (2.14)), and the definition (4.34) of \bar{W} , while the above equality follows from the definition (4.12) of $\tilde{\gamma}_z$; compare (4.35) with [11, Lemma 5.1].

We must next establish the inequality

$$\hat{\mathbb{P}}(z - |\Delta| \leq \bar{W} \leq z) \leq \tilde{\tau} e^{-(1-\pi_1)z/\theta}; \tag{4.36}$$

cf. [11, Lemma 5.2]. Consider two cases:

$$(i) \delta > \delta_0 \quad \text{and} \quad (ii) 0 < \delta \leq \delta_0 \leq w$$

(recall the restriction on the number δ_0 in (2.14)). In the first case, when $\delta > \delta_0$, one has

$$\begin{aligned} \hat{\mathbb{P}}(z - |\Delta| \leq \bar{W} \leq z) & \leq \mathbb{P}(\bar{W} \geq (1 - \pi_1)z) \leq \mathbb{E} e^{\bar{W}/\theta} e^{-(1-\pi_1)z/\theta} \\ & \leq \frac{\delta}{\delta_0} \text{PU}_{\text{exp}}\left(\frac{1}{\theta}, w, 1, \varepsilon_1\right) e^{-(1-\pi_1)z/\theta} \leq c_3 \delta e^{-(1-\pi_1)z/\theta} \leq \tilde{\tau} e^{-(1-\pi_1)z/\theta}; \end{aligned}$$

here (4.5) and (4.4) are used for the third inequality above (as well as the definitions (4.17) and (2.7) of ε_1 and σ_3), and the definitions (4.16) and (4.13) of c_3 and $\tilde{\tau}$ are used for the last two inequalities there. Thus, (4.36) is established when $\delta > \delta_0$.

Consider now the second case, when $0 < \delta \leq \delta_0 \leq w$. Let

$$f_{\Delta}(u) := \begin{cases} 0 & \text{if } u < z - |\Delta| - \delta, \\ e^{u/\theta}(u - z + |\Delta| + \delta) & \text{if } z - |\Delta| - \delta \leq u < z + \delta, \\ e^{u/\theta}(|\Delta| + 2\delta) & \text{if } u \geq z + \delta \end{cases}$$

be defined similarly to [11, (5.16)]. Then, by the independence of $(\Delta_i, \overline{W} - \bar{\xi}_i)$ and ξ_i ,

$$\mathbb{E} W f_{\Delta}(\overline{W}) = G_1 + G_2, \tag{4.37}$$

where

$$G_1 := \sum_{i=1}^n \mathbb{E} \xi_i (f_{\Delta}(\overline{W}) - f_{\Delta}(\overline{W} - \bar{\xi}_i)), \quad G_2 := \sum_{i=1}^n \mathbb{E} \xi_i (f_{\Delta}(\overline{W} - \bar{\xi}_i) - f_{\Delta_i}(\overline{W} - \bar{\xi}_i)).$$

Also, using an obvious modification of the arguments associated with [11, (5.17)–(5.19)], one has

$$G_1 \geq G_{1,1} - G_{1,2}, \tag{4.38}$$

where

$$G_{1,1} := c_* \exp\left\{\frac{1}{\theta}((1 - \pi_1)z - \delta)\right\} \hat{\mathbb{P}}(z - |\Delta| \leq \overline{W} \leq z), \tag{4.39}$$

$$G_{1,2} := \mathbb{E} \int_{|t| \leq \delta} e^{(\overline{W} - \delta)/\theta} |\overline{M}(t) - \mathbb{E} \overline{M}(t)| dt,$$

$$\overline{M}(t) := \sum_{i=1}^n \overline{M}_i(t), \quad \text{and} \quad \overline{M}_i(t) := \xi_i (\mathbb{I}\{-\bar{\xi}_i \leq t \leq 0\} - \mathbb{I}\{0 < t \leq -\bar{\xi}_i\});$$

in particular, the factor c_* in the expression (4.39) for $G_{1,1}$ arises when one uses the relations $\int_{|t| \leq \delta} \mathbb{E} \overline{M}(t) dt = \sum_i \mathbb{E} |\xi_i| (\delta \wedge |\xi_i|) \geq c_*$, which in turn follow by the condition $\delta \leq \delta_0 \leq w$ of case (ii) and (2.9); cf. [11, (5.19)]. Further,

$$\int_{|t| \leq \delta} \mathbb{E} (\overline{M}(t) - \mathbb{E} \overline{M}(t))^2 dt \leq \sum_{i=1}^n \mathbb{E} \int_{|t| \leq \delta} \overline{M}_i(t)^2 dt = \sum_{i=1}^n \mathbb{E} \xi_i^2 (\delta \wedge |\bar{\xi}_i|) \leq \delta,$$

so that two applications of the Cauchy-Schwarz inequality yield

$$\begin{aligned} G_{1,2} &\leq \mathbb{E} \left(\int_{|t| \leq \delta} e^{2(\overline{W} - \delta)/\theta} dt \right)^{1/2} \left(\int_{|t| \leq \delta} (\overline{M}(t) - \mathbb{E} \overline{M}(t))^2 dt \right)^{1/2} \\ &\leq \left(2\delta \mathbb{E} e^{2(\overline{W} - \delta)/\theta} \right)^{1/2} \sqrt{\delta} \\ &\leq \left(2 \text{PU}_{\exp}\left(\frac{2}{\theta}, w, 1, \varepsilon_1\right) \right)^{1/2} e^{-\delta/\theta} \delta = \sqrt{2} \text{PU}_{\exp}\left(\frac{2}{\theta}, w, \frac{1}{\sqrt{2}}, \varepsilon_1\right) e^{-\delta/\theta} \delta, \end{aligned} \tag{4.40}$$

where the last inequality follows from (4.5) and (4.4), in view of the definitions (4.34) and (4.17) of \overline{W} and ε_1 ; the equality in (4.40) follows from the easily verified identity

$$\text{PU}_{\exp}(\lambda, y, B, \varepsilon)^\alpha = \text{PU}_{\exp}\left(\lambda, y, \alpha^{1/2} B, \varepsilon\right) \quad \text{for any } \alpha > 0. \tag{4.41}$$

Next (cf. [11, (5.21)]),

$$|G_2| \leq \sum_{i=1}^n \mathbb{E} |\xi_i e^{(\overline{W} - \bar{\xi}_i)/\theta} (\Delta - \Delta_i)| \leq \sum_{i=1}^n \|\xi_i e^{(\overline{W} - \bar{\xi}_i)/\theta}\|_3 \|\Delta - \Delta_i\|_{3/2}$$

$$\begin{aligned}
 &= \sum_{i=1}^n \mathbb{E}^{1/3} e^{\frac{3}{\theta}(\bar{W}-\bar{\xi}_i)} \|\xi_i\|_3 \|\Delta - \Delta_i\|_{3/2} \\
 &\leq \text{PU}_{\text{exp}}\left(\frac{3}{\theta}, w, \frac{1}{\sqrt{3}}, \varepsilon_1\right) \sum_{i=1}^n \|\xi_i\|_3 \|\Delta - \Delta_i\|_{3/2}.
 \end{aligned} \tag{4.42}$$

Also,

$$\mathbb{E} W f_{\Delta}(\bar{W}) \leq \mathbb{E} (|\Delta| + 2\delta) |W| e^{\bar{W}/\theta} \leq (\|\Delta\|_{3/2} + 2\delta) \|W e^{\bar{W}/\theta}\|_3. \tag{4.43}$$

In the proof in [11], the term $\mathbb{E} W^2 e^{\bar{W}}$ was bounded from above; in our case, more work is required to bound the last factor in (4.43). Specifically, we apply Cramér’s tilt transform to the ξ_i ’s, using at that results of [56, 57, 58].

Let $\boldsymbol{\xi} := (\xi_1, \dots, \xi_n)$, and for any real $c > 0$ let $\hat{\boldsymbol{\xi}} := (\hat{\xi}_1, \dots, \hat{\xi}_n)$ be a random vector such that

$$\mathbb{P}(\hat{\boldsymbol{\xi}} \in E) = \frac{\mathbb{E} e^{c\bar{W}} \mathbb{I}\{\boldsymbol{\xi} \in E\}}{\mathbb{E} e^{c\bar{W}}}$$

for all Borel sets $E \subseteq \mathbb{R}^n$. Then the $\hat{\xi}_i$ ’s are necessarily independent r.v.’s; moreover, if $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is any nonnegative Borel function, then

$$\mathbb{E} f(\hat{\boldsymbol{\xi}}) = \frac{\mathbb{E} f(\boldsymbol{\xi}) e^{c\bar{W}}}{\mathbb{E} e^{c\bar{W}}}. \tag{4.44}$$

By [57, Proposition 2.6,(I)], $\mathbb{E} \hat{\xi}_i$ is nondecreasing in c , so that $\mathbb{E} \hat{\xi}_i \geq \mathbb{E} \xi_i = 0$, and so, by [57, Corollary 2.7],

$$\left| \sum_i \mathbb{E} \hat{\xi}_i \right| = \sum_i \mathbb{E} \hat{\xi}_i \leq \frac{e^{cw} - 1}{w} \sum_i \mathbb{E} \xi_i^2 = \frac{e^{cw} - 1}{w}.$$

If the ξ_i ’s are assumed to have symmetric distributions, then [58, Theorem 1] allows for the factor $e^{cw} - 1$ above to be replaced by $\sinh(cw)$; cf. Remark 2.8. Choose now

$$c = \frac{3}{\theta}.$$

Then, by [56, Theorem 2.1],

$$\begin{aligned}
 \mathbb{E} e^{c\bar{\xi}_i} &= \mathbb{E} e^{c(\xi_i \wedge w)} = \mathbb{E} e^{cw(1 \wedge \xi_i/w)} \geq L_W(cw, \|\xi_i\|_2/w) \\
 &\geq L_W(cw, \max_i \|\xi_i\|_2/w) = a_1^{-1},
 \end{aligned}$$

where a_1 is as defined in (4.18); the last inequality above follows because $L_W(c, \sigma)$ in [56, (2.9)] is nonincreasing in σ ; the condition $c = \frac{3}{\theta}$ was used here in the above display only for the last equality. So,

$$\mathbb{E} |\hat{\xi}_i|^\alpha = \frac{\mathbb{E} |\xi_i|^\alpha e^{c\bar{\xi}_i}}{\mathbb{E} e^{c\bar{\xi}_i}} \leq a_1 e^{cw} \mathbb{E} |\xi_i|^\alpha$$

for $\alpha \in \{2, 3\}$, with $\sum_i \mathbb{E} \hat{\xi}_i^2 \leq a_1 e^{cw}$ a consequence of this. Next,

$$\begin{aligned} \|\sum_i \hat{\xi}_i\|_3 &\leq \|\sum_i (\hat{\xi}_i - \mathbb{E} \hat{\xi}_i)\|_3 + |\sum_i \mathbb{E} \hat{\xi}_i| \\ &\leq (1.316 \sum_i \mathbb{E} |\hat{\xi}_i|^3)^{1/3} + 2^{1/3} (\sum_i \mathbb{E} \hat{\xi}_i^2)^{1/2} + (e^{3w/\theta} - 1)/w \quad (4.45) \\ &\leq (1.316 a_1 e^{3w/\theta} \sigma_3^3)^{1/3} + 2^{1/3} (a_1 e^{3w/\theta})^{1/2} + (e^{3w/\theta} - 1)/w, \end{aligned}$$

where (4.3) is used in the second inequality above. Letting $f(x_1, \dots, x_n) \equiv |\sum_i x_i|^3$ in (4.44) and using (4.5), (4.4), and (4.41) once more, one has

$$\begin{aligned} \|W e^{\overline{W}/\theta}\|_3 &= \left(\mathbb{E} |\sum_i \xi_i|^3 e^{3\overline{W}/\theta}\right)^{1/3} = \left(\mathbb{E} e^{3\overline{W}/\theta} \mathbb{E} |\sum_i \hat{\xi}_i|^3\right)^{1/3} \quad (4.46) \\ &\leq \text{PU}_{\text{exp}}\left(\frac{3}{\theta}, w, \frac{1}{\sqrt{3}}, \varepsilon_1\right) \|\sum_i \hat{\xi}_i\|_3. \end{aligned}$$

Thus, recalling the case condition $\delta \leq \delta_0$, we have

$$\begin{aligned} \mathbb{P}(z - |\Delta| \leq \overline{W} \leq z, |\Delta| \leq \pi_1 z) &= \frac{1}{c_z} e^{-(1-\pi_1)z/\theta} e^{\delta/\theta} G_{1,1} \\ &\leq \frac{1}{c_z} e^{-(1-\pi_1)z/\theta} e^{\delta/\theta} (G_{1,2} + |G_2| + \mathbb{E} W f_{\overline{\Delta}}(\overline{W})) \\ &\leq (c_1 \sum_i \|\xi_i\|_3 \|\Delta - \Delta_i\|_{3/2} + c_2 \|\Delta\|_{3/2} + c_3 \delta) e^{-(1-\pi_1)z/\theta}, \end{aligned}$$

where the equality comes from the definition (4.39) of $G_{1,1}$, the first inequality follows from (4.37) and (4.38), and the second inequality follows from (4.40), (4.42), (4.43), (4.46), and (4.45), along with the definitions (4.14), (4.15), and (4.16) of c_1 , c_2 , and c_3 . Thus, in view of the definition (4.13) of $\tilde{\tau}$, the inequality (4.36) is proved for the other case, $\delta \leq \delta_0$.

Replace now \mathbb{P} with $\hat{\mathbb{P}}$ in (4.33), so that (4.35) and (4.36) imply

$$\hat{\mathbb{P}}(W \leq z) - \hat{\mathbb{P}}(\tilde{T} \leq z) \leq \tilde{\gamma}_z + \tilde{\tau} e^{-(1-\pi_1)z/\theta}.$$

In a similar fashion, one bounds $\hat{\mathbb{P}}(\tilde{T} \leq z) - \hat{\mathbb{P}}(W \leq z)$ from above, establishing (4.10).

Appendix A: An explicit nonuniform bound in the multivariate delta method

The bound PU_{exp} on the exponential moment of a sum of independent random variables found in (4.5) can be used in conjunction with the Markov inequality to obtain a bound, PU_{tail} , on certain tail probabilities of the sum; the following lemma quotes expressions found in [69, 54] for this bound. These expressions will be needed in applications of Theorem A.2.

Lemma A.1. *For any real $x, y > 0, B > 0, m$, and $\varepsilon \in (0, 1]$, let*

$$u := \frac{(x - m)_+ y}{B^2} \quad \text{and} \quad \kappa := \frac{(x - m)_+}{y}.$$

Further, let $S = \sum_{i=1}^n \zeta_i$, where ζ_1, \dots, ζ_n are independent real-valued r.v.'s satisfying (4.4). Then

$$\mathbb{P}(S > x) \leq \text{PU}_{\text{tail}}(u, \kappa, \varepsilon) := \inf_{\lambda \geq 0} e^{-\lambda(x-m)+} \text{PU}_{\text{exp}}(\lambda, y, B, \varepsilon), \tag{A.1}$$

with

$$\text{PU}_{\text{tail}}(0, \kappa, \varepsilon) = 1,$$

$$\text{PU}_{\text{tail}}(u, \kappa, 1) = \text{BH}_{\text{tail}}(u, \kappa)$$

whenever $u > 0$ (where BH_{tail} is as defined in (4.7)), and

$$\begin{aligned} \text{PU}_{\text{tail}}(u, \kappa, \varepsilon) = \\ \exp \left[\frac{\kappa}{2(1-\varepsilon)u} \left((1-\varepsilon)^2 \left[1 + \text{W} \left(\frac{\varepsilon}{1-\varepsilon} \exp \frac{\varepsilon+u}{1-\varepsilon} \right) \right]^2 - (\varepsilon+u)^2 - (1-\varepsilon^2) \right) \right] \end{aligned} \tag{A.2}$$

whenever $u > 0$ and $\varepsilon < 1$; in (A.2), W is Lambert's product-log function with domain restricted to the positive real numbers (so that for positive w and z one has $\text{W}(z) = w$ if and only if $z = we^w$).

One also has the alternative identity

$$\text{PU}_{\text{tail}}(u, \kappa, \varepsilon) = \inf_{0 < \alpha < 1} \exp[L_1 \vee L_2], \tag{A.3}$$

where

$$L_1 := L_1(\alpha, u, \kappa, \varepsilon) := \kappa \left(1 - \alpha - \alpha \frac{\varepsilon}{1-\varepsilon} - \frac{\alpha(2-\alpha)}{2(1-\varepsilon)} u \right), \tag{A.4}$$

$$\text{with } L_1(\alpha, u, \kappa, 1) := -\infty,$$

and

$$L_2 := L_2(\alpha, u, \kappa, \varepsilon) := \kappa \left(1 - \alpha - \left(1 - \frac{\alpha}{2} + \frac{\varepsilon}{u} \right) \ln \left(1 + (1-\alpha) \frac{u}{\varepsilon} \right) \right), \tag{A.5}$$

$$\text{with } L_2(\alpha, 0, \kappa, \varepsilon) := 0.$$

Moreover, if the ζ_i 's take values in a separable Banach space with norm $\|\cdot\|$, then

$$(A.1) \text{ holds under (4.4) when } S \text{ and } \zeta_i \text{ are replaced by } \|S\| \text{ and } \|\zeta_i\|, \tag{A.6}$$

respectively.

Indeed, (A.1) is essentially [54, Proposition 3.1], with the ‘‘boundary’’ case $\varepsilon = 1$ resulting in the Bennett–Hoeffding bound $\text{BH}_{\text{tail}}(u, \kappa)$. Next, (A.3) (for $\varepsilon < 1$) is established in [69, Corollary 1] and, again, immediately follows for $\varepsilon = 1$ using $\text{BH}_{\text{tail}}(u, \kappa)$. The verity of (A.6) is implied by the remark in [69, p. 343] (see also [68, Theorem 1]).

Theorem A.2. *Let \mathfrak{X} be a Hilbert space, let f satisfy (2.1) for some real $\epsilon > 0$, and assume that V, V_1, \dots, V_n are i.i.d. \mathfrak{X} -valued random vectors such that $\mathbb{E}V = 0$, $\tilde{\sigma} = \|L(V)\|_2 > 0$, and $v_3 = \|V\|_3 < \infty$. Further, take any positive real numbers*

$$c_*, \theta, w, \delta_0, \pi_1, \pi_2, \pi_3, z_0, \omega, \kappa_{2,0}, \kappa_{3,0}, \kappa_{2,1}, \kappa_{3,1}, \kappa_2, \kappa_3, \alpha, \varepsilon_*, K_1, K_2, \text{ and } K_3 \quad (\text{A.7})$$

satisfying the constraints

$$c_* < 1, \delta_0 \leq w, \pi_1 + \pi_2 + \pi_3 = 1, \omega \leq \frac{M_\epsilon \epsilon^2}{2\pi_1}, \kappa_3 \geq \frac{3}{2}, \alpha < 1, \varepsilon_* < 1, \hat{\kappa}_2 \geq 2, \hat{\gamma} < 1, \quad (\text{A.8})$$

where

$$\hat{\gamma} := \left(\frac{M_\epsilon^2 \omega}{4\pi_1^2 K_2} \right)^{1/4} + \frac{\kappa_2^2}{K_3} \left(\frac{M_\epsilon \omega}{2\pi_1} \right)^{3/2} \quad (\text{A.9})$$

and

$$\hat{\kappa}_2 := (1 - \hat{\gamma})\kappa_2. \quad (\text{A.10})$$

Also introduce

$$t_2 := \frac{\pi_1 \alpha (2 - \alpha) (1 - \hat{\gamma})^2}{M_\epsilon (1 - \varepsilon_*)} \left(\frac{K_2}{\omega} \right)^{1/2}, \quad t_3 := \frac{\kappa_2^2}{(1 - \hat{\gamma}) K_3} \left(\frac{M_\epsilon \omega}{2\pi_1} \right)^{3/2}, \quad (\text{A.11})$$

$$u_0 := \frac{2\pi_1 (1 - \hat{\gamma})}{M_\epsilon \kappa_2} \left(\frac{K_2}{\omega} \right)^{1/2}, \quad (\text{A.12})$$

$$\tilde{\varepsilon}_1 := \frac{1}{K_1 w}, \quad \text{and} \quad \tilde{a}_1 := 1/L_W(3w/\theta, \tilde{\varepsilon}_1), \quad (\text{A.13})$$

where $L_W(c, B)$ is as in (4.8); further let \tilde{c}_1, \tilde{c}_2 , and \tilde{c}_3 be obtained from c_1, c_2 , and c_3 in (4.14)–(4.16) by replacing there a_1, ε_1 , and σ_p by $\tilde{a}_1, 1 \wedge \tilde{\varepsilon}_1$, and $K_1^{-1/3}$, respectively. Recall also the definition of PU_{tail} in (A.1). Then for all $z \in \mathbb{R}$ and $n \in \mathbb{N}$ such that

$$z_0 \leq z \leq \frac{\omega}{\tilde{\sigma}} \sqrt{n}, \quad (\text{A.14})$$

$$\frac{K_1 \varsigma_3^3}{\sqrt{n}} \leq 1, \quad \frac{K_2 v_2^4}{\tilde{\sigma}^3 z^3 \sqrt{n}} \leq 1, \quad \text{and} \quad \frac{K_3 v_3^3}{\tilde{\sigma}^3 z^3 \sqrt{n}} \leq 1 \quad (\text{A.15})$$

one has

$$\begin{aligned} & \left| \mathbb{P} \left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \\ & \leq \frac{\mathfrak{K}_{n1} \varsigma_3^3 + ((\mathfrak{K}_{n21} \vee \mathfrak{K}_{n22}) v_2^4) \vee (\mathfrak{K}_{n31} v_3^3) + \mathfrak{K}_{n32} v_3^3}{z^3 \sqrt{n}} \\ & \quad + \frac{\mathfrak{K}_{e0} + \mathfrak{K}_{e1} \varsigma_3^3 + \mathfrak{K}_{e2} v_2^3 + \mathfrak{K}_{e3} v_3^3}{e^{(1-\pi_1)z/\theta} \sqrt{n}}, \quad (\text{A.16}) \end{aligned}$$

where

$$\begin{aligned} \mathfrak{R}_{n1} &:= 30.2211 + \frac{1}{\pi_2^3} \\ &+ \frac{\kappa_3^{3/2}}{(w\pi_3)^3} \left(\frac{\kappa_3^{3/2}}{K_1} + \sup_{u \geq \pi_3^2 z_0^2 / \kappa_3} u^{3/2} \text{PU}_{\text{tail}} \left(u, \kappa_3, \frac{\kappa_3}{K_1 \pi_3 z_0} \wedge 1 \right) \right), \end{aligned} \quad (\text{A.17})$$

$$\mathfrak{R}_{n21} := \frac{\omega \exp\{\hat{\kappa}_2(1 - \alpha - \frac{\alpha \varepsilon_*}{1 - \varepsilon_*})\}}{\tilde{\sigma}^3} \left(\frac{M_\varepsilon(1 - \varepsilon_*)}{\pi_1 \alpha (2 - \alpha)(1 - \hat{\gamma})^2} \right)^2 \sup_{t \geq t_2} t^2 e^{-t}, \quad (\text{A.18})$$

$$\mathfrak{R}_{n22} := \frac{\omega}{\tilde{\sigma}^3} \left(\frac{M_\varepsilon \kappa_2}{2\pi_1(1 - \hat{\gamma})} \right)^2 \sup_{u \geq u_0} u^2 \text{PU}_{\text{tail}}(u, \hat{\kappa}_2, \varepsilon_*), \quad (\text{A.19})$$

$$\mathfrak{R}_{n31} := \frac{\kappa_2^2 e^{\hat{\kappa}_2(1 - \alpha)}}{\tilde{\sigma}^3(1 - \hat{\gamma})} \left(\frac{M_\varepsilon \omega}{2\pi_1} \right)^{3/2} \sup_{t \in (0, t_3]} \frac{1}{t} \exp \left[-\hat{\kappa}_2 \left(1 - \frac{\alpha}{2} + t \right) \ln \left(1 + \frac{1 - \alpha}{t} \right) \right], \quad (\text{A.20})$$

$$\mathfrak{R}_{n32} := \left(\frac{\kappa_2}{\tilde{\sigma}} \right)^3 \left(\frac{M_\varepsilon \omega}{2\pi_1} \right)^{3/2}, \quad (\text{A.21})$$

$$\mathfrak{R}_{e0} := \frac{M_\varepsilon \tilde{c}_2}{6\tilde{\sigma}} \left(\frac{1}{\kappa_{3,0}^3 K_1^{2/3}} + \frac{2^{2/3}}{\kappa_{2,0}^3} \right) - \frac{(2c_* - 1)_+^2}{4(1 - c_*)} \tilde{c}_3, \quad (\text{A.22})$$

$$\mathfrak{R}_{e1} := d(c_*) \tilde{c}_3 + \frac{M_\varepsilon \tilde{c}_1}{6\tilde{\sigma}} \left(\frac{1}{\kappa_{3,1}^3 K_1} + \frac{2}{\kappa_{2,1}^3} \right), \quad (\text{A.23})$$

$$\mathfrak{R}_{e2} := \frac{M_\varepsilon}{3\tilde{\sigma}} (2\tilde{c}_1 \kappa_{2,1}^{3/2} + 2^{2/3} \tilde{c}_2 \kappa_{2,0}^{3/2}), \quad (\text{A.24})$$

$$\mathfrak{R}_{e3} := \frac{M_\varepsilon}{3\tilde{\sigma}} \left(\frac{\tilde{c}_1 \kappa_{3,1}^{3/2}}{K_1} + \frac{\tilde{c}_2 \kappa_{3,0}^{3/2}}{K_1^{2/3}} \right), \quad (\text{A.25})$$

$$d: (0, 1) \rightarrow \mathbb{R} \text{ is defined by } d(c_*) = \begin{cases} c_* & \text{if } c_* \in (0, \frac{1}{2}], \\ \frac{1}{4(1 - c_*)} & \text{if } c_* \in (\frac{1}{2}, 1); \end{cases} \quad (\text{A.26})$$

moreover, each of the expressions in (A.17)–(A.25) is finite.

Remark A.3. Suppose here that $L(V)$ is symmetric. Then the statement of Theorem A.2 holds when the replacement mentioned in Remark 2.8 is made in the expression (4.15) for c_2 and, accordingly, in the expression for \tilde{c}_2 defined right after (A.13). Also, one can take \mathfrak{R}_{n1} in (A.16) to be defined as

$$\begin{aligned} \mathfrak{R}_{n1} &:= 30.2211 + \frac{1}{2\pi_2^3} \\ &+ \frac{\kappa_3^{3/2}}{2(w\pi_3)^3} \left(\frac{\kappa_3^{3/2}}{2K_1} + \sup_{u \geq \pi_3^2 z_0^2 / \kappa_3} u^{3/2} \text{PU}_{\text{tail}} \left(u, \kappa_3, \frac{\kappa_3}{K_1 \pi_3 z_0} \wedge 1 \right) \right), \end{aligned}$$

because one can then use $\mathbb{P}(\xi_1 > t) \leq \varsigma_3^3 / (2t^3 \sqrt{n})$ in place of $\mathbb{P}(\xi_1 > t) \leq \varsigma_3^3 / (t^3 \sqrt{n})$ to improve bounds on the terms in γ_z .

Remark A.4. That all the expressions denoted by \mathfrak{K} with indices and defined by formulas (A.17)–(A.25) are finite is easily verifiable by inspection, except perhaps for \mathfrak{K}_{n1} , \mathfrak{K}_{n22} , and \mathfrak{K}_{n31} , whose definitions in (A.17), (A.19), and (A.20) involve comparatively complicated suprema. However, as shown in [65], conditions $\kappa_3 \geq \frac{3}{2}$ and $\hat{\kappa}_2 \geq 2$ in (A.8) suffice for these three suprema, and hence for \mathfrak{K}_{n1} , \mathfrak{K}_{n22} , and \mathfrak{K}_{n31} , to be finite.

Proof of Theorem A.2. Take any $z \in \mathbb{R}$ and $n \in \mathbb{N}$ such that (A.14) and (A.15) hold. The conditions of Theorem 2.5 are met, with $X_i = V_i/n$, so that (2.19) and (2.16) imply

$$\left| \mathbb{P}\left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{30.2211\varsigma_3^3}{z^3\sqrt{n}} + \gamma_z + \tau e^{-(1-\pi_1)z/\theta}. \tag{A.27}$$

We shall first demonstrate that

$$\tau \leq \frac{1}{\sqrt{n}} \left(\mathfrak{K}_{e0} + \mathfrak{K}_{e1}\varsigma_3^3 + \mathfrak{K}_{e2}v_2^3 + \mathfrak{K}_{e3}v_3^3 \right), \tag{A.28}$$

where $\mathfrak{K}_{e0}, \dots, \mathfrak{K}_{e3}$ are as in (A.22)–(A.25). By the first inequality of (A.15) (recall also (2.21), (4.17), and (A.13))

$$\sigma_3 = \frac{\varsigma_3}{n^{1/6}} \leq K_1^{-1/3}, \quad \varepsilon_1 \leq \tilde{\varepsilon}_1, \quad \text{and} \quad \max_i \frac{\|\xi_i\|_2}{w} = \frac{1}{w\sqrt{n}} \leq \frac{\varsigma_3^3}{w\sqrt{n}} \leq \frac{1}{K_1 w}.$$

Then, recalling that $\text{PU}_{\text{exp}}(\lambda, y, B, \varepsilon)$ and $L_W(c, B)$ are nondecreasing with respect to ε and B , respectively, we see that $a_1 \leq \tilde{a}_1$ and $c_j \leq \tilde{c}_j$ for $j = 1, 2, 3$. By Remark 2.3 and (4.24), we see that (2.9) is satisfied when

$$\delta = \frac{d(c_*)\varsigma_3^3 - (2c_* - 1)_+^2/(4(1 - c_*))}{\sqrt{n}},$$

where d is as in (A.26). Then (2.18) and (4.25) imply

$$\tau \leq \frac{1}{\sqrt{n}} \left(\frac{M_\varepsilon}{2\tilde{\sigma}} \left(\tilde{c}_1\varsigma_3 \left(\frac{v_3^2}{K_1} + 2v_2^2 \right) + \tilde{c}_2 \left(\frac{v_3^2}{K_1^{2/3}} + 2^{2/3}v_2^2 \right) \right) + \tilde{c}_3 \left(d(c_*)\varsigma_3^3 - \frac{(2c_* - 1)_+^2}{4(1 - c_*)} \right) \right); \tag{A.29}$$

the inequalities $v_3^2/\sqrt{n} \leq v_3^2/K_1$ and $v_3^2/n^{1/3} \leq v_3^2/K_1^{2/3}$ were used above, with these following from $n \geq K_1^2\varsigma_3^3 \geq K_1^2$ (which follows from (A.15)). Finally, use Young’s inequality to see that

$$\varsigma_3^i v_\alpha^2 \leq \frac{1}{3} \frac{\varsigma_3^{3i}}{\kappa_{\alpha,i}^3} + \frac{2}{3} \kappa_{\alpha,i}^{3/2} v_\alpha^3 \text{ for } (\alpha, i) \in \{2, 3\} \times \{0, 1\}; \tag{A.30}$$

applying (A.30) to the various terms in the bound of (A.29) yields the desired inequality (4.26).

Consider next the problem of bounding the terms in γ_z . First,

$$\sum_i \mathbb{P}(\xi_i > \pi_2 z) \leq \frac{1}{\pi_2^3} \frac{\zeta_3^3}{z^3 \sqrt{n}}. \tag{A.31}$$

Next, use truncation and Lemma A.1 (with $x = \pi_3 z$, $y = x_3/\kappa_3$, $B = 1$, $m = 0$, and $\varepsilon = \sigma_3^3/(B^2 y)$) to obtain

$$\begin{aligned} &\mathbb{P}(W - \xi_i \geq \pi_3 z) \\ &\leq \mathbb{P}\left(\max_i \xi_i > \frac{\pi_3 z}{\kappa_3}\right) + \mathbb{P}\left(\sum_{j \neq i} \xi_j \mathbf{I}\left\{\xi_j \leq \frac{\pi_3 z}{\kappa_3}\right\} > \pi_3 z\right) \\ &\leq n \mathbb{P}\left(\xi_1 > \frac{\pi_3 z}{\kappa_3}\right) + \text{PU}_{\text{tail}}\left(\frac{\pi_3^2 z^2}{\kappa_3}, \kappa_3, \frac{\sigma_3^3}{\pi_3 z / \kappa_3} \wedge 1\right) \\ &\leq \frac{1}{z^3} \left(\frac{\kappa_3^3}{K_1 \pi_3^3} + \frac{\kappa_3^{3/2}}{\pi_3^3} \sup_{u \geq \pi_3^2 z_0^2 / \kappa_3} u^{3/2} \text{PU}_{\text{tail}}\left(u, \kappa_3, \frac{\kappa_3}{K_1 \pi_3 z_0} \wedge 1\right)\right); \end{aligned}$$

note we have again used the first inequality of (A.15), the fact that PU_{tail} is nondecreasing in ε (cf. (4.6)), and the assumption that $z > z_0$ (cf. (A.14)). Then, since $\sum_i \mathbb{P}(\xi_i > w) \leq \zeta_3^3/(w^3 \sqrt{n})$,

$$\begin{aligned} \frac{30.2211 \zeta_3^3}{z^3 \sqrt{n}} + \gamma_z &\leq \frac{\mathfrak{K}_{n1} \zeta_3^3}{z^3 \sqrt{n}} + \frac{v_3^3}{n^2 y_2^3} \cdot \left(\frac{\omega \sqrt{n}}{\tilde{\sigma} z}\right)^{3/2} + \mathbb{P}(\|\bar{V}_{y_2}\| > x_2) \\ &= \frac{\mathfrak{K}_{n1} \zeta_3^3 + \mathfrak{K}_{n32} v_3^3}{z^3 \sqrt{n}} + \mathbb{P}(\|\bar{V}_{y_2}\| > x_2), \end{aligned} \tag{A.32}$$

with the above inequality following from (4.29) and (A.14) (recall also the definition (A.17) of \mathfrak{K}_{n1}), and the equality following from the definitions (4.28) and (A.21) of y_2 and \mathfrak{K}_{n32} , respectively.

Considering (A.27), (A.28), and (A.32), the proof will be complete once we show that

$$\mathbb{P}(\|\bar{V}_{y_2}\| > x_2) \leq \frac{((\mathfrak{K}_{n21} \vee \mathfrak{K}_{n22}) v_2^4) \vee (\mathfrak{K}_{n31} v_3^3)}{z^3 \sqrt{n}}. \tag{A.33}$$

Recalling the definitions in (4.28), we have

$$\begin{aligned} \mathbb{E}\|\bar{V}_{y_2}\| &\leq \mathbb{E}\|\bar{V}\| + \mathbb{E}\|\bar{V}_{y_2} - \bar{V}\| \\ &\leq \|\bar{V}\|_2 + \frac{1}{n} \sum_i \mathbb{E}\|V_i \mathbf{I}\{\|V_i\| > n y_2\}\| \\ &\leq \frac{v_2}{\sqrt{n}} + \frac{v_3^3}{n^2 y_2^3} \\ &= x_2 \left(\frac{v_2}{c_x (\tilde{\sigma} z)^{1/2} n^{1/4}} + \frac{\kappa_2^2 v_3^3}{c_x^3 (\tilde{\sigma} z)^{3/2} n^{5/4}}\right) \\ &= x_2 \left(\frac{1}{c_x} \left(\frac{v_2^4}{\tilde{\sigma}^3 z^3 \sqrt{n}}\right)^{1/4} \left(\frac{\tilde{\sigma} z}{\sqrt{n}}\right)^{1/4} + \frac{\kappa_2^2}{c_x^3} \frac{v_3^3}{\tilde{\sigma}^3 z^3 \sqrt{n}} \left(\frac{\tilde{\sigma} z}{\sqrt{n}}\right)^{3/2}\right) \\ &\leq x_2 \left(\frac{1}{c_x} \left(\frac{\omega}{K_2}\right)^{1/4} + \frac{\kappa_2^2 \omega^{3/2}}{c_x^3 K_3}\right) = \hat{\gamma} x_2, \end{aligned}$$

where (A.14) and (A.15) are used to obtain the last inequality above, and the definition (A.9) of $\hat{\gamma}$ is used for the last equality. Then, since $\hat{\gamma} < 1$ is assumed in (A.8), invoke Lemma A.1 (with $x = x_2$, $y = y_2$, $B = v_2/\sqrt{n}$, $m = \hat{\gamma}x_2$, and $\varepsilon = v_3^3/(n^2 B^2 y)$) to see that

$$\mathbb{P}(\|\bar{V}_{y_2}\| > x_2) \leq \text{PU}_{\text{tail}}(\hat{u}, \hat{\kappa}_2, \varepsilon_2) \quad (\text{A.34})$$

where

$$\hat{u} := \frac{c_x^2(1-\hat{\gamma})}{\kappa_2} \frac{\tilde{\sigma}z}{v_2^2} \sqrt{n}, \quad \hat{\kappa}_2 := (1-\hat{\gamma})\kappa_2, \quad \text{and} \quad \varepsilon_2 := \frac{\kappa_2 v_3^3}{c_x v_2^2 (\tilde{\sigma}z)^{1/2} n^{3/4}} \wedge 1. \quad (\text{A.35})$$

The inequality in (A.33) is proved by taking any $\varepsilon_* \in (0, 1)$ and then considering two cases: (i) $\varepsilon_2 \in (\varepsilon_*, 1]$ and (ii) $\varepsilon_2 \in (0, \varepsilon_*]$. Assume first that $\varepsilon_2 \in (\varepsilon_*, 1]$. By (A.35) and (A.3),

$$\text{PU}_{\text{tail}}(\hat{u}, \hat{\kappa}_2, \varepsilon_2) \leq \exp\{L_1(\alpha, \hat{u}, \hat{\kappa}_2, \varepsilon_2)\} \vee \exp\{L_2(\alpha, \hat{u}, \hat{\kappa}_2, \varepsilon_2)\} \quad (\text{A.36})$$

for any $\alpha \in (0, 1)$. Now introduce

$$r_2^2 := \frac{1}{\hat{u}} = \frac{\kappa_2}{c_x^2(1-\hat{\gamma})} \left(\frac{v_2^4}{\tilde{\sigma}^3 z^3 \sqrt{n}} \right)^{1/2} \left(\frac{\tilde{\sigma}z}{\sqrt{n}} \right)^{1/2} \leq \frac{\kappa_2 \omega^{1/2}}{c_x^2(1-\hat{\gamma})} \left(\frac{v_2^4}{\tilde{\sigma}^3 z^3 \sqrt{n}} \right)^{1/2} \quad (\text{A.37})$$

$$\leq \frac{\kappa_2 \omega^{1/2}}{K_2^{1/2} c_x^2(1-\hat{\gamma})} = \frac{1}{u_0} \quad (\text{A.38})$$

and

$$r_3^3 := \frac{\varepsilon_2}{\hat{u}} \leq \frac{\kappa_2^2}{c_x^3(1-\hat{\gamma})} \frac{v_3^3}{\tilde{\sigma}^3 z^3 \sqrt{n}} \left(\frac{\tilde{\sigma}z}{\sqrt{n}} \right)^{3/2} \leq \frac{\kappa_2^2 \omega^{3/2}}{c_x^3(1-\hat{\gamma})} \frac{v_3^3}{\tilde{\sigma}^3 z^3 \sqrt{n}} \quad (\text{A.39})$$

$$\leq \frac{\kappa_2^2 \omega^{3/2}}{c_x^3(1-\hat{\gamma}) K_3} = t_3, \quad (\text{A.40})$$

where (A.14) is used to establish the inequalities in (A.37) and (A.39), and (A.15) and (A.11) are used for (A.38) and (A.40).

Next, in view of (A.38), (A.10), and (A.11), one has

$$\begin{aligned} \frac{\hat{\kappa}_2 \alpha (2-\alpha)}{2(1-\varepsilon_2)} \hat{u} &\geq \frac{\hat{\kappa}_2 \alpha (2-\alpha)}{2(1-\varepsilon_*)} \frac{c_x^2(1-\hat{\gamma})}{\kappa_2} \left(\frac{K_2}{\omega} \right)^{1/2} \\ &= \frac{\pi_1 \alpha (2-\alpha) (1-\hat{\gamma})^2}{M_\varepsilon (1-\varepsilon_*)} \left(\frac{K_2}{\omega} \right)^{1/2} = t_2. \end{aligned}$$

So, the case condition $\varepsilon_2 \in (\varepsilon_*, 1]$ together with the definitions of (A.4) and (A.37) of L_1 and r_2^2 imply

$$e^{L_1} \leq e^{\hat{\kappa}_2(1-\alpha-\alpha\varepsilon_*/(1-\varepsilon_*))} \left(\frac{2(1-\varepsilon_*)}{\hat{\kappa}_2 \alpha (2-\alpha)} \right)^2 \left(\sup_{t \geq t_2} t^2 e^{-t} \right) r_2^4 \leq \mathfrak{K}_{n21} \frac{v_2^4}{z^3 \sqrt{n}}, \quad (\text{A.41})$$

where the last inequality follows by the definition (A.18) of \mathfrak{K}_{n21} and (A.37) (on recalling also that $\hat{\kappa}_2 = (1 - \hat{\gamma})\kappa_2$). Note that if $\varepsilon_2 = 1$ then, by the definition, $L_1 = -\infty$, which makes (A.41) trivial (using the convention $\exp\{-\infty\} := 0$).

Again by the case condition $\varepsilon_2 \in (\varepsilon_*, 1]$, now together with (A.5) and (A.40),

$$e^{L_2} \leq e^{\hat{\kappa}_2(1-\alpha)} \left(\sup_{t \in (0, t_3]} \frac{1}{t} \exp \left[-\hat{\kappa}_2 \left(1 - \frac{\alpha}{2} + t \right) \ln \left(1 + \frac{1-\alpha}{t} \right) \right] \right) r_3^3 \leq \mathfrak{K}_{n31} \frac{v_3^3}{z^3 \sqrt{n}}, \tag{A.42}$$

where the last inequality follows by the definition (A.20) of \mathfrak{K}_{n31} and (A.39). Now, upon combining (A.36), (A.41), and (A.42), we obtain the result (A.33) in the case $\varepsilon_2 \in (\varepsilon_*, 1]$.

Consider the remaining case, when $\varepsilon_2 \in (0, \varepsilon_*]$. Then, by (4.6), (A.35), (A.37), (A.38), and the definition (A.19) of \mathfrak{K}_{n22} ,

$$\begin{aligned} \text{PU}_2 &\leq \text{PU}_{\text{tail}}(\hat{u}, \hat{\kappa}_2, \varepsilon_2) \leq \text{PU}_{\text{tail}}(\hat{u}, \hat{\kappa}_2, \varepsilon_*) \\ &\leq r_2^4 \left(\sup_{u \geq u_0} u^2 \text{PU}_{\text{tail}}(u, \hat{\kappa}_2, \varepsilon_*) \right) \leq \mathfrak{K}_{n22} \frac{v_2^4}{z^3 \sqrt{n}}. \end{aligned} \tag{A.43}$$

Thus, (A.43) yields (A.33) in the case $\varepsilon_2 \in (0, \varepsilon_*]$ as well. As was noted, the demonstration of (A.33) completes the proof. \square

Appendix B: Optimality of the restriction $z = O(\sqrt{n})$ for the nonuniform bound

Proposition B.1. *Let $\mathfrak{X} = \mathbb{R}$ and $f(x) \equiv x + x^2$, so that (2.1) is satisfied when $L(x) \equiv x$, $M_\epsilon = 2$, and $\epsilon = 1$. Let V, V_1, \dots, V_n 's be real-valued symmetric i.i.d. r.v.'s with density $|v|^{-4} \ln^{-2} |v|$ for all $|v| \geq v_0$, where the real number $v_0 > 1$ and the density values on $(-v_0, v_0)$ are chosen so that $\|V\|_2 = 1$; note that then $\|V\|_3 < \infty$. For any triple $b := (b_1, b_2, b_3)$ of positive real numbers, let $\mathbf{NZ}(b)$ denote the set of all pairs $(n, z) \in \mathbb{N} \times (0, \infty)$ for which the inequality (2.25) with b_1, b_2, b_3 in place of the three instances of \mathfrak{C} holds. Then there exists a constant $\omega(b) \in (0, \infty)$ depending only on b such that (2.24) holds for all pairs $(n, z) \in \mathbf{NZ}$.*

Remark B.2. Let $r \in (0, 3)$. Then an application of Chebyshev's inequality to the first two terms in the bound of (2.25) yields

$$\begin{aligned} &\left| \mathbb{P} \left(\frac{f(\bar{V})}{\bar{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \\ &\leq \mathfrak{C} \left(\frac{\mathbb{E} \|V\|^r \mathbf{I}\{\|V\| > \mathfrak{C}z\sqrt{n}\}}{z^r n^{r/2-1}} + \frac{\mathbb{E} \|V\|^r \mathbf{I}\{\|V\| > \mathfrak{C}\sqrt{n}\}}{z^3 n^{r/2-1}} + \frac{1}{z^3 \sqrt{n}} + \frac{1}{e^z \sqrt{n}} \right) \end{aligned} \tag{B.1}$$

for any z satisfying (2.24). The arguments of the proof of Proposition B.1 can be used to demonstrate that the bound of (B.1) (larger than that in (2.25))

generally fails to hold if $z/\sqrt{n} \rightarrow \infty$. Using Chebyshev’s inequality when $r = 3$ yields

$$\left| \mathbb{P}\left(\frac{f(\bar{V})}{\bar{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}. \tag{B.2}$$

One might hope that a bound of the form in (B.2) could hold for all f satisfying the smoothness condition (2.1) and for all $z > 0$. However, another modification of the proof of Proposition B.1 demonstrates that (B.2) fails to be true whenever

$$\frac{z}{\sqrt{n} \ln^\alpha n} \rightarrow \infty \quad \text{for any fixed } \alpha > \frac{1}{3}; \tag{B.3}$$

the extra log factor above is needed because the bound in (B.2) is worse than that in (B.1).

Proof of Proposition B.1. Let $S = \bar{V}$, so that $\sigma = \|L(S)\|_2 = 1/\sqrt{n}$, $T = f(S)/\sigma = \sqrt{n}(S + S^2)$, and $W = L(S)/\sigma = \sqrt{n}S$. To obtain a contradiction, assume that Proposition B.1 is false. Then for some triple $b \in (0, \infty)^3$ and each value of $\omega \in \mathbb{N}$ there is a pair $(n, z) = (n_\omega, z_\omega) \in \mathbf{NZ}(b)$ such that $z > \omega\sqrt{n}$. Now, for the rest of the proof of Proposition B.1, let $\omega \rightarrow \infty$, so that

$$\zeta := z/\sqrt{n} \rightarrow \infty;$$

further let

$$\vartheta := \zeta^{1/2}n = z^{1/2}n^{3/4},$$

so that $\vartheta/n = \zeta^{1/2} \rightarrow \infty$. Note that for $v > v_0$

$$\mathbb{P}(V > v) = \int_v^\infty \frac{du}{u^4 \ln^2 u} \asymp \frac{1}{v^3 \ln^2 v}$$

as $v \rightarrow \infty$, which follows by l’Hospital’s rule.

So,

$$\begin{aligned} n \mathbb{P}(\|V\| > \mathfrak{C}z\sqrt{n}) &\asymp \frac{n}{z^3 n^{3/2} \ln^2(z\sqrt{n})} = \frac{\ln^2(\zeta^{1/2}n)}{\zeta^{3/2} \ln^2(\zeta n)} \frac{n}{\vartheta^3 \ln^2 \vartheta} \\ &= o(n \mathbb{P}(V > \vartheta)), \\ \frac{n \mathbb{P}(\|V\| > \mathfrak{C}\sqrt{n})}{z^3} &\asymp \frac{n}{z^3 n^{3/2} \ln^2 \sqrt{n}} = \frac{\ln^2(\zeta^{1/2}n)}{\zeta^{3/2} \ln^2 \sqrt{n}} \frac{n}{\vartheta^3 \ln^2 \vartheta} = o(n \mathbb{P}(V > \vartheta)), \\ \frac{1}{(z\sqrt{n})^3} &= \frac{\ln^2(\zeta^{1/2}n)}{\zeta^{3/2} n} \frac{n}{\vartheta^3 \ln^2 \vartheta} = o(n \mathbb{P}(V > \vartheta)), \\ \frac{1}{e^z \sqrt{n}} &= \frac{\zeta^{3/2} n^{3/2} \ln^2(\zeta^{1/2}n)}{e^{\zeta \sqrt{n}}} \frac{n}{\vartheta^3 \ln^2 \vartheta} = o(n \mathbb{P}(V > \vartheta)), \end{aligned}$$

and

$$1 - \Phi(z) \asymp \frac{1}{ze^{z^2/2}} = \frac{\zeta^{1/2} n^{3/2} \ln^2(\zeta^{1/2}n)}{e^{\zeta^2 n/2}} \frac{n}{\vartheta^3 \ln^2 \vartheta} = o(n \mathbb{P}(V > \vartheta)). \tag{B.4}$$

Then (2.25) and (4.32) imply that $|\mathbb{P}(T \leq z) - \Phi(z)|$ and $|\mathbb{P}(W \leq z) - \Phi(z)|$ are both $o(n\mathbb{P}(V > \vartheta))$. Now let $\Delta = T - W = \sqrt{n}S^2$, so that

$$\mathbb{P}(\Delta > 2z) \leq \mathbb{P}(T > z) + \mathbb{P}(-W > z) = \mathbb{P}(T > z) + \mathbb{P}(W > z) = o(n\mathbb{P}(V > \vartheta)), \tag{B.5}$$

by (B.4).

On the other hand, by [15, Lemma 2.3],

$$\mathbb{P}(\Delta > 2z) = \mathbb{P}(\sqrt{n}S^2 > 2z) = \mathbb{P}(|\sum_i V_i| > \sqrt{2}\vartheta) \geq \frac{1}{2} (1 - e^{-\psi})$$

for large enough n , where

$$\psi := n\mathbb{P}(|V| > \sqrt{2}\vartheta) = 2n\mathbb{P}(V > \sqrt{2}\vartheta).$$

Since $\vartheta/n = \zeta^{1/2} \rightarrow \infty$, one has $\psi = o(n^{-2}) \rightarrow 0$, whence

$$\mathbb{P}(\Delta > 2z) \geq \frac{\psi}{3} > \frac{2}{3 \cdot 2^3} n\mathbb{P}(V > \vartheta)$$

for large enough n , which contradicts (B.5).

The statements of Remark B.2 are proved with only a few modifications to the above arguments, using the relation

$$\mathbb{E}\|V\|^r \mathbb{I}\{\|V\| > v\} \asymp \frac{1}{v^{3-r} \ln^2 v}$$

as $v \rightarrow \infty$, for any $r \in (0, 3)$. In order to show that (B.2) fails to hold simultaneously with (B.3), let V have density $1/(|v|^4 \ln^{3\alpha} |v|)$ for $|v| \geq v_0 > 1$ (and still assume that V is symmetric, with v_0 and density on $(-v_0, v_0)$ chosen to ensure that $\|V\|_2 = 1$), $\zeta := z/(\sqrt{n} \ln^\alpha n)$, and $\vartheta := \zeta^{1/2} n = z^{1/2} n^{3/4} / \ln^{\alpha/2} n$. After these redefinitions, it is easy to verify that

$$\frac{1}{z^3 n^{1/2}} = \frac{\ln^{3\alpha}(\zeta^{1/2} n)}{\zeta^{3/2} \ln^{3\alpha} n} \frac{n}{\vartheta^3 \ln^{3\alpha} \vartheta} \asymp \frac{\ln^{3\alpha}(\zeta^{1/2} n)}{\zeta^{3/2} \ln^{3\alpha} n} n\mathbb{P}(V > \vartheta) = o(n\mathbb{P}(V > \vartheta)),$$

from which (B.5) follows and the contradiction is derived as done previously. \square

Appendix C: Proof of explicit Berry–Esseen bounds for the Pearson statistic

Proof of Corollary 3.8. For $\alpha \geq 1$, let

$$y_\alpha := \|Y\|_\alpha \quad \text{and} \quad z_\alpha := \|Z\|_\alpha.$$

Also adopt the notation of Theorem 3.4, with $\rho = 0$, so that $V = (Y, Z, Y^2 - 1, Z^2 - 1, YZ)$, $L(V) = YZ$, and $\tilde{\sigma} = \|YZ\|_2$. Take any natural number N_0 and any real number $b_3 > 0$, and consider the two cases: (i) $n \leq N_0 - 1$ and (ii) $n \geq N_0$.

In the first case, when $n \leq N_0 - 1$, note that $1 \leq (y_6^6 + z_6^6)/2$ (since $1 = y_2 \leq y_6$ and $1 = z_2 \leq z_6$) and $\tilde{\sigma}^3 \leq (y_4 z_4)^3 \leq y_6^3 z_6^3 \leq (y_6^6 + z_6^6)/2$ (which follows by Hölder's and Young's inequalities). Then

$$|\mathbb{P}(\sqrt{n}R/\tilde{\sigma} \leq z) - \Phi(z)| \leq 1 \leq \frac{\sqrt{N_0 - 1}}{\sqrt{n}} \leq \frac{y_6^6 + z_6^6}{\sqrt{n}} \left(B_{0,1} + \frac{B_{3,1}}{\tilde{\sigma}^3} \right), \tag{C.1}$$

where

$$(B_{0,1}, B_{3,1}) := \frac{\sqrt{N_0 - 1}}{2(1 + b_3)} (1, b_3). \tag{C.2}$$

Suppose then that $n \geq N_0$. Take any $\epsilon \in (0, \sqrt{3}/2)$ and $c_* \in [\frac{1}{2}, 1)$ so that the conditions of Theorem 2.11 are satisfied (cf. the discussion following (3.8)); also introduce the parameter $\kappa > 0$. Recall the notation in (2.21), so that

$$\begin{aligned} \varsigma_3 &= \|YZ\|_3/\tilde{\sigma} \leq y_6 z_6/\tilde{\sigma}, \quad \varsigma_3^3 \leq \frac{1}{2}(y_6^6 + z_6^6)/\tilde{\sigma}^3, \\ 1 \leq v_2^3 \leq v_3^3 &\leq \sup_{(y,z) \in \mathbb{R}^2} \frac{(y^2 + z^2 + (y^2 - 1)^2 + (z^2 - 1)^2 + y^2 z^2)^{3/2}}{1 - y^2 + 1 - z^2 + y^6 + z^6} (y_6^6 + z_6^6) \\ &= \frac{3^{3/2}}{2} (y_6^6 + z_6^6), \\ v_2^2 \leq v_3^2 &\leq 1 + \frac{2}{3^{3/2}} v_3^3 \leq \frac{1}{2}(y_6^6 + z_6^6) + \frac{2}{3^{3/2}} v_3^3 \leq \frac{3}{2}(y_6^6 + z_6^6), \\ \varsigma_3 v_2^2 \leq \varsigma_3 v_3^2 &\leq y_6 z_6 v_3^2/\tilde{\sigma} \leq (y_6^3 z_6^3 + \frac{2}{3^{3/2}} v_3^3)/\tilde{\sigma} \leq \frac{3}{2}(y_6^6 + z_6^6)/\tilde{\sigma}; \end{aligned}$$

in the last two lines we use the following instance of Young's inequality: $ab \leq a^3 + 2(b/3)^{3/2}$ for $a \geq 0$ and $b \geq 0$. Then (2.29) implies

$$\left| \mathbb{P}\left(\frac{R}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{y_6^6 + z_6^6}{\sqrt{n}} \left(A_0 + \frac{A_1}{\tilde{\sigma}} + \frac{A_2}{\tilde{\sigma}^2} + \frac{A_3}{\tilde{\sigma}^3} \right) \leq \frac{y_6^6 + z_6^6}{\sqrt{n}} \left(B_{0,2} + \frac{B_{3,2}}{\tilde{\sigma}^3} \right), \tag{C.3}$$

where

$$\begin{aligned} A_0 &:= \frac{1}{2} (\mathfrak{K}_0)_+ + \frac{3^{3/2}(2+1/\sqrt{N_0})}{2\epsilon^3 N_0}, \quad A_1 := \frac{3}{2} (\mathfrak{K}_{20} + \mathfrak{K}_{30}) \tilde{\sigma}, \\ A_2 &:= \frac{3}{2} (\mathfrak{K}_{21} + \mathfrak{K}_{31}) \tilde{\sigma}, \quad A_3 := \frac{1}{2} \mathfrak{K}_1, \end{aligned} \tag{C.4}$$

with N_0 replacing n in the expressions $\mathfrak{K}_1, \dots, \mathfrak{K}_{3,1}$,

$$B_{0,2} := A_0 + \frac{2}{3} \kappa^{-3/2} A_1 + \frac{1}{3} \kappa^{-3} A_2, \quad \text{and} \quad B_{3,2} := A_3 + \frac{1}{3} \kappa^3 A_1 + \frac{2}{3} \kappa^{3/2} A_2. \tag{C.5}$$

Then (C.1) and (C.3) yield the desired inequality (3.11) if we let

$$B_0 := B_{0,1} \vee B_{0,2} \quad \text{and} \quad B_3 := B_{3,1} \vee B_{3,2}. \tag{C.6}$$

We shall show that, for f as in (3.8),

$$\begin{aligned} (2.1) \text{ holds for any pair} \\ (\epsilon, M_\epsilon) \in \{(0.06, 1.094), (0.17, 1.365), (0.25, 1.688), (0.30, 1.962)\}. \end{aligned} \tag{C.7}$$

Then, substituting the values of the parameters b_3 , N_0 , ϵ , c_* , and κ given in the table below into the expressions for B_0 and B_3 in (C.6) (which depend on the expressions in (C.2), (C.5), (C.4), and (2.30)), one will see that (3.11) holds for any of the pairs (B_0, B_3) listed in (3.12).

b_3	N_0	ϵ	c_*	κ	B_0	B_3
1	209	0.25	0.77	0.983	3.61	3.61
8	405	0.3	0.877	1.745	1.12	8.94
1/8	900	0.17	0.6115	0.4416	13.33	1.69
27	965	0.3	0.909	2.339	0.56	14.97
1/27	5674	0.06	0.5635	0.28273	36.32	1.37

To complete the proof of Corollary 3.8, it now remains to verify (C.7). Toward that end, take any $\epsilon \in (0, \sqrt{3}/2)$, and recall the definition (3.8) of f (with $\rho = 0$) to see that

$$\begin{aligned} f(x_1, x_2, x_3, x_4, x_5) &\equiv f(-x_1, -x_2, x_3, x_4, x_5) \\ &\equiv -f(-x_1, x_2, x_3, x_4, -x_5) \\ &\equiv -f(x_1, -x_2, x_3, x_4, -x_5) \\ &\equiv f(x_2, x_1, x_4, x_3, x_5) \end{aligned}$$

for any $\mathbf{x} \in \mathbb{R}^5$ such that $\|\mathbf{x}\| \leq \epsilon$. The above identities then imply

$$M_\epsilon^* := \sup_{\|\mathbf{x}\| \leq \epsilon} \|f''(\mathbf{x})\| = \sup[\|f''(\mathbf{x})\| : \mathbf{x} \in B_\epsilon \cap \tilde{\mathbb{R}}^5];$$

here B_ϵ denotes the open ϵ -ball about the origin and

$$\tilde{\mathbb{R}}^5 := [\mathbf{x} \in \mathbb{R}^5 : \text{Sgn}(x_1) = \text{Sgn}(x_2) = \text{Sgn}(x_5) \text{ and } x_3 \leq x_4],$$

where $\text{Sgn}(x) := \mathbb{I}\{x \geq 0\} - \mathbb{I}\{x < 0\}$.

Next take any positive $m \in \mathbb{N}$, and let $\delta_\epsilon := \epsilon/m$. For any $\mathbf{u} = (u_1, \dots, u_5) \in \mathbb{Z}^5$, let

$$C_{\mathbf{u}} := \prod_{j=1}^5 [u_j \delta_\epsilon, (u_j + 1) \delta_\epsilon], \quad \text{and} \quad \mathbf{c}_{\mathbf{u}} := ((u_1 + \frac{1}{2}) \delta_\epsilon, \dots, (u_5 + \frac{1}{2}) \delta_\epsilon);$$

that is, $C_{\mathbf{u}}$ is the cube of side length δ_ϵ with its “southwest” corner at the point $\delta_\epsilon \mathbf{u}$ and center at $\mathbf{c}_{\mathbf{u}}$. Introduce also the set

$$\begin{aligned} U &:= \left\{ \mathbf{u} \in \mathbb{Z}^5 \cap \tilde{\mathbb{R}}^5 : B_\epsilon \cap C_{\mathbf{u}} \neq \emptyset \right\} \\ &= \left\{ \mathbf{u} \in \mathbb{Z}^5 \cap \tilde{\mathbb{R}}^5 : \sum_{i=1}^5 (u_i + \frac{1}{2} - \frac{1}{2} \text{Sgn}(u_i))^2 < m^2 \right\}, \end{aligned}$$

so that $B_\epsilon \cap \tilde{\mathbb{R}}^5 \subseteq \bigcup_{\mathbf{u} \in U} C_{\mathbf{u}}$. Then

$$M_\epsilon^* \leq \max_{\mathbf{u} \in U} \sup_{\mathbf{x} \in C_{\mathbf{u}}} \|f''(\mathbf{x})\| \leq \max_{\mathbf{u} \in U} \left(\|f''(\mathbf{c}_{\mathbf{u}})\| + \sup_{\mathbf{x} \in C_{\mathbf{u}}} \|f''(\mathbf{x}) - f''(\mathbf{c}_{\mathbf{u}})\|_F \right)$$

$$\leq \max_{\mathbf{u} \in U} \left(\|f''(\mathbf{c}_{\mathbf{u}})\| + \sqrt{5} \frac{\delta_\epsilon}{2} \sup_{\mathbf{x} \in C_{\mathbf{u}}} \|f'''(\mathbf{x})\|_F \right), \tag{C.8}$$

where

$$\|f'''(\mathbf{x})\|_F := \left(\sum_{i,j,k=1}^5 (f_{ijk}(\mathbf{x}))^2 \right)^{1/2}$$

and $f_{ijk} = \partial^3 f / (\partial x_i \partial x_j \partial x_k)$; here we assume that m is chosen large enough (whence δ_ϵ is small enough) so as to ensure f_{ijk} exists and is continuous on each cube $C_{\mathbf{u}}$ (i.e. $\min_{\mathbf{u} \in U} \inf_{\mathbf{x} \in C_{\mathbf{u}}} [(1 + x_3 - x_1^2) \wedge (1 + x_4 - x_2^2)] > 0$).

Take now any $\mathbf{u} \in U$, and then take any $\mathbf{x} \in \text{int } C_{\mathbf{u}}$, so that $x_j \neq 0$ for any $j \in \{1, \dots, 5\}$. It is easy to see with a computer algebra system (CAS) that

$$\|f'''(\mathbf{x})\|_F^2 = \frac{3\tilde{x}_3\tilde{x}_4}{64} p(\tilde{\mathbf{x}}),$$

where $\tilde{\mathbf{x}} := (\tilde{x}_1, \dots, \tilde{x}_5) := \left(x_1, x_2, \frac{1}{1 + x_3 - x_1^2}, \frac{1}{1 + x_4 - x_2^2}, x_5 \right),$ (C.9)

and p is a polynomial, namely, the sum of 172 monomials with integer coefficients; note that \tilde{x}_3 and \tilde{x}_4 are both positive. Further, $p(\tilde{\mathbf{x}})$ can be bounded from above by bounding each of the 172 monomials. To do that, for $j \in \{1, 2, 5\}$ introduce

$$\tilde{x}_{j,1} := \left(u_j + \frac{1}{2} + \frac{1}{2} \text{Sgn}(u_j) \right) \delta_\epsilon \quad \text{and} \quad \tilde{x}_{j,-1} := \left(u_j + \frac{1}{2} - \frac{1}{2} \text{Sgn}(u_j) \right) \delta_\epsilon,$$

so that $|\tilde{x}_{j,-1}| \leq |\tilde{x}_j| \leq |\tilde{x}_{j,1}|$; also, for $j \in \{3, 4\}$ let

$$\tilde{x}_{j,1} := \frac{1}{1 + u_j \delta_\epsilon - \tilde{x}_{j-2,1}^2} \quad \text{and} \quad \tilde{x}_{j,-1} := \frac{1}{1 + (u_j + 1) \delta_\epsilon - \tilde{x}_{j-2,-1}^2},$$

so that $0 < \tilde{x}_{j,-1} \leq \tilde{x}_j \leq \tilde{x}_{j,1}$. Then, for any nonnegative integers d_1, \dots, d_5 , any integer a , and $s := \text{Sgn}(a) \text{Sgn}(u_1)^{d_1} \text{Sgn}(u_2)^{d_2} \text{Sgn}(u_5)^{d_5}$,

$$a \tilde{x}_1^{d_1} \tilde{x}_2^{d_2} \tilde{x}_3^{d_3} \tilde{x}_4^{d_4} \tilde{x}_5^{d_5} = s |a| |\tilde{x}_1|^{d_1} \dots |\tilde{x}_5|^{d_5} \leq s |a| |\tilde{x}_{1,s}|^{d_1} \dots |\tilde{x}_{5,s}|^{d_5} \\ = a \tilde{x}_{1,s}^{d_1} \tilde{x}_{2,s}^{d_2} \tilde{x}_{3,s}^{d_3} \tilde{x}_{4,s}^{d_4} \tilde{x}_{5,s}^{d_5}, \tag{C.10}$$

which follows since $\tilde{x}_j \geq 0$ whenever $u_j \geq 0$ (and $\tilde{x}_j \leq 0$ whenever $u_j < 0$) for $j \in \{1, 2, 5\}$. Replacing each of the monomial summands in $p(\tilde{\mathbf{x}})$ with their upper bound in (C.10), we see from (C.9) that

$$\|f'''(\mathbf{x})\|_F \leq \frac{\sqrt{3\tilde{x}_{3,1}\tilde{x}_{4,1}}}{8} \sqrt{p_{\text{Sgn}(u_1)}(\tilde{x}_{1,1}, \dots, \tilde{x}_{5,1}, \tilde{x}_{1,-1}, \dots, \tilde{x}_{5,-1})}, \tag{C.11}$$

where p_1 and p_{-1} are each polynomials in the 10 variables (in fact, p_{-1} is a polynomial in only the five variables $\tilde{x}_{1,1}, \dots, \tilde{x}_{5,1}$, as it turns out that $s = 1$ for each of the monomials of $p(\tilde{\mathbf{x}})$ for $\mathbf{u} \in U$ with $u_1 < 0$).

Thus, combining (C.8) and (C.11), one has

$$M_\epsilon^* \leq \max_{\mathbf{u} \in U} \left(\|f''(\mathbf{c}_{\mathbf{u}})\| + \frac{\epsilon \sqrt{15\tilde{x}_{3,1}\tilde{x}_{4,1}}}{16m} \sqrt{p_{\text{Sgn}(u_1)}(\tilde{x}_{1,1}, \dots, \tilde{x}_{5,1}, \tilde{x}_{1,-1}, \dots, \tilde{x}_{5,-1})} \right).$$

One can then write a program in a CAS which will give an algebraic number for the latter upper bound (and then to bound that algebraic number with a rational). In particular, upon letting $m = 19$ and implementing the bound above for $\epsilon \in \{\frac{6}{100}, \frac{17}{100}, \frac{25}{100}, \frac{30}{100}\}$, (C.7) follows. \square

Appendix D: Asymptotic behavior of the uniform and nonuniform Berry–Esseen bounds

The bounds in Theorems 2.11 and A.2 are complicated in appearance. However, in particular applications such as the one presented in Corollary 3.8, the resulting bounds are of much simpler structure, with explicit numerical constants, which are also rather moderate in size, especially in the uniform bounds. The following corollary shows that the asymptotic behavior of the uniform and nonuniform BE-type bounds given in Theorems 2.11 and A.2 is rather simple as well (especially in the “nonuniform” case) and the corresponding constants are again moderate in size.

Corollary D.1. *Assume that the conditions of Theorem 2.11 hold, and also that f'' is twice continuously differentiable in a neighborhood of the origin. Then*

$$\limsup_{n \rightarrow \infty} \sup_{z \in \mathbb{R}} \sqrt{n} \left| \mathbb{P}\left(\frac{f(\bar{V})}{\bar{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq 0.63925 + 0.83554\varsigma_3^3 + \frac{y_*}{2} + \frac{1}{2} \sqrt{(\varsigma_3^3 - 1)(\varsigma_3^3 - 1 + 2y_*)} \tag{D.1}$$

$$\leq 0.13925 + 1.33554\varsigma_3^3 + y_*, \tag{D.2}$$

where

$$y_* := \frac{\|f''(0)\|}{\bar{\sigma}} \left(\left(\frac{2}{\pi}\right)^{1/6} + \varsigma_3 \right) v_2^2.$$

Also, for any positive increasing unbounded function g on \mathbb{N}

$$\limsup_{n \rightarrow \infty} \sup_{g(n) \leq z \leq \sqrt{n}/g(n)} z^3 \sqrt{n} \left| \mathbb{P}\left(\frac{f(\bar{V})}{\bar{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq 30.2211\varsigma_3^3. \tag{D.3}$$

In the case of a linear statistic, the simple expression in (D.2) obviously further simplifies to $0.13925 + 1.33554\varsigma_3^3$.

It will be possible to replace the factor 30.2211 in (D.3) by any improved constant factor that one will be able to obtain in place of 30.2211 in the nonuniform BE inequality (2.19) for linear statistics. A substantial improvement of this constant factor appears quite possible; see e.g. [60, 61].

As one can see, in the expressions of the asymptotic uniform bounds in (D.1) the higher moment v_3 disappears, and in the asymptotic nonuniform bound in (D.3) the moment v_2 disappears as well; however, Corollary D.1 inherits the condition $v_3 < \infty$ from Theorems 2.11 and A.2 – where, as seen from Remark 3.6, this condition is essential in general.

Proof of Corollary D.1. Let $n \rightarrow \infty$. Following the lines of the proof of (2.29), one can see that the bound there equals

$$\frac{0.13925 + 0.33554\varsigma_3^3}{\sqrt{n}} + \frac{4\delta}{2c_*} + \frac{\mathfrak{C}}{\sqrt{n}} + \frac{\mathfrak{K}_\epsilon}{\sqrt{n}}, \tag{D.4}$$

where $\mathfrak{C} := (\mathfrak{K}_{20} + \mathfrak{K}_{21}\varsigma_3)v_2^2 + (\mathfrak{K}_{30} + \mathfrak{K}_{31}\varsigma_3)v_3^2$ is an upper bound on $\frac{1}{2c_*}(\mathbb{E}|W\bar{\Delta}| + \sum_i \mathbb{E}|\xi_i(\bar{\Delta} - \Delta_i)|)$ – cf. (4.9). Restricting c_* to be in $[\frac{1}{2}, 1)$ and then letting δ be as in (4.24), so that $\delta \rightarrow 0$, by [65, Remark 2.2] the term 4δ in the bound (D.4) may be replaced by

$$2\delta + \frac{\delta^2}{c_*} + 2\delta\sqrt{\frac{\delta}{c_*} + \frac{\delta^2}{4c_*^2} + \frac{\mathfrak{C}}{\sqrt{n}}} \sim 2\delta.$$

So, the term $\mathfrak{K}_1 = 0.33554 + \frac{1}{2c_*(1-c_*)}$ in (2.29) can be replaced by one asymptotic to $0.33554 + \frac{1}{4c_*(1-c_*)}$ and similarly the term \mathfrak{K}_0 may be replaced by $0.13925 - \frac{(2c_*-1)^2}{4c_*(1-c_*)}$.

Let now $\epsilon = \epsilon_n = n^{-1/8}$; the assumed continuity of f'' implies $M_\epsilon \downarrow \|f''(0)\|$, and from (2.30) we see that $\mathfrak{K}_\epsilon \downarrow 0$. Moreover, then $(\mathfrak{K}_{20}, \mathfrak{K}_{21}, \mathfrak{K}_{30}, \mathfrak{K}_{31}) \rightarrow \frac{\|f''(0)\|}{2c_*\tilde{\sigma}}((2/\pi)^{1/6}, 1, 0, 0)$.

Thus,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sup_{z \in \mathbb{R}} \sqrt{n} \left| \mathbb{P}\left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \\ & \leq 0.13925 + 0.33554\varsigma_3^3 + \frac{\varsigma_3^3 - (2c_* - 1)^2}{4c_*(1 - c_*)} + \frac{\|f''(0)\|}{2c_*\tilde{\sigma}} \left(\left(\frac{2}{\pi}\right)^{1/6} + \varsigma_3 \right) v_2^2. \end{aligned}$$

Since

$$\min_{c_* \in [1/2, 1)} \left(\frac{\varsigma_3^3 - (2c_* - 1)^2}{4c_*(1 - c_*)} + \frac{y_*}{2c_*} \right) = \frac{1 + \varsigma_3^3 + y_* + \sqrt{(\varsigma_3^3 - 1)(\varsigma_3^3 - 1 + 2y_*)}}{2},$$

the inequality (D.1) follows. As for (D.2), it follows because the square root term above is no greater than $\tilde{c}_3^3 - 1 + y_*$.

To prove (D.3), fix any real $\tilde{\theta} > 0$ and let $z_0 = g(n)$, $\omega = \tilde{\sigma}/g(n)$, $K_1 = \sqrt{n}/\varsigma_3^3$, $K_2 = \tilde{\sigma}^3 z_0^3 \sqrt{n}/v_2^4$, and $K_3 = \tilde{\sigma}^3 z_0^3 \sqrt{n}/v_3^3$, so that conditions (A.14) and (A.15) hold for all $z \in [g(n), \sqrt{n}/g(n)]$. Then, for $z \geq z_0$ and large enough n we have $z^3 e^{-z/\tilde{\theta}} \leq z_0^3 e^{-z_0/\tilde{\theta}} \rightarrow 0$. One can clearly choose values for the corresponding parameters so that (i) $\mathfrak{K}_{e0}, \dots, \mathfrak{K}_{e3}$ be absolutely bounded; (ii) $\mathfrak{K}_{n21}, \mathfrak{K}_{n22}, \mathfrak{K}_{n31}$, and \mathfrak{K}_{n32} all vanish in the limit (since $\omega \downarrow 0$); and (iii) $\mathfrak{K}_{n1} \rightarrow 30.2211 + \pi_2^{-3}$. Moreover, one can replace the factor ς_3^3 on the right-hand side of inequality (A.31) by the asymptotically much smaller expression $\mathbb{E} \left(\frac{L(V)}{\tilde{\sigma}} \right)^3 \mathbb{I}\{L(V) > \pi_2 \tilde{\sigma} z \sqrt{n}\} = o(\varsigma_3^3)$. Then the limit of the corresponding improved expression for \mathfrak{K}_{n1} becomes just 30.2211, instead of $30.2211 + \pi_2^{-3}$. Now (D.3) follows by Theorem A.2. \square

Appendix E: Compactness of the covariance operator

Here we give a short proof that the covariance operator of a random vector with finite second moment is compact. Let X be a random vector taking values in a separable Hilbert space \mathbb{H} such that $\mathbb{E}\|X\|^2 < \infty$ and $\mathbb{E}X = \mu$. Then the covariance operator $R: \mathbb{H} \rightarrow \mathbb{H}$ is defined by

$$Rx := \mathbb{E}\langle x, X - \mu \rangle (X - \mu) = \mathbb{E} \overline{\langle X - \mu, x \rangle} (X - \mu);$$

let us assume w.l.o.g. that $\mu = 0$. Note that R is both self-adjoint and non-negative-definite: for all $x, y \in \mathbb{H}$

$$\langle Rx, y \rangle = \mathbb{E}\langle x, X \rangle \langle X, y \rangle = \mathbb{E} \overline{\langle y, X \rangle} \overline{\langle X, x \rangle} = \overline{\langle Ry, x \rangle} = \langle x, Ry \rangle$$

and

$$\langle Rx, x \rangle = \mathbb{E}\langle x, X \rangle \langle X, x \rangle = \mathbb{E} \|\langle x, X \rangle\|^2 \geq 0.$$

Now let $(e_j)_{j \in \mathbb{N}}$ be any orthonormal basis of \mathbb{H} , so that $X = \sum_j \langle X, e_j \rangle e_j$. Further take any $x \in \mathbb{H}$, so that $Rx = \mathbb{E}\langle x, X \rangle \sum_j \langle X, e_j \rangle e_j$. For $n \in \mathbb{N}$, define the operator R_n by $R_n x = \mathbb{E}\langle x, X \rangle \sum_{j=1}^n \langle X, e_j \rangle e_j$, and note that the range of R_n is finite-dimensional. Moreover, if $\|x\| \leq 1$, then

$$\begin{aligned} \|(R - R_n)x\| &= \left\| \mathbb{E}\langle x, X \rangle \sum_{j=n+1}^{\infty} \langle X, e_j \rangle e_j \right\| \\ &\leq \mathbb{E} \left\| \langle x, X \rangle \sum_{j=n+1}^{\infty} \langle X, e_j \rangle e_j \right\| \\ &\leq \mathbb{E}\|X\| \sqrt{\sum_{j=n+1}^{\infty} \langle X, e_j \rangle^2} \xrightarrow{n \rightarrow \infty} 0; \end{aligned}$$

the limit holds by dominated convergence, since

$$\sqrt{\sum_{j=n+1}^{\infty} \langle X, e_j \rangle^2} \leq \sqrt{\sum_{j=1}^{\infty} \langle X, e_j \rangle^2} = \|X\|.$$

As x was arbitrary and the above majorant of $\|(R - R_n)x\|$ does not depend on x , it follows that $\|R - R_n\| \rightarrow 0$; that is, R is the limit (in the operator norm) of a sequence of finite-dimensional linear operators on \mathbb{H} , and so is compact.

Appendix F: On the spectral decomposition of a covariance operator of a random vector in an arbitrary separable Hilbert space

Let X be a random vector in a separable Hilbert space $(H, \langle \cdot, \cdot \rangle)$ with $\mathbb{E}\|X\|^2 < \infty$. Let R be the covariance operator of X . So, R is self-adjoint. Obviously, any self-adjoint operator is normal. Hence, by [37, Theorem 2.10, page 260],

$$R = \sum_{\lambda \in \Lambda} \lambda P_\lambda, \tag{F.1}$$

where Λ is the (necessarily at most countable) set of all (necessarily nonnegative) eigenvalues of R ; (in the case when the set Λ is infinite) the sum converges in the operator norm; and, for each $\lambda \in \Lambda$, P_λ is the orthoprojector onto the eigenspace (say E_λ) of λ , which is necessarily of a finite dimension $n_\lambda := \dim E_\lambda = \text{tr } P_\lambda$ if $\lambda \neq 0$. At that,

$$\sum_{\lambda \in \Lambda} P_\lambda = I, \quad (\text{F.2})$$

the identity operator, and the eigenspaces E_λ are pairwise mutually orthogonal:

$$P_\lambda P_\mu = I\{\lambda = \mu\}P_\lambda \quad (\text{F.3})$$

for all λ and μ in Λ .

Moreover, for each $\lambda \in \Lambda$, let B_λ be any orthonormal basis of E_λ , so that $B := \bigcup_{\lambda \in \Lambda} B_\lambda$ is an orthonormal basis of H . Then $\text{tr } R = \sum_{\lambda \in \Lambda} \lambda n_\lambda = \sum_{\lambda \in \Lambda} \sum_{e \in B_\lambda} \langle Re, e \rangle = \sum_{e \in B} \mathbb{E} |\langle e, X \rangle|^2 = \mathbb{E} \sum_{e \in B} |\langle e, X \rangle|^2 = \mathbb{E} \|X\|^2 < \infty$, so that $\sum_{\lambda \in \Lambda} \lambda n_\lambda < \infty$. So, the set Λ of all eigenvalues of R may have at most one limit point, and any limit point of Λ must be 0.

The spectrum $\text{sp } R$ of R is defined as the set of all $z \in \mathbb{C}$ such that the linear operator $R - zI$ does not have a bounded inverse. It follows that $\text{sp } R$ coincides with Λ if $\dim H < \infty$ and with $\Lambda \cup \{0\}$ if $\dim H = \infty$. The complementary set $\text{res } R := \mathbb{C} \setminus \text{sp } R$ is called the resolvent set. Let $B(H)$ denote the Banach space of all bounded linear operators $A: H \rightarrow H$.

One can now define the resolvent $\mathcal{R}: \text{res } R \rightarrow B(H)$ by the formula

$$\mathcal{R}(z) := (R - zI)^{-1} = \sum_{\lambda \in \Lambda} \frac{1}{\lambda - z} P_\lambda; \quad (\text{F.4})$$

the latter equality can be easily verified in view of (F.1), (F.3), and (F.2), because $R - zI = \sum_{\lambda \in \Lambda} (\lambda - z) P_\lambda$.

Take now any nonzero $\lambda \in \Lambda$, which is necessarily an isolated point of the set Λ . So, there is an open disc D_λ in \mathbb{C} such that $\lambda \in D_\lambda$ but no other point of the set $\Lambda \cup \{0\}$ is in the closure of D_λ . Let Γ_λ be the boundary of D_λ . Then, by (F.4) and the Cauchy integral theorem,

$$P_\lambda = -\frac{1}{2\pi i} \int_{\Gamma_\lambda} \mathcal{R}(z) dz, \quad (\text{F.5})$$

whence

$$\lambda = \frac{1}{n_\lambda} \text{tr } P_\lambda = \frac{1}{n_\lambda} \text{tr } RP_\lambda = -\frac{1}{2\pi i n_\lambda} \int_{\Gamma_\lambda} \text{tr } R\mathcal{R}(z) dz. \quad (\text{F.6})$$

Formulas (F.5) and (F.6) are important, because it is comparatively easy to analyze the resolvent.

References

- [1] ANASTASIOU, A. and LEY, C. (2015). New simpler bounds to assess the asymptotic normality of the maximum likelihood estimator. <http://arxiv.org/abs/1508.04948>.

- [2] ANASTASIOU, A. and REINERT, G. (2015). Bounds for the normal approximation of the maximum likelihood estimator. <http://arxiv.org/abs/1411.2391>.
- [3] BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.
- [4] BENTKUS, V., BLOZNELIS, M. and GÖTZE, F. (1996). A Berry-Esséen bound for Student's statistic in the non-i.i.d. case. *J. Theoret. Probab.* **9** 765–796. [MR1400598 \(97e:60036\)](#)
- [5] BENTKUS, V. and GÖTZE, F. (1993). On smoothness conditions and convergence rates in the CLT in Banach spaces. *Probab. Theory Related Fields* **96** 137–151. [MR1227029 \(94f:60010\)](#)
- [6] BENTKUS, V. and GÖTZE, F. (1996). The Berry-Esseen bound for Student's statistic. *Ann. Probab.* **24** 491–503. [MR1387647 \(97f:62021\)](#)
- [7] BENTKUS, V., JING, B.-Y., SHAO, Q.-M. and ZHOU, W. (2007). Limiting distributions of the non-central t -statistic and their applications to the power of t -tests under non-normality. *Bernoulli* **13** 346–364. [MR2331255 \(2008e:60048\)](#)
- [8] BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451. [MR0471142 \(57 ##10880\)](#)
- [9] BOLTHAUSEN, E. and GÖTZE, F. (1993). The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21** 1692–1710. [MR1245764 \(94i:60036\)](#)
- [10] CHEN, L. H. Y. and FANG, X. (2011). Multivariate normal approximation by Stein's method: the concentration inequality approach (preprint). <http://arxiv.org/abs/1111.4073>.
- [11] CHEN, L. H. Y. and SHAO, Q.-M. (2007). Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* **13** 581–599. [MR2331265](#)
- [12] CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. [MR2724863 \(2011h:62167\)](#)
- [13] CUPIDON, J., EUBANK, R., GILLIAM, D. and RUYMGAART, F. (2008). Some properties of canonical correlations and variates in infinite dimensions. *J. Multivariate Anal.* **99** 1083–1104. [MR2419340 \(2009j:62152\)](#)
- [14] CUPIDON, J., GILLIAM, D. S., EUBANK, R. and RUYMGAART, F. (2007). The delta method for analytic functions of random operators with application to functional data. *Bernoulli* **13** 1179–1194. [MR2364231 \(2009g:62060\)](#)
- [15] DE ACOSTA, A. and SAMUR, J. D. (1979). Infinitely divisible probability measures and the converse Kolmogorov inequality in Banach spaces. *Studia Math.* **66** 143–160. [MR565155 \(81d:60004\)](#)
- [16] DOWSON, D. C. and LANDAU, B. V. (1982). The Fréchet distance between multivariate normal distributions. *J. Multivariate Anal.* **12** 450–455. [MR666017 \(84a:62078\)](#)
- [17] EUBANK, R. L. and HSING, T. (2008). Canonical correlation

- for stochastic processes. *Stochastic Process. Appl.* **118** 1634–1661. [MR2442373 \(2009m:62179\)](#)
- [18] FISHER, T. J., SUN, X. and GALLAGHER, C. M. (2010). A new test for sphericity of the covariance matrix for high dimensional data. *J. Multivariate Anal.* **101** 2554–2570. [MR2719881 \(2012d:62169\)](#)
- [19] GAINES, G., KAPHLE, K. and RUYMGAART, F. (2011). Application of a delta-method for random operators to testing equality of two covariance operators. *Math. Methods Statist.* **20** 232–245. [MR2908760](#)
- [20] GAMBOA, F., JANON, A., KLEIN, T., LAGNOUX-RENAUDIE, A. and PRIEUR, C. (2013). Statistical inference for Sobol pick freeze Monte Carlo method (preprint). <http://arxiv.org/pdf/1303.6447.pdf>.
- [21] GILLIAM, D. S., HOHAGE, T., JI, X. and RUYMGAART, F. (2009). The Fréchet derivative of an analytic function of a bounded operator with some applications. *Int. J. Math. Math. Sci. Art.* ID 239025, 17. [MR2496753 \(2010a:47030\)](#)
- [22] GINÉ, E., GÖTZE, F. and MASON, D. M. (1997). When is the Student t -statistic asymptotically standard normal? *Ann. Probab.* **25** 1514–1531. [MR1457629 \(98j:60033\)](#)
- [23] GIVENS, C. R. and SHORTT, R. M. (1984). A class of Wasserstein metrics for probability distributions. *Michigan Math. J.* **31** 231–240. [MR752258 \(85m:60009\)](#)
- [24] GÖTZE, F. (1986). On the rate of convergence in the central limit theorem in Banach spaces. *Ann. Probab.* **14** 922–942. [MR841594 \(87i:60013\)](#)
- [25] GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. [MR1106283 \(92g:60028\)](#)
- [26] HALL, P. and WANG, Q. (2004). Exact convergence rate and leading term in central limit theorem for Student's t statistic. *Ann. Probab.* **32** 1419–1437. [MR2060303 \(2005e:62025\)](#)
- [27] HE, G., MÜLLER, H.-G. and WANG, J.-L. (2004). Methods of canonical analysis for functional data. *J. Statist. Plann. Inference* **122** 141–159. Contemporary data analysis: theory and methods. [MR2057919](#)
- [28] HEYDE, C. C. (1997). *Quasi-likelihood and its application. Springer Series in Statistics*. Springer-Verlag, New York A general approach to optimal parameter estimation. [MR1461808 \(99f:62003\)](#)
- [29] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. [MR0144363 \(26 ##1908\)](#)
- [30] HOFFMANN-JØRGENSEN, J. and PISIER, G. (1976). The law of large numbers and the central limit theorem in Banach spaces. *Ann. Probab.* **4** 587–599. [MR0423451 \(54 ##11429\)](#)
- [31] HORGAN, D. and MURPHY, C. C. (2013). On the Convergence of the Chi Square and Noncentral Chi Square Distributions to the Normal Distribution. *IEEE Communications Letters* **17** 2233–2236.
- [32] HORN, R. A. and JOHNSON, C. R. (1985). *Matrix analysis*. Cambridge University Press, Cambridge. [MR832183 \(87e:15001\)](#)
- [33] JAIN, N. C. and MARCUS, M. B. (1975). Integrability of infinite sums of

- independent vector-valued random variables. *Trans. Amer. Math. Soc.* **212** 1–36. [MR0385995 \(52 #6854\)](#)
- [34] JI, X. and RUYMGAART, F. H. (2008). Fréchet-differentiation of functions of operators with application to testing the equality of two covariance operators. In *Journal of Physics: Conference Series* **124** 012028. IOP Publishing.
- [35] JIN, G. and MATTHEWS, D. (2014). Reliability demonstration for long-life products based on degradation testing and a Wiener process model. *IEEE Transactions on Reliability* **63** 781–797.
- [36] JOHN, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127. [MR0275568 \(43 #1321\)](#)
- [37] KATO, T. (1995). *Perturbation theory for linear operators. Classics in Mathematics*. Springer-Verlag, Berlin Reprint of the 1980 edition. [MR1335452 \(96a:47025\)](#)
- [38] KOROLJUK, V. S. and BOROVSKICH, Y. V. (1994). *Theory of U-statistics. Mathematics and its Applications* **273**. Kluwer Academic Publishers Group, Dordrecht. Translated from the 1989 Russian original by P. V. Malyshev and D. V. Malyshev and revised by the authors. [MR1472486 \(98e:60033\)](#)
- [39] KOSOROK, M. R. (2008). *Introduction to empirical processes and semi-parametric inference. Springer Series in Statistics*. Springer, New York. [MR2724368 \(2012b:62005\)](#)
- [40] KOTEVSKI, Z. and MITREVSKI, P. (2013). Hybrid fluid modeling approach for performance analysis of P2P live video streaming systems. *Peer-to-Peer Networking and Applications* **7** 410–426.
- [41] KÜCHLER, U. and TAPPE, S. (2013). Tempered stable distributions and processes. *Stochastic Process. Appl.* **123** 4256–4293. [MR3096354](#)
- [42] LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. [MR1926169 \(2003g:62101\)](#)
- [43] LI, K. (2014). Second-order asymptotics for quantum hypothesis testing. *Ann. Statist.* **42** 171–189. [MR3178460](#)
- [44] MICHEL, R. (1981). On the constant in the nonuniform version of the Berry-Esseen theorem. *Z. Wahrsch. Verw. Gebiete* **55** 109–117. [MR606009 \(82c:60042\)](#)
- [45] MOLAVIANJAZI, E. (2015). Private communication.
- [46] MUIRHEAD, R. J. (1982). *Aspects of multivariate statistical theory*. John Wiley & Sons, Inc., New York Wiley Series in Probability and Mathematical Statistics. [MR652932 \(84c:62073\)](#)
- [47] NAGAO, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. [MR0339405 \(49 #4164\)](#)
- [48] NOETHER, G. E. (1955). On a theorem of Pitman. *Ann. Math. Statist.* **26** 64–68. [MR0070122 \(16,1133g\)](#)
- [49] OLKIN, I. and PUKELSHEIM, F. (1982). The distance between two random vectors with given dispersion matrices. *Linear Algebra Appl.* **48** 257–263. [MR683223 \(84f:62062\)](#)
- [50] PETROV, V. V. (1975). *Sums of independent random variables*.

- Springer-Verlag, New York. Translated from the Russian by A. A. Brown, *Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82*. MR0388499 (52 ##9335)
- [51] PFANZAGL, J. (1971). The Berry-Esseen bound for minimum contrast estimates. *Metrika* **17** 82–91. MR0295467 (45 ##4533)
- [52] PFANZAGL, J. (1972/73). The accuracy of the normal approximation for estimates of vector parameters. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **25** 171–198. MR0329093 (48 ##7435)
- [53] PINELIS, I. Comparing the asymptotic closeness of the distributions of Pearson's and Fisher's correlation statistics to normality (draft).
- [54] PINELIS, I. (2009). On the Bennett-Hoeffding inequality. <http://arxiv.org/abs/0902.4058>; a shorter version appeared in [62].
- [55] PINELIS, I. (2011). Improved nonuniform Berry-Esseen-type bounds (preprint). <http://arxiv.org/abs/1109.0741>.
- [56] PINELIS, I. (2011). Exact lower bounds on the exponential moments of truncated random variables. *J. Appl. Probab.* **48** 547–560. MR2840315 (2012f:60083)
- [57] PINELIS, I. (2011). Exact bounds on the truncated-tilted mean, with applications (preprint). <http://arxiv.org/abs/1103.3683>.
- [58] PINELIS, I. (2012). An exact bound on the truncated-tilted mean for symmetric distributions (preprint). <http://arxiv.org/abs/1205.5234>.
- [59] PINELIS, I. (2013). An optimal bound on the quantiles of a certain kind of distributions (preprint). <http://arxiv.org/pdf/1301.0294.pdf>.
- [60] PINELIS, I. (2013). On the nonuniform Berry-Esseen bound. <http://arxiv.org/abs/1301.2828>.
- [61] PINELIS, I. (2013). More on the nonuniform Berry-Esseen bound. <http://arxiv.org/abs/1302.0516>.
- [62] PINELIS, I. (2014). On the Bennett-Hoeffding inequality. *Ann. Inst. H. Poincaré Probab. Statist.* **50** 15–27.
- [63] PINELIS, I. (2015). Exact Rosenthal-type bounds. *Ann. Probab.* **43** 2511–2544. MR3395468
- [64] PINELIS, I. (2016). Optimal-order bounds on the rate of convergence to normality for maximum likelihood estimators. <http://arxiv.org/abs/1601.02177>.
- [65] PINELIS, I. and MOLZON, R. (2016). Optimal-order bounds on the rate of convergence to normality in the multivariate delta method (preprint). <http://arxiv.org/abs/0906.0177v5>.
- [66] PINELIS, I. F. (1986). Probability inequalities for sums of independent random variables with values in a Banach space. *Math. Notes* **39** 241–244.
- [67] PINELIS, I. F. (2015). Rosenthal-type inequalities for martingales in 2-smooth Banach spaces. *Theory Probab. Appl.* **59** 699–706.
- [68] PINELIS, I. F. and SAKHANENKO, A. I. (1986). Remarks on inequalities for large deviation probabilities. *Theory Probab. Appl.* **30** 143–148.
- [69] PINELIS, I. F. and UTEV, S. A. (1989). Exact exponential bounds for sums of independent random variables. *Theory Probab. Appl.* **34** 340–346. MR1005745 (91a:60053)

- [70] RIPPL, T., MUNK, A. and STURM, A. (2013). Limit laws of the empirical Wasserstein distance: Gaussian distributions (preprint). <http://arxiv.org/pdf/1507.04090v1.pdf>.
- [71] RÖMISCH, W. (2006). Delta method, infinite dimensional. *Encyclopedia of Statistical Sciences*.
- [72] ROSENTHAL, H. P. (1970). On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables. *Israel J. Math.* **8** 273–303. [MR0271721 \(42 ##6602\)](#)
- [73] SHAO, Q.-M., WANG, Q. et al. (2013). Self-normalized limit theorems: A survey. *Probability Surveys* **10** 69–93.
- [74] SHEVTSOVA, I. (2011). On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands (preprint). <http://arxiv.org/abs/1111.6554>.
- [75] SMALL, C. G. (2010). *Expansions and asymptotics for statistics. Monographs on Statistics and Applied Probability* **115**. CRC Press, Boca Raton, FL. [MR2681183 \(2011m:62005\)](#)
- [76] SRIVASTAVA, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.* **35** 251–272. [MR2328427](#)
- [77] WASSERMAN, L., KOLAR, M. and RINALDO, A. (2014). Berry-Esseen bounds for estimating undirected graphs. *Electron. J. Stat.* **8** 1188–1224.
- [78] ZALESSKII, B. A. (1988). On the accuracy of normal approximation in Banach spaces. *Theory Probab. Appl.* **33** 239–247. [MR954572 \(89h:60014\)](#)
- [79] ZALESSKII, B. A. (1990). The accuracy of Gaussian approximation in Banach spaces. *Theory Probab. Appl.* **34** 747–748. [MR1036728 \(91e:60021\)](#)
- [80] ZEIFMAN, A., KOROLEV, V., SATIN, Y., KOROTYSHEVA, A. and BENING, V. (2014). Perturbation bounds and truncations for a class of Markovian queues. *Queueing Syst.* **76** 205–221. [MR3158075](#)