



Michigan Technological University
Create the Future Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's
Reports - Open

Dissertations, Master's Theses and Master's
Reports

2012

Berry-Esseen bounds for nonlinear statistics, and asymptotic relative efficiency between correlation statistics

Raymond E. Molzon
Michigan Technological University

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Mathematics Commons](#)

Copyright 2012 Raymond E. Molzon

Recommended Citation

Molzon, Raymond E., "Berry-Esseen bounds for nonlinear statistics, and asymptotic relative efficiency between correlation statistics", Dissertation, Michigan Technological University, 2012.
<https://doi.org/10.37099/mtu.dc.etds/200>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Mathematics Commons](#)

BERRY-ESSEEN BOUNDS FOR NONLINEAR STATISTICS, AND ASYMPTOTIC
RELATIVE EFFICIENCY BETWEEN CORRELATION STATISTICS

By
Raymond E. Molzon

A DISSERTATION
Submitted in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
In Mathematical Sciences

MICHIGAN TECHNOLOGICAL UNIVERSITY
2012

© 2012 Raymond E. Molzon

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Mathematical Sciences.

Department of Mathematical Sciences

Dissertation Advisor: *Dr. Iosif Pinelis*

Committee Member: *Dr. Thomas Drummer*

Committee Member: *Dr. Jianping Dong*

Committee Member: *Dr. Susan Amato-Henderson*

Department Chair: *Dr. Mark Gockenbach*

Contents

List of Figures	vii
List of Tables	ix
Preface	xi
Acknowledgments	xiii
Abstract	1
Introduction	3
1 Berry-Esseen bounds for general nonlinear statistics, with applications to Pearson’s and non-central Student’s and Hotelling’s	5
1.1 Introduction	5
1.2 Approximation of the distributions of general abstract nonlinear statistics by the distributions of linear ones	7
1.3 Berry-Esseen bounds for smooth nonlinear functions of sums of independent random vectors	12
1.4 Applications	16
1.4.1 “Quadratic” statistic	20
1.4.2 Student’s T	21
1.4.2.1 Central T and the self-normalized sum	23
1.4.3 Pearson’s R	28
1.4.4 Non-central Hotelling’s T^2 statistic	30
1.5 Proofs	32
1.5.1 Proofs of results from Section 1.2	32
1.5.2 Proofs of results from Section 1.3	37
1.5.3 Proofs of results from Section 1.4	42
1.A An explicit nonuniform bound	46
1.B Optimality of the restriction $z = O(\sqrt{n})$ for the nonuniform bound	52
1.C Proofs of bounds with explicit numerical constants, using a computer algebra system (CAS)	55
1.D On Fisher’s z transform	69
2 A Berry-Esseen type bound for the null distribution of the F-statistic from fixed effects general linear models	71

2.1	Introduction	71
2.1.1	Notation and basic results	73
2.2	Bounding the convergence rate of the F -statistic	74
2.3	Applications	77
2.3.1	Simple linear regression	78
2.3.2	One-way ANOVA	79
2.3.3	Two-way ANOVA	80
2.4	Proof of Theorem 2.2.1	81
3	Relative efficiency between Kendall's, Spearman's, and Pearson's correlation statistics	87
3.1	Introduction	87
3.2	An expression for the ARE between the correlation statistics T , S , and R	89
3.3	On $\text{ARE}_{T,S} = 1$ for tests of independence	92
3.4	The ARE under some specific models	96
3.4.1	Mixtures of independence and complete dependence (MICD)	97
3.4.1.1	On $\text{ARE}_{T,S}(0) = 1$ in the MICD models	99
3.4.1.2	ARE's between R , S , and T in the MICD models	101
3.4.1.3	The Neyman-Pearson test for the MICD AS model	102
3.4.2	Durbin-Stuart models	102
3.4.3	Transformations of the bivariate normal model	104
3.4.3.1	Bivariate lognormal model	105
3.4.3.2	Hyperbolic sine model	106
3.4.3.3	"Power tails" model	107
3.4.3.4	Generalized gamma model	108
3.5	Simulations of the relative efficiency	110
4	Monotonicity of the asymptotic relative efficiency between common correlation statistics in the bivariate normal model	115
4.1	Introduction	115
4.2	Monotonicity properties of the ARE in the bivariate normal model	116
4.3	Proofs	122
4.3.1	Berry-Esseen bounds and expressions for the asymptotic means and variances of R , S , and T	122
4.3.2	Proofs of monotonicity	126
	Future work	137
	References	139

List of Figures

1.1	Comparison of asymptotic bounds for quadratic statistic	21
3.1	Supports of the MICD models	98
3.2	ARE for the MICD models	100
3.3	ARE for Durbin-Stuart models	103
3.4	ARE for the bivariate normal model	105
3.5	ARE for bivariate lognormal model	107
3.6	ARE for hyperbolic sine model	107
3.7	ARE for power tails model	108
3.8	ARE for generalized gamma model	108
3.9	Plots of $\widehat{n}_R/\tilde{n}_R$ for RE simulations	111
3.10	Plots of \widehat{RE}	112
4.1	Plots of $ARE_{R,T}$, $ARE_{T,S}$, $ARE_{R,S}$ in the bivariate normal model	117
4.2	Illustration of piecewise quadratic bounds of Corollary 4.2.3	119
4.3	Illustration of Corollary 4.2.6	121

List of Tables

1.1	Constants associated with nonuniform bound in (1.83)	26
1.2	Constants associated with nonuniform bound in (1.83) when Y is symmetric .	28
1.3	Parameters associated with triples (A_3, A_4, A_6) in (1.80) and (1.86)	57
1.4	Parameters associated with Corollary 1.4.12, for $w_g = 0$	61
1.5	Parameters associated with Corollary 1.4.12, for $w_g = 1$	62
1.6	Parameters associated with Remark 1.4.16, for $w_g = 0$	63
1.7	Parameters associated with Remark 1.4.16, for $w_g = 1$	64
1.8	Parameters associated with pairs (B_0, B_3) in (1.92) and (1.93)	66
3.1	Description of the four MICD models	97
3.2	Asymptotic moments of correlation statistics in the MICD models	99

Preface

The main body of this dissertation consists of four scholarly papers which were collaborative efforts between the author and his graduate school advisor, Dr. Iosif Pinelis. Each dissertation chapter is a version of a paper which has either been submitted to an academic journal or is in preparation for journal submission. Differences between the material printed here and the submitted (or to be submitted) manuscripts are primarily of a typographical nature.

Chapter 1, “Berry-Esseen bounds for general nonlinear statistics, with applications to Pearson’s and non-central Student’s and Hotelling’s”, is a version of a paper of the same title which has been submitted to the journal *Bernoulli* with main author Iosif Pinelis and myself as co-author. There is one significant alteration made to the paper printed here and the submitted manuscript. Namely, (1.26) here relies on a result by Michel [81], whereas the corresponding inequality in the submitted paper cites the paper [89] by Nefedova and Shevtsova. Since submission of the paper, we found an error in the proof of the needed result from [89], and so the paper printed here relies on the work in [81]; this change is then reflected in alterations to Corollary 1.4.4, Corollary 1.4.12, Remarks 1.4.15 and 1.4.16, and Theorem 1.A.2. The bulk of the paper was authored by myself, while the motivation and ideas used in the proofs were due primarily to Dr. Pinelis.

Chapter 2, “A Berry-Esseen type bound for the null distribution of the F -statistic from fixed effects general linear models”, is a paper that is currently in preparation for submission to the pre-print server <http://arxiv.org> and eventual submission to an academic journal. Dr. Pinelis had primarily an advisory role in the generation of this paper, though his input was also crucial in some of the details to the main proof.

Chapter 3, “Relative efficiency between Kendall’s, Spearman’s, and Pearson’s correlation statistics”, is also currently in preparation for submission to <http://arxiv.org> and an academic journal, with Pinelis as main author and Molzon as co-author. Most of the writing in this paper is due to myself; my other main contribution to the paper is the proof of Theorem 3.3.1, while Pinelis provided the bulk of the motivation behind other results and proofs.

Chapter 4, “Monotonicity of the asymptotic relative efficiency between common correlation statistics in the bivariate normal model”, is a version of a paper which has been previously submitted to several journals and also to the pre-print server <http://arxiv.org>. After previous submissions of the paper were rejected for publication, a much slimmer version of this paper is in preparation for submission to another academic journal. The main ideas behind the proofs of the results in the paper are due to Pinelis, while my major contributions to the paper consisted of writing and creation of the Mathematica notebooks used in the proofs.

Acknowledgments

This dissertation is the culmination of several years of the author's work and perseverance, though several individuals deserve special mention for their advisory and supportive roles in the story that has led to this point. My appreciation goes out to all of the close family and friends who had faith in my abilities to finish this degree program. I thank my parents for instilling in me from an early age (wittingly or not) the virtues of honor, humility, and a constant readiness for and appreciation of hard work; the value of their encouragement in my academic pursuits cannot be overstated. I express deep gratitude to all of the dedicated educators in my life, with special thanks to Bruce Bright, Mark Servis, Dr. Janice Glime, Dr. Thomas Drummer, Dr. Susan Amato-Henderson, Dr. Shuanglin Zhang, and Dr. Iosif Pinelis; all of these people have played an important advisory role at various points in my student life, and have helped to cement a love for learning that will be with me to my grave. Along these lines, I would also like to thank the faculty, staff, and fellow graduate students of the Mathematical Sciences department at Michigan Technological University who have helped to create a positive and productive learning environment; singled out for acknowledgement is Dr. Mark Gockenbach for his excellent introduction to the L^AT_EX typesetting language, as well as his support in his role of department chair. This dissertation could not have been written (at least by my own hands) without enormous input from my graduate academic advisor, Dr. Iosif Pinelis. His wisdom and patience were indispensable qualities during our years of collaboration, and I look forward to the day when my mastery of mathematics places me amongst his peers.

Abstract

Four papers, written in collaboration with the author's graduate school advisor, are presented. In the first paper, uniform and non-uniform Berry-Esseen (BE) bounds on the convergence to normality of a general class of nonlinear statistics are provided; novel applications to specific statistics, including the non-central Student's, Pearson's, and the non-central Hotelling's, are also stated. In the second paper, a BE bound on the rate of convergence of the F -statistic used in testing hypotheses from a general linear model is given. The third paper considers the asymptotic relative efficiency (ARE) between the Pearson, Spearman, and Kendall correlation statistics; conditions sufficient to ensure that the Spearman and Kendall statistics are equally (asymptotically) efficient are provided, and several models are considered which illustrate the use of such conditions. Lastly, the fourth paper proves that, in the bivariate normal model, the ARE between any of these correlation statistics possesses certain monotonicity properties; quadratic lower and upper bounds on the ARE are stated as direct applications of such monotonicity patterns.

Introduction

The kernel of this dissertation could be said to have originated with a simple conjecture made by the author's advisor, Dr. Iosif Pinelis, in a course on nonparametric statistics. Namely, it was hypothesized that the Kendall and Spearman rank correlation statistics (referred to here as T and S , respectively) are asymptotically equally efficient as test statistics in a test for independence between two random variables (r.v.'s), in the sense that the asymptotic relative efficiency (ARE) between the two statistics is equal to 1.

Suppose that we are interested in testing the null hypothesis $\theta = \theta_0$ against the alternative $\theta = \theta_1$, and that there are two (sequences of) test statistics T_1 and T_2 which can be used in this testing problem. It is natural to say that T_1 is more efficient than T_2 if $n_{T_1} < n_{T_2}$, where n_{T_j} is the minimal sample size required for the test based on T_j to achieve some prescribed level α and size β ; the relative efficiency of T_1 to T_2 , denoted by RE_{T_1, T_2} , is then defined to be the ratio n_{T_2}/n_{T_1} . While the RE is a natural tool for comparing two test statistics, its practicality is diminished by the fact that power functions for tests are generally difficult to calculate. Fortunately, in many situations we may allow θ_1 to approach θ_0 (while keeping α and β fixed) in a manner which ensures the existence of a limiting value of RE_{T_1, T_2} . This limit is called the Pitman asymptotic relative efficiency (ARE), and can be calculated by the formula

$$\text{ARE}_{T_1, T_2}(\theta_0) = \left(\frac{\sigma_{T_2}(\theta_0)/\mu'_{T_2}(\theta_0)}{\sigma_{T_1}(\theta_0)/\mu'_{T_1}(\theta_0)} \right)^2,$$

where $\mu_{T_j}(\theta) := \lim_{n \rightarrow \infty} \mathbb{E}_\theta T_j$ and $\sigma_{T_j}(\theta) := \sqrt{\lim_{n \rightarrow \infty} n \text{Var}_\theta T_j}$, under some fairly mild conditions.

The validity of the above equation relies in part on the assumption that either test statistic converges to normality at a rate which is uniform with respect to the model parameter θ . Much to our surprise, we could not find any results in the literature where this uniform convergence was explicitly addressed in applications of the ARE to specific models; our search for this uniform convergence for the statistics T and S led us to the topic of Berry-Esseen (BE) bounds. A BE bound, for our purposes, is a bound on the difference between the standard normal distribution function (d.f.) Φ and the d.f. of some statistic of n random variables (r.v.'s) X_1, \dots, X_n ; the bound should decrease to 0 as n approaches ∞ , and it will also generally depend on certain properties of the vector (X_1, \dots, X_n) . The simplest such bound is when the X_j 's are assumed independent, identically distributed (i.i.d.), zero-mean, and unit-variance, in which case the distance to normality of the normalized sum $\sum_i X_i/\sqrt{n}$ can be bounded by $A \mathbb{E}|X_1|^3/\sqrt{n}$ for some absolute constant A .

The literature contains several types of BE bounds, with some variations obtained by altering the assumptions about the i.i.d. property possessed by the sample. In particular,

BE bounds for U - and V -statistics (of which T and S are instances) have been developed by numerous researchers. While such bounds would have been sufficient for our purposes of comparing T and S via the ARE, we also wished to compare these statistics to the Pearson correlation statistic R ; while R is generally only used as an estimator or test statistic when the sampled population is assumed to be normal, it is desirable to have some notion of the robustness of R to deviations from the normality assumption.

In our search for a BE bound on the distribution of R , we came across the paper [19] by Chen and Shao, where a BE bound for a general class of nonlinear statistics (which also includes U - and V -statistics) is proved to exist. While this result could have been applied to the statistic R , it provided a bound of the wrong order; that is, to get a bound of order $\mathcal{O}(1/\sqrt{n})$ the existence of the fourth moments of relevant r.v.'s needed to be assumed, which bucked the trend of other BE bounds where the existence of only the third absolute moment was required. Several years of work eventually culminated in the first chapter of this dissertation, which significantly generalized the results of [19] to provide BE bounds of the correct order on a class of nonlinear statistics (which includes the Student statistic T , the Pearson statistic R , and also the non-central Hotelling statistic T^2).

The aforementioned work prompted the author to search for other commonly used statistics which did not yet have BE bounds in the literature. This search yielded no results concerning the rate of convergence to a limit distribution of the F -statistic used to test hypotheses in a general linear model; the absence of such results was the motivation behind the second chapter of this dissertation, where a BE bound on the distance between the null distribution of the F -statistic and an appropriate χ^2 distribution is proved to exist. Sufficient conditions for this bound to be on the order of $\mathcal{O}(1/\sqrt{n})$ are also considered there, and are shown to hold under some popular applications of the general linear model.

Returning to the original motivation behind the study of BE bounds, we considered the ARE between any of the three statistics R , S , and T . The fruits of this work are embodied in the last two chapters of this dissertation. In the third chapter, it is proved that the ARE between T and S is not always equal to 1 in the test for independence; for a certain class of models, we provide a necessary and sufficient condition for $\text{ARE}_{T,S} = 1$ in the independence test, and a few bivariate models are constructed which provide examples where $\text{ARE}_{T,S} \neq 1$. The results of simulations of the RE between these correlation statistics are also reported there, and they suggest that the ARE is a practical alternative to the RE (in that $\text{ARE} - 1$ and $\text{RE} - 1$ are estimated to share the same sign for the models considered there). We also prove, in the fourth chapter of the dissertation, that each of $\text{ARE}_{R,T}$, $\text{ARE}_{T,S}$, and $\text{ARE}_{R,S}$ are strictly increasing on the unit interval (where the argument to any of these functions is the null correlation ρ_0). Actually, much stronger results are proved which allow us to bound any of these ARE's by piecewise quadratic functions.

The path that led to each of these four papers contained several dead ends, though it also opened into territory which had previously been unexplored by either of us. The dissertation is concluded with some remarks on these topics and their potential to yield other novel results.

Chapter 1

Berry-Esseen bounds for general nonlinear statistics, with applications to Pearson's and non-central Student's and Hotelling's

1.1 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. [92]), 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.

Kendall's and Spearman's correlation coefficients are instances of U -statistics, for which BE bounds are well known; see e.g. [73]. 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 [10] for independent identically distributed (i.i.d.) random variables (r.v.'s) and by Bentkus, Bloznelis and Götze [8] 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 [41].) For more recent developments concerning the Student statistic, see e.g. the 2005 paper by Shao [134] and the 2011 preprint by Pinelis [111].

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 $\mathcal{O}(n^{-1/3})$ 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. (1.87). 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

The material contained in this chapter has been submitted to *Bernoulli*.

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.

Using (instead of the mentioned Rosenthal inequality) exponential inequalities for sums of random vectors due to Pinelis and Sakhnenko [119] or Pinelis [99, 100], for each $p \in (2, 3)$, under the assumption of the finiteness of the p th moment of the norm of the V_i 's, one can obtain a uniform bound of the form $\mathcal{O}(1/n^{p/2-1})$, which is similar to the BE bound for a linear statistic with a comparable moment restriction. However, the corresponding constant factor in the $\mathcal{O}(1/n^{p/2-1})$ will then explode to infinity as $p \uparrow 3$. As for $p \geq 3$, this method produces bounds of order $\mathcal{O}((\ln n)^{3/2}/\sqrt{n})$ (for $p = 3$) and $\mathcal{O}((\ln n)/\sqrt{n})$ (for $p > 3$), with the extra logarithmic factors.

While any of these bounds would have sufficed as far as the ARE is concerned, we became interested in obtaining an optimal-rate BE bound for the Pearson statistic. Soon after that, we came across the remarkable paper by Chen and Shao [19]. Suppose that T is any nonlinear statistic and W is any linear one, and let $\Delta := T - W$; then make the simple observation that

$$-\mathbf{P}(z - |\Delta| \leq W \leq z) \leq \mathbf{P}(T \leq z) - \mathbf{P}(W \leq z) \leq \mathbf{P}(z \leq W \leq z + |\Delta|)$$

for all $z \in \mathbb{R}$. Chen and Shao [19] offer a Stein-type method to provide relatively simple bounds on the two concentration probabilities in the above inequality, hence bounding the distance between T and W ; the reader is referred e.g. to [6] for illustrations of the elegance and power of Stein's method to a wide array of problems. Chen and Shao provided a number of applications of their general results.

However, in the applications that we desired, such as to Pearson's R , it was difficult to deal with $\Delta = T - W$, as defined above. The simple cure applied here was to allow for any $\Delta \geq |T - W|$, so that, for $T = f(\bar{V})$, $W = L(\bar{V})$, and smooth enough f , the random variable Δ could be taken as $\|\bar{V}\|^2$ (up to some multiplicative constant). This allowed for a BE bound of order $\mathcal{O}(1/\sqrt{n})$, though under the excessive moment restriction that $\mathbf{E}\|V_i\|^4 < \infty$.

To obtain a BE bound of the "optimal" order $\mathcal{O}(1/\sqrt{n})$ using only the assumption $\mathbf{E}\|V_i\|^3 < \infty$, we combine the Chen-Shao technique with a Cramér-type tilt transform. Yet another modification was made by introducing a second level of truncation, to obtain a bound of order $\mathcal{O}(1/n^{p/2-1})$ in the case when $\mathbf{E}\|V_i\|^p < \infty$ for $p \in (2, 3)$. Thus we obtain our first group of main results (presented in Section 1.2), on the closeness in distribution of general abstract nonlinear statistics to linear ones. These results may be represented by Theorem 1.2.4, which provides a "nonuniform" upper bound on $|\mathbf{P}(T > z) - \mathbf{P}(W > z)|$ (that is, an upper bound which decreases to 0 in $|z|$), for a general abstract nonlinear statistic T and a general linear statistic W ; a "uniform" bound on $|\mathbf{P}(T > z) - \mathbf{P}(W > z)|$ is given by Theorem 1.2.1.

The other kind of main results, based on Theorems 1.2.1 and 1.2.4, is presented in Section 1.3. For instance, Theorem 1.3.5 provides a nonuniform upper bound on $|\mathbf{P}(f(S) > z) - \mathbf{P}(L(S) > z)|$ and thus may be considered as a bound on the rate of convergence in the delta method for vector statistics; it is the latter bound that took more of our time and effort.

Finally, as applications of the delta-method bounds given in Section 1.3, we present (in Section 1.4) BE-type bounds for the Pearson statistic, as well as for the noncentral Student and Hotelling ones. No such BE bounds appear to be previously known. As for the known

BE bounds for the central Student statistic (obtained by specialized methods, targeting this specific statistic), it turns out that our bounds (even though based on the mentioned results for general nonlinear statistics) compare well with the former ones.

To obtain the delta-method bounds stated in Sections 1.3 and their applications presented in Section 1.4, we use a number of previously known results, including precise exponential and Rosenthal-type bounds developed by Pinelis, Sakhanenko, and Utev [117, 119, 118, 120, 99] and also a number of other known results due to Bennett [7], Hoeffding [51], de Acosta and Samur [28], Nefedova and Shevtsova [89], Ibragimov and Sharakhmetov [57], and Shevtsova [136]. There we also use the recent results developed in [114, 108, 107, 111, 109, 112, 113].

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 organized as follows.

- In Section 1.2, we state and discuss the mentioned upper bounds on $|\mathbb{P}(T > z) - \mathbb{P}(W > z)|$ for general T and W .
- In Section 1.3, the mentioned Theorem 1.3.5 and other results are stated, providing general bounds on the rate of convergence in the vector delta method, that is, bounds on $|\mathbb{P}(f(S) > z) - \mathbb{P}(L(S) > z)|$.
- Applications to several commonly used statistics, namely the non-central Student T , the Pearson R , and the non-central Hotelling T^2 are stated in Section 1.4.
- Proofs of results from Sections 1.2 and 1.3, as well as selected results from Section 1.4, are deferred to Section 1.5.

Certain results and proofs are relegated to appendices.

- The statement and proof of an explicit (and quite complicated in appearance) nonuniform bound on the distance to normality of $f(\bar{V})$ in an i.i.d. setting is provided in Appendix 1.A.
- The nonuniform bounds developed in this paper are valid under the restriction that $z = \mathcal{O}(\sqrt{n})$ (in the i.i.d. case); in Appendix 1.B we prove that this restriction cannot generally be discarded or even relaxed.
- Appendix 1.C contains the proofs of bounds from Section 1.4 which, for practical purposes, make the use of a computer algebra system (CAS) preferable.
- In Appendix 1.D, we discuss the potential application of the bounds presented in Section 1.3 to the Fisher z -transform of the Pearson statistic.

1.2 Approximation of the distributions of general abstract nonlinear statistics by the distributions of linear ones

Let X_1, \dots, X_n be independent r.v.'s with values in some measurable space \mathfrak{X} , and let $T: \mathfrak{X}^n \rightarrow \mathbb{R}$ be a Borel-measurable function. For brevity, let T also stand for $T(X_1, \dots, X_n)$,

the statistic of the random sample $(X_i)_{i=1}^n$. Further let

$$\xi_i := g_i(X_i) \quad \text{and} \quad \eta_i := h_i(X_i) \quad (1.1)$$

for $i = 1, \dots, n$, where $g_i: \mathfrak{X} \rightarrow \mathbb{R}$ and $h_i: \mathfrak{X} \rightarrow \mathbb{R}$ are Borel-measurable functions. Assume that

$$\mathbb{E} \xi_i = 0 \text{ for all } i = 1, \dots, n, \text{ and } \sum_{i=1}^n \mathbb{E} \xi_i^2 = 1. \quad (1.2)$$

Consider the linear statistic

$$W := \sum_{i=1}^n \xi_i. \quad (1.3)$$

Further, take an arbitrary $c_* \in (0, 1)$ and let δ be any real number such that

$$\sum_{i=1}^n \mathbb{E} |\xi_i| (\delta \wedge |\xi_i|) \geq c_*; \quad (1.4)$$

note that such a number δ always exists (because the limit of the left-hand side of (1.4) as $\delta \uparrow \infty$ is 1). Necessarily, $\delta > 0$.

Theorem 1.2.1. *Let Δ be any r.v. such that $|\Delta| \geq |T - W|$ almost surely (a.s.), and for each $i = 1, \dots, n$, let Δ_i be any r.v. such that X_i and $(\Delta_i, W - \xi_i)$ are independent. Take any real number $w > 0$, and let $\overline{\Delta}$ be any r.v. such that*

$$\overline{\Delta} = \Delta \text{ a.s. on the event } \left\{ \max_{1 \leq i \leq n} \eta_i \leq w \right\}. \quad (1.5)$$

Then for all $z \in \mathbb{R}$

$$|\mathbb{P}(T > z) - \mathbb{P}(W > z)| \leq \frac{1}{2c_*} \left(4\delta + \mathbb{E}|W\overline{\Delta}| + \sum_{i=1}^n \mathbb{E} |\xi_i (\overline{\Delta} - \Delta_i)| \right) + \mathbb{P}(\max_i \eta_i > w), \quad (1.6)$$

where δ satisfies (1.4).

Remark 1.2.2. Sacrificing some simplicity in appearance, one can improve the bound in (1.6) by replacing the term 4δ there with

$$2\delta + \frac{\delta^2}{c_*} + 2\delta \sqrt{\frac{1}{2c_*} \left(2\delta + \frac{\delta^2}{2c_*} + \mathbb{E}|W\overline{\Delta}| + \sum_i \mathbb{E} |\xi_i (\overline{\Delta} - \Delta_i)| \right)}; \quad (1.7)$$

the validity of (1.6) after such a replacement will be shown in the proof of Theorem 1.2.1. Evidently, when the upper bound in (1.6) is small, the expression (1.7) will behave like 2δ , in place of 4δ in (1.6).

Remark 1.2.3. Inequality (1.6) above is a rather straightforward generalization of the result (2.3) in Theorem 2.1 by Chen and Shao [19]. The modifications we have made are as follows. First, Δ was defined in [19] as simply equal to $T - W$. Then, in the applications given in our present paper, it becomes problematic to bound the term $\mathbb{E} |\xi_i (T - W - \Delta_i)|$ (which would arise in place of the term $\mathbb{E} |\xi_i (\overline{\Delta} - \Delta_i)|$ in (1.6)). Using the more general condition $|\Delta| \geq |T - W|$ instead of $\Delta = T - W$ allows one to choose a possibly larger Δ so that $\mathbb{E} |\xi_i (\Delta - \Delta_i)|$ be more amenable to analysis. However, if that Δ should happen to be “too

large,” our second generalization allows one to truncate Δ to within acceptable constraints by using the additional truncation level w , as well as $\bar{\Delta}$ and $P(\max_i \eta_i > w)$. The third difference is that in [19] c_* was chosen to be $\frac{1}{2}$; the more general condition $c_* \in (0, 1)$ results in improved explicit constants in the applications.

Before stating the “nonuniform” counterpart of Theorem 1.2.1, let us introduce some notation. For any real a and b , let $a \wedge b$ and $a \vee b$ denote the minimum and maximum, respectively, of a and b ; use also the notation $a_+ := a \vee 0$. For any real-valued r.v. ξ and any $p \in [1, \infty)$, let $\|\xi\|_p := E^{1/p} |\xi|^p$. For the ξ_i ’s as in (1.1), also let

$$\sigma_p := \left(\sum_{i=1}^n \|\xi_i\|_p^p \right)^{1/p} = \left(\sum_{i=1}^n E |\xi_i|^p \right)^{1/p}. \quad (1.8)$$

In proving, and even stating, the forthcoming results of the current paper, we will need several tools for estimating moments and tail probabilities. Let here $\zeta := (\zeta_1, \dots, \zeta_n)$, where ζ_1, \dots, ζ_n are independent real-valued r.v.’s, $S := \sum_i \zeta_i$, and

$$G_\zeta(z) := \sum_{i=1}^n P(\zeta_i > z) \quad \text{for all } z \in \mathbb{R}. \quad (1.9)$$

If the ζ_i ’s are zero-mean, then for each real $\alpha \geq 2$ there exist positive constants $\mathfrak{A}_\mathbb{R}(\alpha)$ and $\mathfrak{B}_\mathbb{R}(\alpha)$, depending only on α , such that

$$\|S\|_\alpha^\alpha \leq \mathfrak{A}_\mathbb{R}^\alpha(\alpha) \sum_i \|\zeta_i\|_\alpha^\alpha + \mathfrak{B}_\mathbb{R}^\alpha(\alpha) \left(\sum_i \|\zeta_i\|_2^2 \right)^{\alpha/2}. \quad (1.10)$$

Such a result will be referred to in this paper as a Rosenthal-type inequality, since it was first obtained by Rosenthal in [127, Theorem 3]; however, the constants there were too large.

Next, we shall need upper bounds on the tail probabilities. Suppose now that $G_\zeta(y) = 0$ for some $y > 0$, i.e. each of the ζ_i ’s is bounded a.s. by y . Then [120, Theorem 2] implies that for any $\lambda \geq 0$

$$E \exp\{\lambda(S - m)\} \leq \text{PU}_{\text{exp}}(\lambda, y, B, \varepsilon) := \exp\left\{ \frac{\lambda^2}{2} B^2(1 - \varepsilon) + \frac{e^{\lambda y} - 1 - \lambda y}{y^2} B^2 \varepsilon \right\}, \quad (1.11)$$

where $B = (\sum_i E \zeta_i^2)^{1/2} < \infty$, $m = ES$, $\varepsilon = \sum_i E(\zeta_i)_+^p / (B^2 y^{p-2}) \in (0, 1)$, and $p \in [2, 3]$. Further, an application of the Markov inequality and (1.11) yield

$$P(S \geq x) \leq \text{PU}_{\text{tail}}(x, y, B, m, \varepsilon) := \inf_{\lambda \geq 0} e^{-\lambda(x-m)} \text{PU}_{\text{exp}}(\lambda, y, B, \varepsilon) \quad \text{for any } x \in \mathbb{R}. \quad (1.12)$$

As functions of the real numbers $\lambda \geq 0$, $y > 0$, $B > 0$, $\varepsilon \in (0, 1)$, x , and m , the bounds PU_{exp} and PU_{tail} possess certain monotonicity properties: PU_{tail} is clearly nondecreasing in $m \in \mathbb{R}$, and from the inequality $e^t - 1 - t - t^2/2 \geq 0$ for all $t \geq 0$ it follows that

$$\text{PU}_{\text{exp}}, \text{ and hence } \text{PU}_{\text{tail}}, \text{ are nondecreasing in } B \text{ and in } \varepsilon. \quad (1.13)$$

Thus, we see the inequalities in (1.11) and (1.12) hold under the relaxed (and more

convenient) conditions

$$\sum_i \mathbb{P}(\zeta_i > y) = 0, \quad (\sum_i \mathbb{E} \zeta_i^2)^{1/2} \leq B, \quad \mathbb{E} S \leq m, \quad \text{and} \quad \frac{\sum_i \mathbb{E}(\zeta_i)_+^p}{B^2 y^{p-2}} \leq \varepsilon \in (0, 1]; \quad (1.14)$$

that (1.11) is true when $\varepsilon = 1$ is a result by Bennett [7] and Hoeffding [51], and we let $\text{BH}_{\text{exp}}(\lambda, y, B) := \text{PU}_{\text{exp}}(\lambda, y, B, 1)$ and $\text{BH}_{\text{tail}}(x, y, B, m) := \inf_{\lambda > 0} e^{-\lambda(x-m)} \text{BH}_{\text{exp}}(\lambda, y, B)$. The bounds PU_{exp} and PU_{tail} can be much less than BH_{exp} and BH_{tail} , respectively, when ε is significantly less than 1. Expressions for PU_{tail} are given in [120, Corollary 1] and [114, Proposition 3.1], and Lemma 1.A.1 will present these in a manner useful for the applications considered in the present paper. We remark here that an exponential bound on $\mathbb{E} e^{\lambda(S-m)}$ (and hence also $\mathbb{P}(S \geq x)$) which incorporates the moments $\mathbb{E}(\zeta_i)_+^p$ with $p > 3$ is stated in [120, Theorem 6], though the resulting expression is considerably more complicated in appearance than the bound in (1.11).

In the proof of Theorem 1.2.4 stated below, we shall also have cause to find a lower bound for the exponential moment of a Winsorized r.v. Particularly, suppose that ξ is a zero-mean r.v. with $\sqrt{\mathbb{E} \xi^2} \leq B$ for some $B \in (0, \infty)$. Then for any $c > 0$, [108, Theorem 2.1] states that

$$\mathbb{E} \exp\{c(1 \wedge \xi)\} \geq L_{W;c,B} := \frac{a_{c,B}^2 e^{c c_{c,B}} + B^2 e^{-c a_{c,B}}}{a_{c,B}^2 + B^2}, \quad (1.15)$$

where $a_{c,B}$ is the unique positive root of the function $a \mapsto \frac{a}{c} (2(e^{c+ac} - 1) - ac) - B^2$. In fact, as shown in [108], $L_{W;c,B}$ 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 $B \in (0, \infty)$.

Theorem 1.2.4. *Let Δ be any r.v. such that $|\Delta| \geq |T - W|$ a.s. For each $i = 1, \dots, n$, let Δ_i be any r.v. such that X_i and $(\Delta_i, (X_j : j \neq i))$ are independent, and assume that the mentioned Borel-measurable functions g_i and h_i are such that $g_i \leq h_i$, so that $\xi_i \leq \eta_i$. Take any real $p \in [2, 3]$ and let $q := \frac{p}{p-1}$, so that $\frac{1}{p} + \frac{1}{q} = 1$; also take any real numbers*

$$c_* \in (0, 1), \quad \theta > 0, \quad w > 0, \quad \delta_0 \in (0, w], \quad \pi_1 > 0, \quad \pi_2 > 0, \quad \text{and} \quad \pi_3 > 0 \quad \text{such that} \quad \pi_1 + \pi_2 + \pi_3 = 1. \quad (1.16)$$

Then for all $z \geq 0$

$$|\hat{\mathbb{P}}(T > z) - \hat{\mathbb{P}}(W > z)| \leq \gamma_z + \tau e^{-(1-\pi_1)z/\theta}, \quad (1.17)$$

where

$$\hat{\mathbb{P}}(E) := \mathbb{P}(E \cap \{|\Delta| \leq \pi_1 z\}) \quad \text{for any event } E, \quad (1.18)$$

$$\gamma_z := G_\xi(\pi_2 z) + \sum_{i=1}^n \mathbb{P}(W - \xi_i \geq \pi_3 z) \mathbb{P}(\eta_i > w), \quad (1.19)$$

$$\tau := c_1 \sum_{i=1}^n \|\xi_i\|_p \|\overline{\Delta} - \Delta_i\|_q + c_2 \|\overline{\Delta}\|_q + c_3 \delta, \quad (1.20)$$

G_ξ is defined by (1.9), $\overline{\Delta}$ is any r.v. satisfying (1.5), δ is any number such that (1.4) holds,

$$c_1 := \frac{1}{c_*} \text{PU}_{\text{exp}}\left(\frac{p}{\theta}, w, \frac{1}{\sqrt{p}}, \varepsilon_1\right) e^{\delta_0/\theta}, \quad (1.21)$$

$$c_2 := c_1 \left(\mathfrak{A}_{\mathbb{R}}(p) (1.32 a_1 e^{pw/\theta})^{1/p} \sigma_p + \mathfrak{B}_{\mathbb{R}}(p) (a_1 e^{pw/\theta})^{1/2} + (e^{pw/\theta} - 1)/w \right), \quad (1.22)$$

$$c_3 := \left(2c_2 + \frac{1}{c_*} \sqrt{2} \text{PU}_{\text{exp}}\left(\frac{2}{\theta}, w, \frac{1}{\sqrt{2}}, \varepsilon_1\right)\right) \vee \left(\frac{1}{\delta_0} \text{PU}_{\text{exp}}\left(\frac{1}{\theta}, w, 1, \varepsilon_1\right)\right) \quad (1.23)$$

$$\varepsilon_1 := \frac{\sigma_p^p}{w^{p-2}} \wedge 1, \quad (1.24)$$

$$a_1 := 1/L_{W; pw/\theta, \max_i \|\xi_i\|_2/w}. \quad (1.25)$$

Remark 1.2.5. We shall use (1.17) in conjunction with the obvious inequality

$$|\mathbf{P}(T > z) - \mathbf{P}(W > z)| \leq \mathbf{P}(|\Delta| > \pi_1 |z|) + |\hat{\mathbf{P}}(T > z) - \hat{\mathbf{P}}(W > z)|.$$

Thus, the use of the measure $\hat{\mathbf{P}}$ in (1.17) will allow us to avoid a “double counting” of the probability $\mathbf{P}(|\Delta| > \pi_1 |z|)$ when Theorem 1.2.4 is used to obtain Theorem 1.3.5.

Remark 1.2.6. The bound (1.17) (as well as other nonuniform bounds presented later in this paper) is stated only for $z \geq 0$, which allows for one-tail expressions $G_\xi(\pi_2 z)$ and $\mathbf{P}(W - \xi_i \geq \pi_3 z)$ to be used in (1.19). In order to obtain the corresponding bound for $z < 0$, all that is needed is to replace T and g_i with $-T$ and $-g_i$, respectively, where the g_i ’s are as in (1.1).

Remark 1.2.7. It is easy to see that the expressions c_1 , c_2 , and c_3 in (1.21)–(1.23) can be bounded by finite positive constants depending only on the values of the parameters p , c_* , θ , w , and δ_0 (and not on the distributions of the X_i ’s). This follows because PU_{exp} is nondecreasing in ε (recall (1.13)) and $a_1 \leq 1/L_{W; pw/\theta, 1/w}$ (since $\max_i \|\xi_i\|_2 \leq \|W\|_2 = 1$ and $L_{W; c, B}$ is nonincreasing in B). Thus, one may refer to c_1 , c_2 , and c_3 as *pre-constants*.

Remark 1.2.8. If we add the assumption that the ξ_i ’s are all symmetric(ally distributed) to the hypotheses of Theorem 1.2.4, then, according to the main result of [113], $(e^{pw/\theta} - 1)/w$ in (1.22) may be replaced by the smaller quantity $\sinh(pw/\theta)/w$. This sharpening of the inequality (1.17) allows for smaller absolute constants to be obtained in applications of Theorem 1.2.4; cf. the nonuniform bound for the self-normalized sum in Corollary 1.4.12 and Remark 1.4.16.

For $p = 2$, the result of Theorem 1.2.4 is similar to that by Chen and Shao [19, Theorem 2.2]. The bound given by (1.17) turns out to be more precise in the applications given in this paper. In particular, it allows one to weaken conditions on moments. Indeed, in Theorem 1.3.5 one will have $|\overline{\Delta}|$ on the order of $\|S\|^2$ and $|\overline{\Delta} - \Delta_i|$ on the order of $\|X_i\|^2 + \|X_i\| \|S - X_i\|$, where $S := \sum_{i=1}^n X_i$ and the X_i ’s are independent random vectors. So, using Theorem 1.2.4 with $p = 3$ (and hence $q = \frac{3}{2}$) in order to obtain a bound of the classical form $\mathcal{O}(\frac{1}{\sqrt{n(|z|+1)^3}})$, one will need only the third moments of $\|X_i\|$ to be finite. On the other hand, using (1.20) with $p = 2$ to get the same kind of bound would require the finiteness of the fourth moments of $\|X_i\|$.

Expressions in Theorem 1.2.4 are complicated, especially the ones for c_1 , c_2 , and c_3 . However, this may be considered as just another instance of the usual trade-off between accuracy and complexity of the bounds.

Bounds (1.6) and (1.17) 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 [97, Chapter V] or

the paper [109]. For the linear statistic W as in (1.3) with i.i.d. ξ_1, \dots, ξ_n as in (1.2), results due to Shevtsova [136] and Michel [81] imply

$$|\mathbf{P}(W \leq z) - \Phi(z)| \leq (0.33554n\|\xi_1\|_3^3 + 0.13925/\sqrt{n}) \wedge \left(\frac{30.2211}{|z|^3 + 1} n\|\xi_1\|_3^3 \right). \quad (1.26)$$

1.3 Berry-Esseen bounds for smooth nonlinear functions of sums of independent random vectors

In this section, we shall state applications of results of Section 1.2. Assume from hereon that $(\mathfrak{X}, \|\cdot\|)$ is a separable Banach space of type 2; for a definition and properties of such spaces, see e.g. [53, 118]. Let X_1, \dots, X_n be independent random vectors in \mathfrak{X} with $\mathbf{E} X_i = 0$ for $i = 1, \dots, n$, and also let

$$\begin{aligned} S &:= \sum_{i=1}^n X_i, \\ \|X\|_p &:= \mathbf{E}^{1/p} \|X\|^p, \\ s_p &:= \left(\sum_{i=1}^n \|X_i\|_p^p \right)^{1/p} = \left(\sum_{i=1}^n \mathbf{E} \|X_i\|^p \right)^{1/p}, \end{aligned} \quad (1.27)$$

$$G_X(z) := \sum_{i=1}^n \mathbf{P}(\|X_i\| > z), \quad (1.28)$$

for any $p \geq 1$ and $z \geq 0$; compare (1.27) and (1.28) to (1.8) and (1.9), respectively.

Note that the results of [119, Theorem 1] (see also the remark in [120, p. 343]) may be used to derive bounds analogous to those given in (1.11) and (1.12) when the ζ_i 's take values in a separable Banach space. Particularly,

$$(1.12) \text{ and } (1.11) \text{ hold under (1.14) when } S \text{ and } \zeta_i \text{ are replaced by } \|S\| \text{ and } \|X_i\|, \quad (1.29)$$

respectively.

Since \mathfrak{X} is of type 2 and the X_i 's are zero-mean, there exists a constant $D := D(\mathfrak{X}) \in (0, \infty)$ such that

$$\|S\|_2 \leq D s_2. \quad (1.30)$$

We shall assume that D is chosen to be minimal with respect to this property; so, $D = 1$ with the equality in (1.30) whenever \mathfrak{X} is a Hilbert space. By [118, Theorem 2], one also has the Rosenthal-type inequality

$$\|S\|_\alpha^\alpha \leq \mathfrak{A}_\mathfrak{X}^\alpha(\alpha) s_\alpha^\alpha + \mathfrak{B}_\mathfrak{X}^\alpha(\alpha) s_2^\alpha \quad (1.31)$$

for any $\alpha \geq 2$ and some pair of constants $(\mathfrak{A}_\mathfrak{X}(\alpha), \mathfrak{B}_\mathfrak{X}(\alpha))$; note that (1.31) generalizes (1.10). In the particular case when \mathfrak{X} is a Hilbert space and $\alpha = 3$, [117, Theorem 1] allows us to use the values

$$\mathfrak{A}_\mathfrak{X}(3) = 1 \quad \text{and} \quad \mathfrak{B}_\mathfrak{X}(3) = 3^{1/3}. \quad (1.32)$$

Remark 1.3.1. The results of this section hold for vector martingales taking values in a 2-smooth separable Banach space; in such a case, one can apply results of [99] instead of the ones of [119] used in the present paper. By [53, 99], every 2-smooth Banach space is

of type 2. It is known that L^p spaces are 2-smooth, and hence of type 2, for all $p \geq 2$ [99, Proposition 2.1].

Let next $f: \mathfrak{X} \rightarrow \mathbb{R}$ be a Borel-measurable functional with $f(0) = 0$, satisfying the following smoothness condition: there exist $\epsilon > 0$, $M_\epsilon > 0$, and a nonzero continuous linear functional $L: \mathfrak{X} \rightarrow \mathbb{R}$ such that

$$|f(x) - L(x)| \leq \frac{M_\epsilon}{2} \|x\|^2 \text{ for all } x \in \mathfrak{X} \text{ with } \|x\| \leq \epsilon; \quad (1.33)$$

thus, L necessarily coincides with the first Fréchet derivative, $f'(0)$, of the function f at 0. Moreover, for the smoothness condition (1.33) 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 $\|x\| \leq \epsilon$.

The following bound for the distribution of $f(S)$ may still look rather abstract and complicated. However, especially in such applications to specific statistics as the ones presented in Corollaries 1.4.10 and 1.4.20, it leads to comparatively simple BE type bounds of a “correct” order of magnitude and with explicit numerical constants of rather moderate sizes.

Theorem 1.3.2. *Let $f: \mathfrak{X} \rightarrow \mathbb{R}$ satisfy (1.33), let X_1, \dots, X_n be independent zero-mean random vectors in \mathfrak{X} , and assume that*

$$\sigma := \|L(S)\|_2 = \left(\sum_i \|L(X_i)\|_2^2 \right)^{1/2} \in (0, \infty). \quad (1.34)$$

Further, take any $p \in (2, 3]$, $c_ \in (0, 1)$, $w > 0$, and let $q := \frac{p}{p-1}$, so that $\frac{1}{p} + \frac{1}{q} = 1$. Then for all $z \in \mathbb{R}$*

$$\begin{aligned} & \left| \mathbb{P}\left(\frac{f(S)}{\sigma} > z\right) - \mathbb{P}\left(\frac{L(S)}{\sigma} > z\right) \right| \\ & \leq \mathbb{P}(\|S\| > \epsilon) + \frac{4c_4\sigma_p^{\tilde{q}} + (\mathfrak{A}_{\mathbb{R}}(p)\sigma_p + \mathfrak{B}_{\mathbb{R}}(p))\mathbf{u} + \sigma_p\mathbf{v}}{2c_*} + G_\eta(w), \end{aligned} \quad (1.35)$$

where

$$c_4 := \left(\frac{(p-2)^{p-2}}{(p-1)^{p-1}(1-c_*)} \right)^{1/(p-2)}, \quad \tilde{q} := \frac{p}{p-2}, \quad (1.36)$$

σ_p and G_η are as in (1.8) and (1.9) with

$$\xi_i = \frac{L(X_i)}{\sigma}, \quad \eta_i = \frac{\|L\| \|X_i\|}{\sigma} \mathbf{1}_{\{2 < p < 3\}}, \quad (1.37)$$

$$\mathbf{u} := \frac{M_\epsilon \sigma}{2\|L\|^2} \times \begin{cases} (\mathfrak{A}_{\mathfrak{X}}^2(3)\lambda_3^2 + \mathfrak{B}_{\mathfrak{X}}^2(3)\lambda_2^2) & \text{if } p = 3, \\ 5w^2(\mathfrak{A}_{\mathfrak{X}}^2(2q)\lambda_p^{p-1} + \mathfrak{B}_{\mathfrak{X}}^2(2q)\lambda_2^2 + \lambda_p^{2p}) & \text{if } p \in (2, 3), \end{cases} \quad (1.38)$$

$$\mathbf{v} := \frac{M_\epsilon \sigma}{2\|L\|^2} \times \begin{cases} (\lambda_3^2 + 2D\lambda_2\lambda_{3/2}) & \text{if } p = 3, \\ w^2(\lambda_p^{p-1} + 4D\lambda_2\lambda_q + 2\lambda_q\lambda_p^p) & \text{if } p \in (2, 3), \end{cases} \quad (1.39)$$

$$\lambda_\alpha := \|L\| \frac{s_\alpha}{\sigma} \times \begin{cases} 1 & \text{if } p = 3, \\ w^{-1} & \text{if } p \in (2, 3). \end{cases} \quad (1.40)$$

Remark 1.3.3. The term $\mathbf{P}(\|S\| > \epsilon)$ in (1.35) can be bounded in a variety of ways. For instance, using Chebyshev's inequality and (1.30), one can write

$$\mathbf{P}(\|S\| > \epsilon) \leq \frac{\|S\|_2^2}{\epsilon^2} \leq \frac{D^2 s_2^2}{\epsilon^2}. \quad (1.41)$$

Alternatively, one can write

$$\mathbf{P}(\|S\| > \epsilon) \leq \frac{\|S\|_p^p}{\epsilon^p} \leq \frac{\mathfrak{A}_{\mathfrak{X}}^p(p) s_p^p + \mathfrak{B}_{\mathfrak{X}}^p(p) s_2^p}{\epsilon^p},$$

using a Rosenthal-type inequality (1.31). An exponential inequality as described in (1.29) can also be used.

Remark 1.3.4. The expressions \mathfrak{u} and \mathfrak{v} in (1.38) and (1.39) are finite for any given $p \in (2, 3]$ whenever $s_p < \infty$, whereas λ_{2q} may be infinite for $p \in (2, 3)$ even when the condition $s_p < \infty$ holds. It is the additional truncation, with $\overline{\Delta}$ instead of Δ , in the bounds of Section 1.2 that allows one to use λ_p instead of λ_{2q} in the terms \mathfrak{u} and \mathfrak{v} when $p < 3$; cf. Remark 1.2.3.

The hardest to obtain result of this section is the nonuniform bound in Theorem 1.3.5 below.

Theorem 1.3.5. *Assume that the conditions of Theorem 1.3.2 are satisfied, and take any real numbers θ , δ_0 , π_1 , π_2 , π_3 , and ω such that the conditions (1.16) hold and*

$$\omega \in \left(0, \frac{M_\epsilon \epsilon^2}{2\pi_1}\right]. \quad (1.42)$$

Let

$$\xi_i := \frac{L(X_i)}{\sigma}, \quad \text{and} \quad \eta_i := \frac{\|L\| \|X_i\|}{\sigma} \mathbf{I}\{2 < p < 3\} + \frac{L(X_i)}{\sigma} \mathbf{I}\{p = 3\}. \quad (1.43)$$

Then for all

$$z \in (0, \omega/\sigma] \quad (1.44)$$

one has

$$\left| \mathbf{P}\left(\frac{f(S)}{\sigma} > z\right) - \mathbf{P}\left(\frac{L(S)}{\sigma} > z\right) \right| \leq \tilde{\gamma}_z + \tilde{\tau} e^{-(1-\pi_1)z/\theta}, \quad (1.45)$$

where

$$\tilde{\gamma}_z := \mathbf{P}\left(\|S\| > \sqrt{\frac{2\pi_1 \sigma z}{M_\epsilon}}\right) + \gamma_z, \quad (1.46)$$

$$\tilde{\tau} := c_1 \sigma_p \mathfrak{v} + c_2 \mathfrak{u} + c_3 c_4 \sigma_p^{\tilde{q}}, \quad (1.47)$$

and γ_z , c_1 , c_2 , c_3 are as in Theorem 1.2.4.

Remark 1.3.6. The restriction (1.44) is of essence. Indeed, if $z \gg \frac{1}{\sigma}$ (that is, if z is much greater than $\frac{1}{\sigma}$) and the event $\{\frac{L(S)}{\sigma} > z\}$ in (1.45) occurs, then $L(S) \gg 1$ 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; cf. e.g. (1.113), in which $\sigma\Delta$, measuring the difference between $\sigma T = f(S)$ and $\sigma W = L(S)$, is on the order of magnitude of $\|S\|^2$ and thus much greater than $L(S)$ when $\|S\| \gg 1$. This heuristics will be implicitly used in Proposition 1.B.1 in Appendix 1.B, which shows that the upper bound $\frac{\omega}{\sigma}$ on z in (1.44) is indeed the best possible up to a constant factor, even when the Banach space \mathfrak{X} is one-dimensional. Note also that (1.42) can be satisfied for any given $\omega \in (0, \infty)$ by (say) taking π_1 to be small enough.

While the expressions for the upper bounds given in Theorems 1.3.2 and 1.3.5 are quite explicit, they may seem complicated (as compared with the classical uniform and nonuniform BE bounds). However, one should realize that here there are a whole host of players: those associated with the function f and the space \mathfrak{X} (like $\|L\|$, M_ϵ , ϵ , and D), the parameters we are free to choose (namely, c_* , θ , w , δ_0 , π_1 , π_2 , π_3 , and ω), and more traditional terms (as s_p , σ , and G_ξ) – each with a significant and rather circumscribed role to play.

One should note that the bounds in Theorems 1.3.2 and 1.3.5 do not depend on the dimension of the space \mathfrak{X} but only on the choice of the norm $\|\cdot\|$ on \mathfrak{X} . One can exercise this choice to an advantage, as e.g. will be done in the application considered in Section 1.4.1. The only restriction on the norm is that the space \mathfrak{X} (possibly even infinite-dimensional) be of type 2; in particular, the bounds will depend on the “smoothness” constant D for the norm and on the corresponding Rosenthal-type inequality constants $(\mathfrak{A}_\mathfrak{X}(\cdot), \mathfrak{B}_\mathfrak{X}(\cdot))$.

Another advantage of the bounds in (1.35) and (1.45) 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 extra zero summands X_i). In fact, (1.35) and (1.45) remain valid when S is the sum of an infinite series of independent zero-mean r.v.’s, i.e. $S = \sum_{i=1}^\infty X_i$, provided that the series converges in an appropriate sense; see e.g. Jain and Marcus [60].

On the other hand, for i.i.d. r.v.’s 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 $\mathbb{E} V = 0$. Here we shall use

$$\overline{V} := \frac{1}{n} \sum_{i=1}^n V_i$$

in place of S (and hence $\frac{1}{n}V_i$ in place of X_i).

Corollary 1.3.7. *Take any $p \in (2, 3]$. Suppose that (1.33) holds,*

$$\tilde{\sigma} := \|L(V)\|_2 > 0,$$

and $\|V\|_p < \infty$. Then for all $z \in \mathbb{R}$

$$\left| \mathbb{P}\left(\frac{f(\overline{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{n^{p/2-1}}; \quad (1.48)$$

moreover, for any $\omega \in (0, \infty)$, $\tilde{\theta} \in (0, \infty)$, and for all

$$z \in \left(0, \frac{\omega}{\tilde{\sigma}} \sqrt{n}\right] \quad (1.49)$$

one has

$$\begin{aligned} & \left| \mathbb{P}\left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \\ & \leq \mathfrak{C} \left(n \mathbb{P}(\|V\| > \mathfrak{C}\sqrt{n}z) + \frac{n \mathbb{P}(\|V\| > \mathfrak{C}\sqrt{n})}{z^p} + \frac{1}{(\sqrt{n}z)^p} + \frac{1}{n^{p/2-1}e^{z/\tilde{\theta}}} \right). \end{aligned} \quad (1.50)$$

Each instance of \mathfrak{C} above is a finite expression that depends only upon p , the space \mathfrak{X} (through the constants D in (1.30) and $(\mathfrak{A}_{\mathfrak{X}}(\cdot), \mathfrak{B}_{\mathfrak{X}}(\cdot))$ in (1.31)), the function f (through (1.33)), the moments $\tilde{\sigma}$, $\|L(V)\|_p$, $\|V\|_q$, $\|V\|_2$, and $\|V\|_p$, with \mathfrak{C} in (1.48) also depending on ω and $\tilde{\theta}$. Also, (1.48) and (1.50) both hold when $\mathbb{P}(\sqrt{n}L(\bar{V})/\tilde{\sigma} \leq z)$ replaces $\Phi(z)$.

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 (in \mathbb{R}) that is needed or most convenient; in fact, as mentioned before, 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 the standard 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 [44] provided an upper bound of the order $\mathcal{O}(1/\sqrt{n})$ 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). Bhattacharya and Holmes [13] obtained a constant which is $\mathcal{O}(k^{5/2})$, and Chen and Fang [17, Theorem 3.5] recently improved this to $\mathcal{O}(k^{1/2})$.

One should also note here such results as the ones obtained by Götze [43] (uniform bounds) and Zaleskiĭ [147, 148] (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 [9] 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.

1.4 Applications

Here we shall apply the results of Section 1.3 to present several novel bounds on the rate of convergence to normality for some commonly used statistics. For the sake of simplicity and brevity, assume throughout this section that

$$p = 3$$

and V, V_1, \dots, V_n are i.i.d. \mathfrak{X} -valued r.v.'s, where \mathfrak{X} is a Hilbert space; also adopt the notation

$$\tilde{\sigma} := \|L(V)\|_2, \quad \varsigma_\alpha := \frac{\|L(V)\|_\alpha}{\tilde{\sigma}}, \quad \text{and} \quad v_\alpha := \|V\|_\alpha \quad \text{for } \alpha \geq 1, \quad (1.51)$$

where L is as in (1.33).

Essentially two types of results will be presented in this section. Theorems 1.4.5, 1.4.17, and 1.4.24, containing uniform and nonuniform BE-type bounds for specific statistics (namely, Student's, Pearson's, and noncentral Hotelling's) are straightforward applications of Corollary 1.3.7, in each specific instance with its own space \mathfrak{X} , function f , and random vector V . Of course, these results inherit from Corollary 1.3.7 the not quite explicit constants \mathfrak{C} , which, recall, were finite expressions depending only upon p , the function f , and the distribution of V , with \mathfrak{C} in the nonuniform bounds also depending on ω ; however, in contrast with Corollary 1.3.7, the \mathfrak{C} 's in Theorems 1.4.5, 1.4.17, and 1.4.24 will no longer depend on the space \mathfrak{X} , since one can use the same constants D in (1.30) and $(\mathfrak{A}_\mathfrak{X}(\cdot), \mathfrak{B}_\mathfrak{X}(\cdot))$ in (1.31) for all Hilbert spaces \mathfrak{X} .

On the other hand, Theorem 1.4.1 will provide a uniform BE-type bound for a normalized statistic $\sqrt{n}f(\bar{V})/\tilde{\sigma}$, with explicit coefficients on each of the terms in the bound. These coefficients, denoted by \mathfrak{K} with two or three subscripts, will in specific applications be variously bounded from above by finite explicit constants which do not depend on n or z ; so, such coefficients may be referred to as *pre-constants*. The corresponding nonuniform bound is much more complicated and therefore will be relegated to Appendix 1.A, where it is stated (and proved) as Theorem 1.A.2. To help the reader follow our indexing of the pre-constants, let us say that the subscript of a pre-constant \mathfrak{K} will be u or e or n, depending on whether the pre-constant appears in a uniform BE-type bound or in an exponentially (in z) decreasing term of a nonuniform BE-type bound or in a power-like decreasing term of a nonuniform BE-type bound, respectively; the remaining subscripts refer to the moments of which the pre-constant is a coefficient.

We then apply the inequalities of Theorems 1.4.1 and 1.A.2 to obtain BE-type bounds for the self-normalized sum and Pearson's correlation coefficient containing only absolute constants and moments of relevant r.v.'s, with a simple (and optimal) dependence on n and z ; these latter bounds are given in Corollaries 1.4.11, 1.4.12, and 1.4.20. The proofs of these three corollaries are somewhat lengthy and technical, and so are placed in Appendix 1.C.

Theorem 1.4.1. *Let \mathfrak{X} be a Hilbert space, let f satisfy (1.33) for some real $\epsilon > 0$, and assume that $\mathbb{E} V = 0$, $\tilde{\sigma} > 0$, and $v_3 < \infty$. Take any real numbers*

$$c_* \in (0, 1), \quad \kappa_{2,0} > 0, \quad \kappa_{3,0} > 0, \quad \kappa_{2,1} > 0, \quad \text{and} \quad \kappa_{3,1} > 0. \quad (1.52)$$

Then

$$\begin{aligned} & \left| \mathbb{P}\left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \\ & \leq \frac{\mathfrak{K}_{u0} + \mathfrak{K}_{u1}\varsigma_3^3 + (\mathfrak{K}_{u20} + \mathfrak{K}_{u21}\varsigma_3)v_2^2 + (\mathfrak{K}_{u30} + \mathfrak{K}_{u31}\varsigma_3)v_3^2 + \mathfrak{K}_{ue}}{\sqrt{n}} \end{aligned} \quad (1.53)$$

$$\leq \frac{\tilde{\mathfrak{K}}_{u0} + \tilde{\mathfrak{K}}_{u1}\varsigma_3^3 + \tilde{\mathfrak{K}}_{u2}v_2^2 + \tilde{\mathfrak{K}}_{u3}v_3^2}{\sqrt{n}} \quad (1.54)$$

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

$$\begin{aligned} \mathfrak{K}_{u0} &:= 0.13925, \quad \mathfrak{K}_{u1} := 0.33554 + \frac{1}{2c_*(1-c_*)}, \\ (\mathfrak{K}_{u20}, \mathfrak{K}_{u21}, \mathfrak{K}_{u30}, \mathfrak{K}_{u31}) &:= \frac{M_\epsilon}{4c_*\tilde{\sigma}} \left(3, 2 + \frac{3^{2/3}}{n^{1/6}}, \frac{3^{1/3}}{n^{1/3}}, \frac{2}{n^{1/2}} \right), \end{aligned} \quad (1.55)$$

$$\mathfrak{K}_{ue} := \frac{v_2^2}{\epsilon^2 n^{1/2}} \bigwedge \frac{3v_2^3 + v_3^3/n^{1/2}}{\epsilon^3 n}, \quad (1.56)$$

$$\begin{aligned} \tilde{\mathfrak{K}}_{u0} &:= \mathfrak{K}_{u0} + \frac{1}{3\kappa_{2,0}^3} \left(\mathfrak{K}_{u20} + \frac{1}{\epsilon^2 n^{1/2}} \right) + \frac{1}{3\kappa_{3,0}^3} \mathfrak{K}_{u30}, \quad \tilde{\mathfrak{K}}_{u1} := \mathfrak{K}_{u1} + \frac{1}{3\kappa_{2,1}^3} \mathfrak{K}_{u21} + \frac{1}{3\kappa_{3,1}^3} \mathfrak{K}_{u31}, \\ \tilde{\mathfrak{K}}_{u2} &:= \frac{2\kappa_{2,0}^{3/2}}{3} \left(\mathfrak{K}_{u20} + \frac{1}{\epsilon^2 n^{1/2}} \right) + \frac{2\kappa_{2,1}^{3/2}}{3} \mathfrak{K}_{u21}, \quad \tilde{\mathfrak{K}}_{u3} := \frac{2\kappa_{3,0}^{3/2}}{3} \mathfrak{K}_{u30} + \frac{2\kappa_{3,1}^{3/2}}{3} \mathfrak{K}_{u31}. \end{aligned} \quad (1.57)$$

Remark 1.4.2. Under the additional assumption that $L(V)$ is symmetric, Theorem 1.4.1 remains true if the quadruple $(\mathfrak{K}_{u20}, \mathfrak{K}_{u21}, \mathfrak{K}_{u30}, \mathfrak{K}_{u31})$ is replaced by the following one, with smaller values:

$$\frac{M_\epsilon}{4c_*\tilde{\sigma}} \left(3^{2/3} C_3^*, 2, \frac{C_3^*}{n^{1/3}}, \frac{1}{n^{1/2}} \right), \quad \text{where } C_3^* := \left(1 + \sqrt{\frac{8}{\pi}} \right)^{1/3}. \quad (1.58)$$

Indeed, [57, Theorem 1] implies $\|W\|_3 \leq C_3^*(1 \vee \sigma_3)$ (recall here (1.3), (1.2), and (1.8)). Since $\mathfrak{K}_{u1} > 1$ for any $c_* \in (0, 1)$, we may assume w.l.o.g. that $\sigma_3^3 = \varsigma_3^3/\sqrt{n} < 1$ and hence $\|W\|_3 \leq C_3^*$. So, following the lines of the proof of Theorem 1.3.2, we see that the upper bound $\mathfrak{A}_{\mathbb{R}}(p)\sigma_p + \mathfrak{B}_{\mathbb{R}}(p)$ on $\|W\|_p$, which appears in (1.35), can be replaced there by C_3^* .

Remark 1.4.3. One can have a “nonuniform” counterpart to Theorem 1.4.1. Indeed, assume that the conditions of Theorem 1.4.1 take place; in particular, let ϵ and M_ϵ be any positive real numbers such that (1.33) holds. Take any positive real numbers z_0 , $\tilde{\theta}$, K_1 , K_2 , and K_3 . Then there exist some finite positive constants \mathfrak{C}_{n1} , \mathfrak{C}_{n21} , \mathfrak{C}_{n22} , \mathfrak{C}_{n31} , \mathfrak{C}_{n32} , \mathfrak{C}_{e0} , \mathfrak{C}_{e1} , \mathfrak{C}_{e2} , and \mathfrak{C}_{e3} , each depending only on ϵ , M_ϵ , z_0 , $\tilde{\theta}$, K_1 , K_2 , and K_3 , such that

$$\begin{aligned} & \left| \mathbb{P} \left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \\ & \leq \frac{\mathfrak{C}_{n1}\varsigma_3^3 + ((\mathfrak{C}_{n21} \vee \mathfrak{C}_{n22})v_2^4) \vee (\mathfrak{C}_{n31}v_3^3) + \mathfrak{C}_{n32}v_3^3}{z^3\sqrt{n}} + \frac{\mathfrak{C}_{e0} + \mathfrak{C}_{e1}\varsigma_3^3 + \mathfrak{C}_{e2}v_2^3 + \mathfrak{C}_{e3}v_3^3}{e^{z/\tilde{\theta}}\sqrt{n}} \end{aligned} \quad (1.59)$$

for all $z \in \mathbb{R}$ and $n \in \mathbb{N}$ such that

$$z_0 \leq z \leq \frac{\omega}{\tilde{\sigma}} \sqrt{n}, \quad \frac{K_1\varsigma_3^3}{\sqrt{n}} \leq 1, \quad \frac{K_2v_2^4}{\tilde{\sigma}^3 z^3 \sqrt{n}} \leq 1, \quad \text{and} \quad \frac{K_3v_3^3}{\tilde{\sigma}^3 z^3 \sqrt{n}} \leq 1. \quad (1.60)$$

The constants $\mathfrak{C}_{n1}, \dots, \mathfrak{C}_{e3}$ in (1.59) are upper bounds on certain corresponding pre-constants $\mathfrak{K}_{n1}, \dots, \mathfrak{K}_{e3}$, explicit expressions for which are given in Theorem 1.A.2. Concerning the conditions in (1.60), note the following:

1. The condition $z \geq z_0$ does not diminish generality, in view of uniform bounds (1.53)

and (1.54).

2. The condition $z \leq \frac{\omega}{\sigma} \sqrt{n}$ is essential and even optimal, up to a constant factor, as shown in Appendix 1.B.
3. The other three conditions in (1.60), involving the constants K_1 , K_2 , and K_3 , will be satisfied when n and z are large enough. As mentioned above, the case when z is not large can be covered using a uniform bound. Finally, the remaining case with “large” z and “small” n can be dealt with based on an appropriate upper bound on large deviation probabilities. In fact, the proof (given in Appendix 1.C) of the nonuniform bound in Corollary 1.4.12 is conducted right along such lines.

The mentioned pre-constants in Theorems 1.4.1 and 1.A.2 are complicated in appearance. However, in particular applications – presented in Corollaries 1.4.11, 1.4.12, and 1.4.20 – these statements will result in bounds 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 1.4.1 and 1.A.2 is quite simple as well, and the corresponding constants are again moderate in size.

Corollary 1.4.4. *Assume that the conditions of Theorem 1.4.1 hold, and also that f'' is twice continuously differentiable in a neighborhood of the origin. Then*

$$\limsup_{n \rightarrow \infty} \sup_{z \in \mathbb{R}} \sqrt{n} \left| \mathbb{P} \left(\frac{f(\bar{V})}{\bar{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \leq 0.13925 + 0.83554 \varsigma_3^3 + \frac{y_*}{4} + \frac{1}{2} \sqrt{\varsigma_3^3 (\varsigma_3^3 + y_*)}, \quad (1.61)$$

where

$$y_* := \frac{\|f''(0)\|}{\bar{\sigma}} (3 + 2\varsigma_3) v_2^2. \quad (1.62)$$

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

$$\limsup_{n \rightarrow \infty} \sup_{g(n) \leq z \leq \sqrt{n}/g(n)} z^3 \sqrt{n} \left| \mathbb{P} \left(\frac{f(\bar{V})}{\bar{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \leq 30.2211 \varsigma_3^3. \quad (1.63)$$

As one can see, in the expressions of the asymptotic uniform bounds in (1.61) the higher moment v_3 disappears, and in the asymptotic nonuniform bound in (1.63) the moment v_2 disappears as well; however, Corollary 1.4.4 inherits the condition $v_3 < \infty$ from Theorems 1.4.1 and 1.A.2 – where, as seen from Remark 1.4.7 and 1.4.19, this condition is essential; cf. also Remark 1.4.13.

For the remainder of the results in this section, \mathfrak{X} will be the Euclidean space \mathbb{R}^k for some natural number k , and the nonlinear functional $f: \mathfrak{X} \rightarrow \mathbb{R}$ will be continuously twice differentiable in some neighborhood about the origin. Thus, for a given (small enough) ϵ , the smoothness condition (1.33) will hold when

$$L = f'(0) \quad \text{and} \quad M_\epsilon = \sup_{\|\mathbf{x}\| \leq \epsilon} \|f''(\mathbf{x})\|, \quad (1.64)$$

where $f'(\mathbf{x})$ and $f''(\mathbf{x})$ are identified with the gradient vector and the Hessian matrix, respectively, of f at some point $\mathbf{x} \in \mathfrak{X}$, and then $\|f''(\mathbf{x})\|$ denotes the spectral norm of

the matrix $f''(\mathbf{x})$. Upon specifying the function f and the relevant r.v. V , the results of Theorems 1.4.5, 1.4.17, and 1.4.24 (uniform and nonuniform bounds without explicit coefficients) will be proved by invoking the results of Corollary 1.3.7.

1.4.1 “Quadratic” statistic

The first application we consider involves a particularly simple nonlinear statistic investigated by Novak in [94, Section 3]. Let $V = (Y, Z)$, $V_1 = (Y_1, Z_1), \dots, V_n = (Y_n, Z_n)$ be i.i.d. r.v.’s with $\mathbb{E} V = 0$, $\mathbb{E} Y^2 = \mathbb{E} Z^2 = 1$. Take any real $\theta > 0$ and let \mathfrak{X} be \mathbb{R}^2 with the norm defined by the formula $\|x\| := \sqrt{x_1^2 + x_2^2/\theta^2}$ for $x = (x_1, x_2) \in \mathfrak{X}$. Next, take any real $c_0 \geq 0$ and let $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ be defined by $f(x_1, x_2) = x_1 + c_0 x_2^2$. Then f satisfies the smoothness condition (1.33) with $L(x_1, x_2) = x_1$ and $M_\epsilon = 2c_0\theta^2 = \|f''(0)\|$, for any $\epsilon > 0$. Consider the statistic

$$Q := \bar{Y} + c_0 \bar{Z}^2 = f(\bar{V}) \quad \text{with} \quad (\bar{Y}, \bar{Z}) = \bar{V} = \frac{1}{n} \sum_{i=1}^n V_i, \quad (1.65)$$

so that the statistic $\sqrt{n}Q = \sum_i (Y_i/\sqrt{n}) + c(\sum_i Z_i/\sqrt{n})^2$ with $c := c_0/\sqrt{n}$ coincides with the quadratic statistic studied in [94]; the X_i ’s and Y_i ’s in [94] are replaced here by Y_i/\sqrt{n} and Z_i/\sqrt{n} , respectively. One may also note that in [94] the condition $\mathbb{E} Z^2 = 1$ was not assumed; however, it can be assumed (as we do) without loss of generality, by adjusting the choice of the factor c_0 .

Now one can use the inequalities $\|Y\|_1 \leq \|Y\|_2 = 1$, $\|Z\|_1 \leq \|Z\|_2 = 1$, $\|YZ\|_1 \leq \|Y\|_2 \|Z\|_2 = 1$, $\|\sqrt{n}Z\|_1 \leq \|\sqrt{n}Z\|_{3/2} \leq \|\sqrt{n}Z\|_2 = 1$, and $\|\sqrt{n}Y\|_3 \leq \|Y\|_3/n^{1/6} + 3^{1/3}$ (cf. (1.10) and (1.32)) in conjunction with [94, Theorem 2] to obtain

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sqrt{n} |\mathbb{P}(\sqrt{n}Q \leq z) - \Phi(z)| &\leq 2 + \left(\frac{9}{\sqrt{2\pi}} + \sqrt{\frac{\pi}{8}} + 1 \right) \|Y\|_3^3 + \left(\sqrt{\frac{\pi}{2}} + 4 \right) c_0 \\ &< 2 + 5.218 \|Y\|_3^3 + 5.254 c_0. \end{aligned} \quad (1.66)$$

On the other hand, Corollary 1.4.4 implies

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sqrt{n} |\mathbb{P}(\sqrt{n}Q \leq z) - \Phi(z)| \\ \leq 0.13925 + 0.83554 \|Y\|_3^3 + \frac{\tilde{y}_*}{4} + \frac{1}{2} \sqrt{\|Y\|_3^3 (\|Y\|_3^3 + \tilde{y}_*)}, \end{aligned} \quad (1.67)$$

where

$$\tilde{y}_* = \inf_{\theta > 0} y_* = 2c_0(3 + 2\|Y\|_3).$$

It is not hard to see that (1.67) works better than (1.66) unless the “nonlinearity coefficient” c_0 in (1.65) is very large. In particular, for (1.66) to be better than (1.67) it is necessary that $\|Y\|_3 > 3.47$ and $c_0 > 149 + 2.67(\|Y\|_3 - 4.5)^2$. Note also that the asymptotic bound in (1.66) was obtained by removing the asymptotically negligible terms, whose contributions were positive, so that the resulting asymptotic bound is smaller than the corresponding non-asymptotic one. Now one can see that the conditions $\|Y\|_3 > 3.47$ and $c_0 > 149 + 2.67(\|Y\|_3 - 4.5)^2$, necessary for (1.66) to be better than (1.67), may obtain only for $n > 1.03 \times 10^6$ in order for the just mentioned non-asymptotic bound on $|\mathbb{P}(\sqrt{n}Q \leq z) - \Phi(z)|$ to

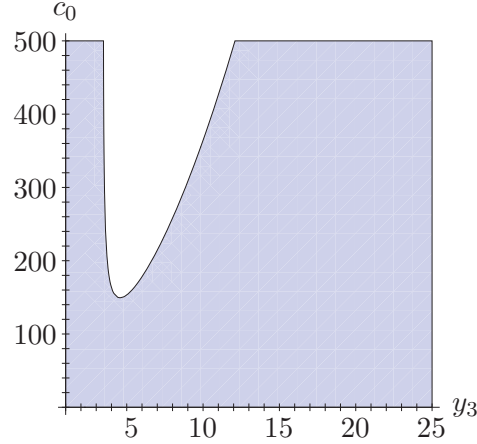


Figure 1.1: Comparison of asymptotic bounds; shaded region is where bound in (1.67) is less than bound in (1.66)

be nontrivial, that is, less than 1. Figure 1.1 shows the set (shaded) of all points $(y_3, c_0) \in [1, 25] \times [0, 500]$ with $y_3 := \|Y\|_3$ for which the asymptotic bound in (1.67) is less than that in (1.66). This and discussion in subsequent Subsubsection 1.4.2.1 suggest that bounds developed in this paper for general nonlinear statistics are competitive with bounds obtained earlier by specialized methods, tailored to a specific statistic or a specific class of statistics.

1.4.2 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

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

let T take an arbitrary value 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 “non-central” 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 $f(\mathbf{x}) := t_0/\sqrt{n} - \mu$ for all other $\mathbf{x} \in \mathfrak{X}$, where t_0 is the “exceptional” value chosen above for T . Since

$$\min_{x_1^2 + x_2^2 \leq \epsilon^2} (1 + x_2 - x_1^2) = \begin{cases} 1 - \epsilon & \text{if } 0 < \epsilon \leq \frac{1}{2}, \\ \frac{3}{4} - \epsilon^2 & \text{if } \epsilon \geq \frac{1}{2}, \end{cases} \quad (1.68)$$

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 $\epsilon \in (0, \sqrt{3}/2)$. Then the smoothness condition (1.33) is satisfied, with $L(\mathbf{x}) = f'(0)(x_1, x_2) = x_1 - \mu x_2/2$ for $\mathbf{x} = (x_1, x_2) \in \mathfrak{X}$, and upon letting

$$V = (Y - \mu, (Y - \mu)^2 - 1) \quad (1.69)$$

we see that $\sqrt{n}f(\bar{V}) = T - \sqrt{n}\mu$. Then Corollary 1.3.7 immediately yields

Theorem 1.4.5. *Take any $\omega > 0$ and assume that $\tilde{\sigma} > 0$ and $v_3 < \infty$, for $\tilde{\sigma}$ and v_p defined in (1.51). Then for all $z \in \mathbb{R}$ and $n \in \mathbb{N}$*

$$\left| \mathbb{P}\left(\frac{T - \sqrt{n}\mu}{\tilde{\sigma}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{\sqrt{n}}, \quad (1.70)$$

where \mathfrak{C} is a finite expression depending only on the distribution of Y ; also, for all real $z > 0$ and $n \in \mathbb{N}$ satisfying (1.49)

$$\left| \mathbb{P}\left(\frac{T - \sqrt{n}\mu}{\tilde{\sigma}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}, \quad (1.71)$$

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

Remark 1.4.6. 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 $\tilde{\sigma} = 0 \Leftrightarrow L(V) = 0$ 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 $p \in (0, 1) \setminus \{\frac{1}{2}\}$.

Much work has been done rather recently concerning the distribution of the central T ; see some references in this regard in Subsubsection 1.4.2.1 below.

On the other hand, the bounds in (1.70) and (1.71) appear to be new for the non-central T . Bentkus, Jing, Shao, and Zhou [11] recently showed 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 1.4.6 concerning the degeneracy condition $\tilde{\sigma} = 0$.

Remark 1.4.7. The condition $\|Y\|_4 < \infty$ in [11] is equivalent to $\|V\|_2 < \infty$, where V is as in (1.69). Therefore, it appears natural to require that $\|V\|_3 < \infty$, or equivalently $\|Y\|_6 < \infty$, in order to obtain a bound of order $\mathcal{O}(1/\sqrt{n})$; cf. the classical BE bound for linear statistics, where the finiteness of the third moment of the summand r.v.’s is usually imposed to achieve

a bound of order $\mathcal{O}(1/\sqrt{n})$. In fact, the asymptotic expansion for the distribution of T up to the order of $\mathcal{O}(1/\sqrt{n})$ (which follows from the general results for nonlinear statistics obtained by Bhattacharya and Ghosh [14]) indeed contains $\|Y\|_6$ whenever the mean μ is nonzero.

The “central”, or “null”, case when $\mu = 0$ is in this sense exceptional, as discussed in Remark 1.4.9. In this case, it is well known that the finiteness of the $\mathbb{E}|Y|^3$ is enough for a uniform BE bound for T . On the other hand, Novak shows at the end of [94] that no nonuniform bound of the form $\mathbb{E}|Y|^3 g(z)/\sqrt{n}$ for the self-normalized sum, a statistic closely related to the central T , can hold for any positive function g such that $g(z) \downarrow 0$ as $z \uparrow \infty$.

1.4.2.1 Central T and the self-normalized sum

The central T is very close to the self-normalized sum

$$T_1 := \frac{Y_1 + \dots + Y_n}{\sqrt{Y_1^2 + \dots + Y_n^2}} = \frac{\sqrt{n} \bar{Y}}{\sqrt{\bar{Y}^2}} = \frac{T}{\sqrt{1 + T^2/n}}. \quad (1.72)$$

In particular, letting $z_n := z/\sqrt{1 + z^2/n}$, one has $\mathbb{P}(T \leq z) = \mathbb{P}(T_1 \leq z_n)$ for all $z \in \mathbb{R}$ and hence

$$\begin{aligned} & \left| \sup_{z \in \mathbb{R}} |\mathbb{P}(T \leq z) - \Phi(z)| - \sup_{z \in \mathbb{R}} |\mathbb{P}(T_1 \leq z) - \Phi(z)| \right| \\ & \leq \sup_{z \in \mathbb{R}} |\Phi(z_n) - \Phi(z)| \leq \sup_{u \in \mathbb{R}} |u^3 \Phi'(u)| / (2n) = (3/(2e))^{3/2} / (n\sqrt{\pi}) < 0.24/n, \end{aligned}$$

which is much less than $1/\sqrt{n}$; cf. [111, Proposition 1.4] and its proof, where Student’s T was defined using the sample standard deviation (as opposed to the empirical standard deviation) of the random sample $(Y_i)_{i=1}^n$.

Slavova [139] appears to have first produced a uniform BE-type bound for T of the optimal order in n , namely of the form C/\sqrt{n} , where C depends only on $\mathbb{E}|Y|^3$. It was only in 1996 that Bentkus and Götze [10, Theorem 1.2] obtained a uniform BE-type bound of the optimal order in n and with the “correct” dependence on the moments; namely, they showed that there exists an absolute constant A such that

$$|\mathbb{P}(T \leq z) - \Phi(z)| \leq An \mathbb{E} \left[\left(\frac{Y}{\sqrt{n}} \right)^2 \wedge \left| \frac{Y}{\sqrt{n}} \right|^3 \right] \quad (1.73)$$

for all $z \in \mathbb{R}$; note that the above bound is no greater than $A \mathbb{E}|Y|^p / n^{p/2-1}$ for any $p \in [2, 3]$. Bentkus, Bloznelis, and Götze [8] provided a similar bound when the Y_i ’s are not necessarily identically distributed (i.d.). Shao [134, Theorem 1.1] obtained a version of (1.73) with explicit absolute constants (and also without the i.d. assumption), which in particular implies that in the i.i.d. case for all $z \in \mathbb{R}$

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq \frac{25\|Y\|_3^3}{\sqrt{n}}. \quad (1.74)$$

Novak [93, 94] obtained BE-type bounds for T_1 ; however, the structure of those bounds is rather complicated.

Nagaev [87, Theorem 1 and (1.18)], stated that for all $z \in \mathbb{R}$

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq \frac{36\|Y\|_3^3 + 9}{\sqrt{n}} \wedge \frac{4.4\|Y\|_3^3 + \|Y\|_4^4/\|Y\|_3^3 + \|Y^2 - 1\|_3^3}{\sqrt{n}} \quad (1.75)$$

when the Y_i 's are i.i.d. However, there are a number of mistakes of various kinds in the proof in [87]; see [111] for details.

Remark 1.4.8. Pinelis [111, Theorem 1.2] obtained a bound of the form

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq \frac{1}{\sqrt{n}} \left(A_3\|Y\|_3^3 + A_4\|Y^2 - 1\|_2 + A_6 \frac{\|Y^2 - 1\|_3^3}{\|Y\|_3^3} \right) \quad (1.76)$$

for all $z \in \mathbb{R}$, where the triple (A_3, A_4, A_6) depends on several parameters whose values may be freely chosen within certain ranges. For instance, a specific choice of the parameters yields $(A_3, A_4, A_6) = (1.53, 1.52, 1.28)$. Thus, all the constant factors A_3, A_4, A_6 in (1.76) can be made rather small. A bound for the general, non-i.i.d. case, similar to (1.76) but with slightly greater constants, was also obtained in [111]; as shown there, that bound in [111] compares well with (1.74), especially after truncation.

A number of important advances concerning limit theorems for the central T and/or T_1 have been made rather recently. For instance, Hall [45] obtained an Edgeworth expansion of the distribution of T . It was only in 1997 that Giné, Götze, and Mason [41] found a necessary and sufficient condition for the Student statistic to be asymptotically standard normal. Shao [132, 133], Nagaev [88], Jing, Shao, and Wang [61], and Wang and Hall [142] studied the probabilities of large deviations. Chistyakov and Götze [20, 21] and Jing, Shao, and Zhou [62] considered the probabilities of moderate deviations. See Giné and Mason [42] and Pang, Zhang, and Wang [96] concerning the law of the iterated logarithm, and Wang and Jing [143] and Robinson and Wang [126] for exponential nonuniform BE bounds. This is of course but a sampling of the recent work done concerning asymptotic properties of the central T and the related self-normalized sums; for work done somewhat earlier, the reader may be referred to the bibliography in [10].

Remark 1.4.9. The central T (as compared with the noncentral one) is special for two reasons: (i) when $\mu = 0$, then $L(V) = Y$ and, to be finite, $\tilde{\sigma}$ needs only the second moment of Y (rather than the fourth) to exist; and (ii) while in general Δ is rather naturally of the order $\sqrt{n}\|\bar{V}\|^2$, Δ is significantly smaller for the central T . Moreover, the first term, $\sqrt{n}L(\bar{V})/\tilde{\sigma}$, in a formal stochastic expansion of the central T is precisely $\sqrt{n}\bar{Y}$ and thus linear in the Y_i 's, whereas for the noncentral T this term contains \bar{Y}^2 . This heuristics is reflected in Corollary 1.4.10 below, which is derived using Theorem 1.2.1, with a better choice of Δ for this specific case than that for the general results of Section 1.3.

Corollary 1.4.10 (to Theorem 1.2.1). *Let Y, Y_1, \dots, Y_n be i.i.d. r.v.'s, with $\mathbb{E}Y = 0$ and $\|Y\|_2 = 1$. Then*

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq \frac{1}{\sqrt{n}} \left(A_3\|Y\|_3^3 + A_4\|Y\|_4^4 - A_0 \right) \quad (1.77)$$

for all $z \in \mathbb{R}$ and any triple

$$(A_3, A_4, A_0) \in \{(3.01, 5.16, 4.75), (3.20, 2.20, 1.14), (3.65, 1.31, -1.45)\}. \quad (1.78)$$

It appears that the bound in (1.77) may in certain cases be competitive with the bound in (1.76) (say with $(A_3, A_4, A_6) = (1.53, 1.52, 1.28)$, as before), even though the bound in (1.76) was obtained by methods specifically designed for T_1 . Therefore, by Remark 1.4.8, the bound in (1.77) may also in certain cases compare well with that in (1.74); see Remark 1.4.13 for some details.

The uniform and nonuniform bounds presented in Corollaries 1.4.11 and 1.4.12, respectively, involve the sixth moments of Y , as they are based on the general results of Theorems 1.4.1 and 1.A.2, with Δ being on the order of magnitude of $\|S\|^2/\sigma = \sqrt{n}\|\overline{V}\|^2$.

Corollary 1.4.11 (to Theorem 1.4.1). *Let Y, Y_1, \dots, Y_n be i.i.d. r.v.'s, with $\mathbb{E}Y = 0$ and $\|Y\|_2 = 1$. Then*

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq \frac{1}{\sqrt{n}} \left(A_3 \|Y\|_3^3 + A_4 \|Y\|_4^6 + A_6 \|Y\|^2 - 1 \right) \quad (1.79)$$

for all $z \in \mathbb{R}$ and either triple

$$(A_3, A_4, A_6) \in \{(3.33, 3.33, 0.17), (5.79, 1.45, 0.26)\}. \quad (1.80)$$

The two triples (A_3, A_4, A_6) in (1.80) are the result of trying to approximately minimize $A_3 \vee (A_4/w_4) \vee (A_6/w_6)$, with weights $(w_4, w_6) \in \{(1, 0.05), (0.25, 0.05)\}$.

One can see that the constants in (1.79)–(1.80) are not much worse than those in (1.77)–(1.78).

Corollary 1.4.12 (to Theorem 1.A.2). *Let $\omega \in \{0.1, 0.5\}$, $w_g \in \{0, 1\}$, and*

$$g(z) := \frac{1}{z^3} + \frac{w_g}{e^{z/2}}. \quad (1.81)$$

Then under the assumptions of Corollary 1.4.11, for all

$$z \in (0, \omega\sqrt{n}] \quad (1.82)$$

one has

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq \frac{g(z)}{\sqrt{n}} (\hat{A}_3 \|Y\|_3^3 + \hat{A}_4 \|Y\|_4^8 + \hat{A}_6 \|Y\|^2 - 1) \quad (1.83)$$

where, for any given pair $(\omega, w_g) \in \{0.1, 0.5\} \times \{0, 1\}$, the triple $(\hat{A}_3, \hat{A}_4, \hat{A}_6)$ is either one of the two triples given in the corresponding block of Table 1.1 below.

One can see that, especially in the case when $w_g = 1$ and $\omega = 0.1$, the sum of the constants \hat{A}_3 , \hat{A}_4 , and \hat{A}_6 is comparable with the constant factor 30.2211 in the nonuniform BE inequality (1.26) for linear statistics. One may also note that the constants \hat{A}_3 , \hat{A}_4 , and \hat{A}_6 in the case when $w_g = 0$ are significantly greater than those for $w_g = 1$. This reflects the fact that, whereas $\frac{1}{e^{z/2}}$ is much smaller than $\frac{1}{z^3}$ for very large $z > 0$, the maximum of the ratio $\frac{1}{e^{z/2}}/\frac{1}{z^3}$ over $z > 0$ is (attained at $z = 6$ and) quite large, about 10.75. Whereas

Table 1.1: Constants associated with nonuniform bound in (1.83)

	$\omega = 0.1$			$\omega = 0.5$		
	\hat{A}_3	\hat{A}_4	\hat{A}_6	\hat{A}_3	\hat{A}_4	\hat{A}_6
$w_g = 1$	38	36	36	48	48	42
	39	20	7	66	33	13
$w_g = 0$	151	148	147	166	166	165
	170	85	29	229	115	45

at least some of the constants \hat{A}_3 , \hat{A}_4 , and \hat{A}_6 are rather large when $w_g = 0$, one can put this into a perspective by recalling that, even in the much simpler case of sums of independent identically distributed r.v.'s, the first explicit constant in the nonuniform BE bound (obtained in [95]) was greater than 1955.

Similarly to their counterparts in [111], the proofs of Corollaries 1.4.11 and 1.4.12 demonstrate a method by which one may obtain a variety of specific numerical constants for the bounds of the form (1.79) and (1.83). In particular, the introduction of the numerous parameters in Theorems 1.4.1 and 1.A.2 allows one to account more accurately for the relations between the possible sizes of the various moments (cf. e.g. the ideas represented by [99, Theorems 5.2, 6.1, 6.2]). On the other hand, such an approach rather understandably results in significantly more complicated expressions.

Remark 1.4.13. The uniform bounds in (1.77) and (1.79) (as well as the nonuniform one in (1.83)) involve moments of orders higher than 3, in contrast with the uniform bound in (1.74), say. However, it appears that the effect of the smaller constants in (1.77)–(1.78) and (1.79)–(1.80) will oftentimes more than counterbalance the “defect” of the higher-order moments. For instance, suppose that $Y \sim \tilde{t}_d$, where \tilde{t}_d denotes the standardized t distribution with d degrees of freedom, where d is any positive real number. This distribution is symmetric. Its tails vary from very heavy ones for small d to the very light tails of the standard normal distribution, corresponding to the limit case $d = \infty$. The absolute moments, say $m_s(d)$, of order s of the distribution \tilde{t}_d will be infinite for all $s \in [d, \infty)$. Then, in particular, the bound in (1.74) will be infinite if $d \leq 3$. On the other hand, one can show that for $Y \sim \tilde{t}_d$ the bound in (1.77) (say with the choice of the triple $(A_3, A_4, A_0) = (3.65, 1.31, -1.45)$ in (1.78)) will be smaller than that in (1.74) for all real $d \geq 4.16$; this can be checked using monotonicity properties of $m_3(d)$ and $m_4(d)$. Namely, $m_4(d) = \frac{3(d-2)}{d-4}$ clearly decreases in $d > 4$. As for $m_3(d)$, one can write $\sqrt{\frac{\pi}{8}}m_3(d) = r(\frac{d-3}{2})$

for $d > 3$, where $r(x) := \frac{\sqrt{x+\frac{1}{2}}\Gamma(x)}{\Gamma(x+\frac{1}{2})}$. So, reasoning as in the proof of [110, Lemma 2.1], one has $(\ln m_3(d))'_d(d-3) = -\int_0^1 \frac{t^{d-3}}{(t+1)^2} dt - \frac{1}{2(d-2)} < 0$ for all $d > 3$, whence $m_3(d)$ decreases in $d > 3$. Note also here that the bound in (1.74) will be nontrivial (that is, less than 1) for some $d \in (0, 4.16)$ only if $n > 25^2 \|Y\|_3^6 = 25^2 m_3(d) > 25^2 m_3(4.16) > 4408$, where $m_3(d)$ stands for the third absolute moment of \tilde{t}_d .

Similarly, the bound in (1.77) (again with $(A_3, A_4, A_0) = (3.65, 1.31, -1.32)$) will be smaller than that in (1.74) when Y has any standardized two-point distribution which is not too skewed – it is enough that $\mathbf{P}(Y = \sqrt{q/p}) = p$, $\mathbf{P}(Y = -\sqrt{p/q}) = q$, $0 < p < 1$, $q := 1 - p$, and $p \wedge q \geq 0.00383$; moreover, if $p \wedge q < 0.00383$ then the bound in (1.74)

will be nontrivial only if $n > 25^2 \|Y\|_3^6 = 25^2 \left(\frac{1}{\sqrt{pq}} - 2\sqrt{pq} \right)^2 > 161322$. Note that any zero-mean distribution is a mixture of zero-mean two-point distributions [105], so that such distributions appear to be of particular interest.

Discussion in Pinelis [111] shows that one could utilize appropriate truncation to further alleviate the presence of the higher-order moments and thus make the comparison even more favorable to the bounds with the smaller constants.

Remark 1.4.14. One may also want to compare, in the case of the statistic T_1 , the asymptotic behavior of our bounds described in Corollary 1.4.4 with the corresponding known asymptotic results. In particular, it follows from [94, (*)] that

$$\limsup_{n \rightarrow \infty} \sup_{z \in \mathbb{R}} |\mathbf{P}(T_1 \leq z) - \Phi(z)| \leq 6.4 \|Y\|_3^3 + 2 \|Y\|_1 \quad (1.84)$$

whenever $\|Y\|_3 < \infty$. On the other hand, taking any real $\theta > 0$, $f(x_1, x_2) := \frac{x_1/\theta}{\sqrt{1+x_2}}$ for $(x_1, x_2) \in \mathbb{R} \times (-1, \infty)$, and $V := (\theta Y, Y^2 - 1)$, one has $\sqrt{n}f(\bar{V}) = T_1$. Choose now $\theta = \|Y\|_4^4 - 1$ (assuming that $\|Y\|_4 \neq 1$ and hence $\|Y\|_4 > 1$; the case $\|Y\|_4 = 1$ can then be treated by continuity, say). Then, by (1.62), $y_* = (3 + 2\|Y\|_3)\sqrt{\|Y\|_4^4 - 1}$. Using this expression for y_* , one can show that both bounds in (1.61) will be smaller than that in (1.84) (and even smaller than $6.4\|Y\|_3^3$) – say, in the case when Y has the standardized t distribution with d degrees of freedom, for any real $d > 6$. The same will be true when Y has any standardized two-point distribution which is not too skewed – it is enough that $\mathbf{P}(Y = \sqrt{q/p}) = p$, $\mathbf{P}(Y = -\sqrt{p/q}) = q$, $0 < p < 1$, $q := 1 - p$, and $p \wedge q \geq .00906$. Note also that in the case of the statistic T_1 one can get an asymptotic bound better than the one just obtained based on Corollary 1.4.4 (which latter is derived from Theorem 1.3.2, which in turn is a corollary to Theorem 1.2.1) – if instead one uses Theorem 1.2.1 directly; cf. Corollary 1.4.10 (to Theorem 1.2.1) vs. Corollary 1.4.11 (to Theorem 1.4.1).

Remark 1.4.15. Consider now the asymptotic behavior of the nonuniform bound for T_1 . Novak [93, Theorem 10] provides an explicit, though complicated in appearance, nonuniform BE-type bound for this statistic. Using [93, (5.10)] and (1.26) (and still assuming that $\mathbf{E}Y = 0$ and $\mathbf{E}Y^2 = 1$, as well as $\mathbf{E}Y^4 < \infty$) one can show that

$$\limsup_{n \rightarrow \infty} \sup_{g(n) \leq z \leq n^{1/6}} z^3 \sqrt{n} |\mathbf{P}(T_1 \leq z) - \Phi(z)| \leq 30.2211 \|Y\|_3^3; \quad (1.85)$$

here g stands for any positive increasing unbounded function on \mathbb{N} . Thus, for the specific statistic T_1 , the asymptotic bound in (1.85) coincides with that in (1.63), obtained for general nonlinear statistics of the form $f(\bar{V})$. Note also that the bound in (1.63) holds for z in the zone $[g(n), n^{1/2}/g(n)]$, which is much wider than the zone $[g(n), n^{1/6}]$ in (1.85) if g is taken to grow slowly enough. On the other hand, Theorem 1.4.1 and then Corollary 1.4.4 contain the moment condition $v_3 < \infty$, which is equivalent, in the specific case of T_1 , to $\mathbf{E}Y^6 < \infty$, which is more stringent than the corresponding condition $\mathbf{E}Y^4 < \infty$ used here to derive (1.85).

Remark 1.4.16. Suppose here that, in addition to the other condition of Corollary 1.4.11, the r.v. Y is symmetric. Then, by Remark 1.4.2, the triples (A_3, A_4, A_6) of the constants in

Corollary 1.4.11 can be replaced by either one of the triples (of smaller constants) in the set

$$\{(3.09, 3.09, 0.16), (5.33, 1.34, 0.27)\}. \quad (1.86)$$

Similarly, by Remark 1.A.3, Table 1.1 can then be replaced by Table 1.2.

Table 1.2: Constants associated with nonuniform bound in (1.83) when Y is symmetric

	$\omega = 0.1$			$\omega = 0.5$		
	\hat{A}_3	\hat{A}_4	\hat{A}_6	\hat{A}_3	\hat{A}_4	\hat{A}_6
$w_g = 1$	35	32	31	48	48	41
	37	19	5	57	29	12
$w_g = 0$	124	124	121	141	138	138
	145	73	22	206	103	42

Proofs of these statements are provided in the proofs of the corresponding corollaries.

1.4.3 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

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

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

$$R := \frac{\sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}} = \frac{\overline{YZ} - \bar{Y} \bar{Z}}{\sqrt{\overline{Y^2} - \bar{Y}^2} \sqrt{\overline{Z^2} - \bar{Z}^2}}, \quad (1.87)$$

where

$$(\bar{Y}, \bar{Z}, \overline{Y^2}, \overline{Z^2}, \overline{YZ}) := \frac{1}{n} \sum_i (Y_i, Z_i, Y_i^2, Z_i^2, Y_i Z_i);$$

let us allow R to take an arbitrary value r_0 if the denominator in (1.87) 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:

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

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 $(1 + x_3 - x_1^2)(1 + x_4 - x_2^2) > 0$, let

$$f(\mathbf{x}) = f(x_1, x_2, x_3, x_4, x_5) = \frac{x_5 + \rho - x_1 x_2}{\sqrt{1 + x_3 - x_1^2} \sqrt{1 + x_4 - x_2^2}} - \rho; \quad (1.88)$$

let $f(\mathbf{x}) := r_0 - \rho$ for all other $\mathbf{x} \in \mathfrak{X}$. Recall (1.68) 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 (1.33) 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 Corollary 1.3.7 immediately yields

Theorem 1.4.17. Take any $\omega > 0$ and assume that $\tilde{\sigma} > 0$ and $v_3 < \infty$. Then for all $z \in \mathbb{R}$ and $n \in \mathbb{N}$

$$\left| \mathbb{P}\left(\frac{R - \rho}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{\sqrt{n}}, \quad (1.89)$$

where \mathfrak{C} is a finite expression depending only on the distribution of the random point (Y, Z) ; also, for all real $z > 0$ and $n \in \mathbb{N}$ satisfying (1.49)

$$\left| \mathbb{P}\left(\frac{R - \rho}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}, \quad (1.90)$$

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

Remark 1.4.18. 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 1.4.19. In order to get a uniform bound of order $\mathcal{O}(1/\sqrt{n})$ in Theorem 1.4.17, it is necessary to assume that $v_3 < \infty$, which is equivalent to $\|Y\|_6 + \|Z\|_6 < \infty$. This moment condition might seem overly restrictive, since only third absolute moments are required to obtain a BE-type bound of the same order for linear statistics (or even for the central Student statistic). 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 $\rho \neq 0$; cf. Remark 1.4.7; for details, one can see [98]. When $\rho = 0$, the most restrictive moment assumption for the existence of the asymptotic expansion is that $\|YZ\|_3 < \infty$.

The bounds in (1.89) and (1.90) 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 [92]. Shen [135] recently provided results concerning probabilities of large deviations for R in the special case when (Y, Z) is a bivariate normal r.v. Formal asymptotic expansions for the density of R follow from the paper by Kollo and Ruul [71].

We next state one particular simplification of the uniform bound in (1.53) when applied to the Pearson statistic in the case when $\rho = 0$.

Corollary 1.4.20 (to Theorem 1.4.1). Assume that $\mathbb{E} YZ = 0$ and $\tilde{\sigma} = \|YZ\|_2 > 0$. Then

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

$$\left| \mathbb{P}\left(\frac{R}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{B_0 + B_3/\tilde{\sigma}^3}{\sqrt{n}} (\|Y\|_6^6 + \|Z\|_6^6), \quad (1.91)$$

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

$$\{(4.08, 4.08), (1.38, 11.02), (14.32, 1.79), (0.71, 19.16), (37.82, 1.41)\}. \quad (1.92)$$

Similarly to the proof of Corollary 1.4.11, that of Corollary 1.4.20 gives a method by which one may obtain a variety of values for the pair (B_0, B_3) . The specific pairs listed in (1.92) are obtained by trying to minimize $B_0 \vee B_3/\tilde{\sigma}^3$ for $\tilde{\sigma} \in \{1, 2, 1/2, 3, 1/3\}$.

Remark 1.4.21. By employing the improvement described in Remark 1.2.2, it is possible to slightly decrease the constants in (1.92); namely, (1.91) holds for any ordered pair (B_0, B_3) in the set

$$\{(4.06, 4.06), (1.38, 10.99), (14.62, 1.83), (0.71, 19.14), (38.70, 1.44)\}.$$

Remark 1.4.22. By Remark 1.4.2, in the case when the r.v. YZ is symmetrically distributed, the constants in (1.92) can be slightly improved; namely, then (1.92) can be replaced by

$$\{(3.74, 3.74), (1.31, 10.47), (13.19, 1.65), (0.68, 18.27), (36.40, 1.35)\}. \quad (1.93)$$

Remark 1.4.23. Bounds similar to the ones in Corollary 1.4.20 can be obtained, e.g., for other statistics related to Pearson's R , including the Fisher z transform. However, for reasons discussed in Appendix 1.D and because the paper is already quite long, we chose not to present such results here.

1.4.4 Non-central Hotelling's T^2 statistic

Let $k \geq 2$ be an integer, and let Y, Y_1, \dots, Y_n be i.i.d. r.v.'s 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

$$T^2 := \bar{Y}^\top (S_Y^2/n)^{-1} \bar{Y} = n \bar{Y}^\top \left(\overline{Y Y^\top} - \bar{Y} \bar{Y}^\top \right)^{-1} \bar{Y}, \quad (1.94)$$

where

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

the generalized inverse is often used in place of the inverse in (1.94), though here we may allow T^2 to take any value t_0^2 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 “non-central” otherwise.

For any nonsingular matrix B , T^2 is invariant under the invertible transformation $Y_i \mapsto$

BY_i , so let us assume w.l.o.g. that

$$\text{Cov } Y = 1,$$

the $k \times k$ identity matrix.

Now let $\mathfrak{X} = \{(x_1, x_2) : x_1 \in \mathbb{R}^k, x_2 \in \mathbb{R}^{k \times k}\}$, equipped with the norm $\|(x_1, x_2)\| := \sqrt{\|x_1\|^2 + \|x_2\|_F^2}$, 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 $1 + x_2 - x_1 x_1^\top$ is nonsingular, let

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

and let $f(\mathbf{x}) := t_0^2/n - \mu^\top \mu$ 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 [54]): $\|B\| \leq \|B\|_F$ for any $k \times k$ matrix B , and $\|B\| < 1$ implies $1 - B$ is nonsingular and $\|(1 - B)^{-1}\| \leq 1/(1 - \|B\|)$. 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 (1.68)), so that the smoothness condition (1.33) holds. Upon letting $V = (Y - \mu, (Y - \mu)(Y - \mu)^\top - 1)$, we see that $n f(\bar{V}) = T^2 - n \mu^\top \mu$. Then Corollary 1.3.7 immediately yields

Theorem 1.4.24. *Take any $\omega > 0$ and assume that $\tilde{\sigma} > 0$ and $v_3 < \infty$. Then for all $z \in \mathbb{R}$ and $n \in \mathbb{N}$*

$$\left| \mathbb{P}\left(\frac{T^2 - n \mu^\top \mu}{\tilde{\sigma} \sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{\sqrt{n}}, \quad (1.95)$$

where \mathfrak{C} is a finite expression depending only on the distribution of Y ; also, for all real $z > 0$ and $n \in \mathbb{N}$ satisfying (1.49)

$$\left| \mathbb{P}\left(\frac{T^2 - n \mu^\top \mu}{\tilde{\sigma} \sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{z^3 \sqrt{n}}, \quad (1.96)$$

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

Remark 1.4.25. The non-degeneracy condition $\tilde{\sigma} > 0$ immediately implies that $\mu \neq 0$, so that Theorem 1.4.24 is applicable only to the non-central 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

$$x_1 = 1 + \|\mu\|^2 + \sqrt{1 + \|\mu\|^2}, \quad x_2 = 1 + \|\mu\|^2 - \sqrt{1 + \|\mu\|^2}, \quad p = \frac{1}{2} \left(1 - \frac{1}{\sqrt{1 + \|\mu\|^2}}\right);$$

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 1.4.6. 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

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

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

Again, the bounds in (1.95) and (1.96) appear to be new; we have found no mention of BE bounds for T^2 in the literature. Probabilities of moderate and large deviations for the central Hotelling T^2 statistic (when $\mu = 0$) were considered by Dembo and Shao [29]. Asymptotic expansions for the generalized T^2 distribution for *normal populations* were given by Itô [58] (for $\mu = 0$), and by Itô [59], Siotani [137], and Muirhead [85] (for any μ); Kano [66] and Fujikoshi [37] give an asymptotic expansion for the distribution of the central T^2 for non-normal populations, and Kakizawa and Iwashita [65] do this for the noncentral T^2 statistic.

1.5 Proofs

All necessary proofs of the theorems and corollaries stated in the previous sections are provided here – except for Corollaries 1.4.11, 1.4.12, and 1.4.20, whose proofs are given in Appendix 1.C.

1.5.1 Proofs of results from Section 1.2

Proof of Theorem 1.2.1. As noted in Remark 1.2.3, the assertion of Theorem 1.2.1 is very similar to that of [19, Theorem 2.1]. From the condition that $|\Delta| \geq |T - W|$ (cf. [19, (5.1)])

$$-P(z - |\Delta| \leq W \leq z) \leq P(T \leq z) - P(W \leq z) \leq P(z \leq W \leq z + |\Delta|) \quad (1.97)$$

for all $z \in \mathbb{R}$. The inequality

$$P(z \leq W \leq z + |\overline{\Delta}|) \leq \frac{1}{2c_*} \left(4\delta + \mathbf{E}|W\overline{\Delta}| + \sum_{i=1}^n \mathbf{E}|\xi_i(\overline{\Delta} - \Delta_i)| \right)$$

is proved by modifying the proof of [19, Theorem 2.1] – replacing their Δ with our $\overline{\Delta}$ and their condition (2.2) with our (1.4). Recalling the condition (1.5) on $\overline{\Delta}$, one has

$$P(z \leq W \leq z + |\Delta|) \leq P(z \leq W \leq z + |\overline{\Delta}|) + P(\max_i \eta_i > w). \quad (1.98)$$

Then $P(z - |\Delta| \leq W \leq z)$ can be bounded in a similar fashion, using $z - |\Delta|$ in place of z , and (1.6) follows.

In order to prove Remark 1.2.2, note that [19, (5.6)] still remains valid when $H_{1,2}$ there is replaced by

$$H_{1,2} = \mathbf{E} \left| \{z \leq W \leq z + |\overline{\Delta}|\} \sum_i (\check{\xi}_i - \mathbf{E} \check{\xi}_i) \right|, \quad \text{with} \quad \check{\xi}_i := |\xi_i|(\delta \wedge |\xi_i|);$$

here, in distinction with the definition of $H_{1,2}$ in [19], the notation $\check{\xi}_i$ is used in place of η_i . Then the Cauchy-Schwarz inequality yields

$$H_{1,2} \leq \sqrt{\mathbb{E} \mathbb{I}\{z \leq W \leq z + |\bar{\Delta}|\}} \sqrt{\sum_i \mathbb{E} \check{\xi}_i^2} \leq \delta \sqrt{\mathfrak{p}}, \quad \text{where } \mathfrak{p} := \mathbb{P}(z \leq W \leq z + |\bar{\Delta}|);$$

cf. [19, (5.8)]. Following through with the remainder of the proof of [19, Theorem 2.1], we have

$$c_* \mathfrak{p} - \delta \mathfrak{p}^{1/2} \leq \mathfrak{b} := \frac{1}{2} \left(2\delta + \mathbb{E}|W\bar{\Delta}| + \sum_i \mathbb{E}|\xi_i(\bar{\Delta} - \Delta_i)| \right).$$

So,

$$\begin{aligned} \mathfrak{p} &\leq \left(\frac{\delta + \sqrt{\delta^2 + 4c_* \mathfrak{b}}}{2c_*} \right)^2 = \frac{2\delta^2 + 4c_* \mathfrak{b} + 2\delta \sqrt{\delta^2 + 4c_* \mathfrak{b}}}{4c_*^2} \\ &= \frac{1}{2c_*} \left(2\mathfrak{b} + \frac{\delta^2}{c_*} + 2\delta \sqrt{\frac{1}{2c_*} \left(2\mathfrak{b} + \frac{\delta^2}{2c_*} \right)} \right); \end{aligned}$$

in view of (1.97) and (1.98), this verifies the improvement provided in Remark 1.2.2. \square

Proof of Theorem 1.2.4. The proof of Theorem 1.2.4 largely follows the lines of that of [19, Theorem 2.2]; for the ease of comparison between the two proofs, we shall use notation similar to that in [19]. The extension to p other than 2 is obtained using a Cramér-tilt absolutely continuous transformation of measure along with the mentioned Rosenthal-type and exponential bounds. Introduce the Winsorized r.v.'s

$$\bar{\xi}_i := \xi_i \wedge w \quad \text{and their sum,} \quad \bar{W} := \sum_{i=1}^n \bar{\xi}_i. \quad (1.99)$$

Note that in the statement of [19, Lemma 5.1] the $\bar{\xi}_i$'s are defined as the truncated r.v.'s $\xi_i \mathbb{I}\{\xi_i \leq w\}$ (with $w = 1$). A problem with this definition arises on page 596 in [19] concerning the assertion there that $\sum_i \mathbb{E}|\xi_i|(\delta \wedge |\bar{\xi}_i|) = \sum_i \mathbb{E}|\xi_i|(\delta \wedge |\xi_i|)$ whenever $\delta \leq 0.07$; indeed, by letting ξ_i take values ± 2 each with probability $\frac{1}{8n}$ and the value 0 with probability $1 - \frac{1}{4n}$, the assertion is seen to be false when $\bar{\xi}_i = \xi_i \mathbb{I}\{\xi_i \leq 1\}$ (while true if $\delta \leq w$ and $\bar{\xi}_i = \xi_i \wedge w$). See [108] for a general discussion on comparative merits of the Winsorization vs. truncation, especially in regard to the Cramér tilt transformation.

Recalling the definition (1.18) of the measure $\hat{\mathbb{P}}$, one has

$$\begin{aligned} \hat{\mathbb{P}}(z - |\Delta| \leq W \leq z) &= \mathbb{P}(z - |\Delta| \leq W \leq z, |\Delta| \leq \pi_1 z) \\ &\leq \sum_{i=1}^n \mathbb{P}(W \geq (1 - \pi_1)z, \eta_i > w) + \mathbb{P}(z - |\Delta| \leq W \leq z, |\Delta| \leq \pi_1 z, \max_i \eta_i \leq w) \\ &\leq \sum_{i=1}^n \mathbb{P}(\xi_i > \pi_2 z) + \sum_{i=1}^n \mathbb{P}(W - \xi_i \geq (1 - \pi_1 - \pi_2)z) \mathbb{P}(\eta_i > w) \\ &\quad + \mathbb{P}(z - |\bar{\Delta}| \leq \bar{W} \leq z, |\bar{\Delta}| \leq \pi_1 z) \\ &= \gamma_z + \mathbb{P}(z - |\bar{\Delta}| \leq \bar{W} \leq z, |\bar{\Delta}| \leq \pi_1 z); \end{aligned} \quad (1.100)$$

here the second inequality follows from the independence of $W - \xi_i$ and η_i , the condition (1.5) on $\bar{\Delta}$, and the definition (1.99) of \bar{W} (recall also the condition that $\xi_i \leq \eta_i$), and the second equality follows from the definitions of γ_z and π_3 in (1.19) and (1.16); cf. [19, Lemma 5.1].

We must next establish the inequality

$$\mathbf{P}(z - |\bar{\Delta}| \leq \bar{W} \leq z, |\bar{\Delta}| \leq \pi_1 z) \leq \tau e^{-(1-\pi_1)z/\theta}; \quad (1.101)$$

cf. [19, 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 (1.16)). In the first case, when $\delta > \delta_0$,

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

here (1.11) and (1.14) are used for the third inequality above (as well as the definitions (1.24) and (1.8) of ε_1 and σ_p), and the definitions (1.23) and (1.20) of c_3 and τ are used for the last two inequalities there. Thus, (1.101) is established when $\delta > \delta_0$.

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

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

be defined similarly to [19, (5.16)]. Then, by the independence of $(\Delta_i, \bar{W} - \bar{\xi}_i)$ and ξ_i ,

$$\mathbf{E} W f_{\bar{\Delta}}(\bar{W}) = G_1 + G_2, \quad (1.102)$$

where

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

Also, using an obvious modification of the arguments associated with [19, (5.17)–(5.19)], one has

$$G_1 \geq G_{1,1} - G_{1,2}, \quad (1.103)$$

where

$$G_{1,1} := c_* \exp\left\{\frac{1}{\theta}((1 - \pi_1)z - \delta)\right\} \mathbf{P}(z - |\bar{\Delta}| \leq \bar{W} \leq z, |\bar{\Delta}| \leq \pi_1 z), \quad (1.104)$$

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

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

in particular, the factor c_* in the expression (1.104) 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 (1.4); cf. [19, (5.19)]. Further,

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

so that two applications of the Cauchy-Schwarz inequality yield

$$\begin{aligned} G_{1,2} &\leq \mathbb{E} \left(\int_{|t| \leq \delta} e^{2(\overline{W}-\delta)/\theta} dt \right)^{1/2} \left(\int_{|t| \leq \delta} (\overline{M}(t) - \mathbb{E} \overline{M}(t))^2 dt \right)^{1/2} \leq \left(2\delta \mathbb{E} e^{2(\overline{W}-\delta)/\theta} \right)^{1/2} \sqrt{\delta} \\ &\leq \left(2 \text{PU}_{\exp} \left(\frac{2}{\theta}, w, 1, \varepsilon_1 \right) \right)^{1/2} e^{-\delta/\theta} \delta = \sqrt{2} \text{PU}_{\exp} \left(\frac{2}{\theta}, w, \frac{1}{\sqrt{2}}, \varepsilon_1 \right) e^{-\delta/\theta} \delta, \end{aligned} \quad (1.105)$$

where the last inequality follows from (1.11) and (1.14) (recalling also the definitions (1.99) and (1.24) of \overline{W} and ε_1); the equality in (1.105) follows from the easily verified identity

$$\text{PU}_{\exp}(\lambda, y, B, \varepsilon)^\alpha = \text{PU}_{\exp}(\lambda, y, \alpha^{1/2} B, \varepsilon) \quad \text{for any } \alpha > 0. \quad (1.106)$$

Next (cf. [19, (5.21)]),

$$\begin{aligned} |G_2| &\leq \sum_{i=1}^n \mathbb{E} |\xi_i e^{(\overline{W}-\xi_i)/\theta} (\overline{\Delta} - \Delta_i)| \\ &\leq \sum_{i=1}^n \|\xi_i e^{(\overline{W}-\xi_i)/\theta}\|_p \|\overline{\Delta} - \Delta_i\|_q \\ &= \sum_{i=1}^n \mathbb{E}^{1/p} e^{\frac{p}{\theta}(\overline{W}-\xi_i)} \|\xi_i\|_p \|\overline{\Delta} - \Delta_i\|_q \\ &\leq \text{PU}_{\exp} \left(\frac{p}{\theta}, w, \frac{1}{\sqrt{p}}, \varepsilon_1 \right) \sum_{i=1}^n \|\xi_i\|_p \|\overline{\Delta} - \Delta_i\|_q. \end{aligned} \quad (1.107)$$

Also,

$$\mathbb{E} W f_{\overline{\Delta}}(\overline{W}) \leq \mathbb{E} (|\overline{\Delta}| + 2\delta) |W| e^{\overline{W}/\theta} \leq (\|\overline{\Delta}\|_q + 2\delta) \|W e^{\overline{W}/\theta}\|_p. \quad (1.108)$$

Chen and Shao [19] bounded $\mathbb{E} W^2 e^{\overline{W}}$ (corresponding to the case when $p = 2$ and $\theta = 2$ in (1.108)) with an absolute constant; in our case, more work is required to bound the last factor in (1.108) for the general p . Specifically, we apply Cramér's tilt transform to the ξ_i 's, using at that results of [108, 107, 113].

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

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

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

$$\mathbb{E} f(\hat{\boldsymbol{\xi}}) = \frac{\mathbb{E} f(\boldsymbol{\xi}) e^{c\overline{W}}}{\mathbb{E} e^{c\overline{W}}}. \quad (1.109)$$

By [107, 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

[107, Corollary 2.7],

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

If the ξ_i 's are assumed to have symmetric distributions, then [113, Theorem 1] allows for the factor $(e^{cw} - 1)/w$ above to be replaced by $\sinh(cw)/w$; cf. Remark 1.2.8. Choose now

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

Then, by [108, Theorem 2.1],

$$\mathbb{E} e^{c\bar{\xi}_i} = \mathbb{E} e^{c(\xi_i \wedge w)} = \mathbb{E} e^{c(1 \wedge \xi_i/w)} \geq L_{W; cw, \|\xi_i\|_2/w} \geq L_{W; cw, \max_i \|\xi_i\|_2/w} = a_1^{-1},$$

where a_1 is as defined in (1.25); the last inequality above follows because $L_{W; c, \sigma}$ in [108, (2.9)] is nonincreasing in σ ; the condition $c = \frac{p}{\theta}$ was used here in the above display only for the last equality. So,

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

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

$$\begin{aligned} \|\sum_i \hat{\xi}_i\|_p &\leq \|\sum_i (\hat{\xi}_i - \mathbb{E} \hat{\xi}_i)\|_p + |\sum_i \mathbb{E} \hat{\xi}_i| \\ &\leq \mathfrak{A}_{\mathbb{R}}(p) (\sum_i \mathbb{E} |\hat{\xi}_i - \mathbb{E} \hat{\xi}_i|^p)^{1/p} + \mathfrak{B}_{\mathbb{R}}(p) (\sum_i \mathbb{E} (\hat{\xi}_i - \mathbb{E} \hat{\xi}_i)^2)^{1/2} + (e^{pw/\theta} - 1)/w \\ &\leq \mathfrak{A}_{\mathbb{R}}(p) (1.32 \sum_i \mathbb{E} |\hat{\xi}_i|^p)^{1/p} + \mathfrak{B}_{\mathbb{R}}(p) (\sum_i \mathbb{E} \hat{\xi}_i^2)^{1/2} + (e^{pw/\theta} - 1)/w \\ &\leq \mathfrak{A}_{\mathbb{R}}(p) (1.32 a_1 e^{pw/\theta} \sigma_p^p)^{1/p} + \mathfrak{B}_{\mathbb{R}}(p) (a_1 e^{pw/\theta})^{1/2} + (e^{pw/\theta} - 1)/w, \end{aligned} \tag{1.110}$$

where (1.10) is used for the second inequality above, and [112, Theorem 2.3(v)–(vi)] is used for the third inequality. Letting $f(x_1, \dots, x_n) \equiv |\sum_i x_i|^p$ in (1.109) and using (1.11), (1.14), and (1.106) once more, one has

$$\begin{aligned} \|W e^{\bar{W}/\theta}\|_p &= \left(\mathbb{E} |\sum_i \xi_i|^p e^{p\bar{W}/\theta} \right)^{1/p} = \left(\mathbb{E} e^{p\bar{W}/\theta} \mathbb{E} |\sum_i \hat{\xi}_i|^p \right)^{1/p} \\ &\leq \text{PU}_{\exp}\left(\frac{p}{\theta}, w, \frac{1}{\sqrt{p}}, \varepsilon_1\right) \|\sum_i \hat{\xi}_i\|_p. \end{aligned} \tag{1.111}$$

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

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

where the equality comes from the definition (1.104) of $G_{1,1}$, the first inequality follows from (1.102) and (1.103), and the second inequality follows from (1.105), (1.107), (1.108), (1.111), and (1.110), along with the definitions (1.21), (1.22), and (1.23) of c_1 , c_2 , and c_3 . Thus, in

view of the definition (1.20) of τ , the inequality (1.101) is proved for the other case, $\delta \leq \delta_0$.

Replace now \mathbf{P} with $\hat{\mathbf{P}}$ in (1.97), so that (1.100) and (1.101) imply

$$\hat{\mathbf{P}}(W \leq z) - \hat{\mathbf{P}}(T \leq z) \leq \gamma_z + \tau e^{-(1-\pi_1)z/\theta}.$$

In a similar fashion, one bounds $\hat{\mathbf{P}}(T \leq z) - \hat{\mathbf{P}}(W \leq z)$ from above, establishing (1.17). \square

1.5.2 Proofs of results from Section 1.3

The uniform and nonuniform BE type bounds in Theorems 1.3.2 and 1.3.5 rely on the corresponding bounds of Section 1.2. Let f be a function satisfying (1.33), and also let X_1, \dots, X_n be independent zero-mean \mathfrak{X} -valued random vectors. Further let $\sigma = \|L(S)\|_2$, as in (1.34), and for $i = 1, \dots, n$ let

$$g_i(x) := \frac{L(x)}{\sigma} \quad \text{and} \quad \xi_i = g_i(X_i) = \frac{L(X_i)}{\sigma},$$

in accordance with (1.1). The choices for the functions h_i (used to define the r.v.'s η_i) will depend on the value of p and the type of bound (uniform or nonuniform) being derived (cf. (1.37) and (1.43)). Next, let

$$T := \frac{f(S)}{\sigma}, \quad W := \sum_i \xi_i = \frac{L(S)}{\sigma},$$

and also

$$\tilde{T} := T \mathbf{I}\{\|S\| \leq \epsilon\} + W \mathbf{I}\{\|S\| > \epsilon\}. \quad (1.112)$$

Finally, let

$$\Delta := \frac{M_\epsilon}{2\sigma} \|S\|^2. \quad (1.113)$$

Then, by (1.33),

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

Adopt some more notation:

$$\tilde{X}_i := X_i \mathbf{I}\{\eta_i \leq w\}, \quad \tilde{S} := \sum_i \tilde{X}_i, \quad (1.114)$$

$$\bar{\Delta} := \frac{M_\epsilon}{2\sigma} \left(\|S\|^2 \mathbf{I}\{p = 3\} + \|\tilde{S}\|^2 \mathbf{I}\{p < 3\} \right), \quad (1.115)$$

$$\Delta_i := \frac{M_\epsilon}{2\sigma} \left(\|S - X_i\|^2 \mathbf{I}\{p = 3\} + \|\tilde{S} - \tilde{X}_i\|^2 \mathbf{I}\{p < 3\} \right). \quad (1.116)$$

Then the assumptions of Theorems 1.2.1 and 1.2.4 are satisfied for the nonlinear statistic \tilde{T} (in place of T) and its linear approximation W ; particularly, $\mathbf{E} \xi_i = 0$, $\mathbf{Var} W = 1$, $|\Delta| \geq |\tilde{T} - W|$, $\bar{\Delta}$ satisfies (1.5), and Δ_i satisfies the condition that X_i and $(\Delta_i, (X_j : j \neq i))$ are independent (which further implies that X_i and $(\Delta_i, W - \xi_i)$ are independent).

Lemma 1.5.1. *Under the conditions of Theorem 1.3.2, $\|\bar{\Delta}\|_q \leq \mathfrak{u}$, where \mathfrak{u} is as defined in (1.38).*

Lemma 1.5.2. *Under the conditions of Theorem 1.3.2, $\sum_{i=1}^n \|\xi_i\|_p \|\bar{\Delta} - \Delta_i\|_q \leq \sigma_p \mathfrak{v}$, where σ_p and \mathfrak{v} are as defined in (1.8) and (1.39), respectively.*

The proofs of these lemmas (and subsequent ones) are deferred to the end of this subsection.

Proof of Theorem 1.3.2. Recall that the conditions of Theorem 1.2.1 hold, with \tilde{T} in place of T , so that (1.112) and (1.6) imply

$$\begin{aligned} |\mathbb{P}(T \leq z) - \mathbb{P}(W \leq z)| &\leq \mathbb{P}(\|S\| > \epsilon) + |\mathbb{P}(\tilde{T} \leq z) - \mathbb{P}(W \leq z)| \\ &\leq \mathbb{P}(\|S\| > \epsilon) + \frac{1}{2c_*} \left(4\delta + \|W\|_p \|\bar{\Delta}\|_q + \sum_{i=1}^n \mathbb{E} \|\xi_i\|_p \|\bar{\Delta} - \Delta_i\|_q \right) + G_\eta(w) \end{aligned} \quad (1.117)$$

for all $z \in \mathbb{R}$. The use of [19, Remark 2.1] allows us to choose

$$\delta = c_4 \sigma_p^{\tilde{q}}, \quad (1.118)$$

in accordance with (1.4), where \tilde{q} and c_4 are as defined in (1.36). Along with (1.117), use Lemmas 1.5.1 and 1.5.2, and apply the Rosenthal-type inequality (1.31) to obtain $\|W\|_p \leq \mathfrak{A}_\mathbb{R}(p)\sigma_p + \mathfrak{B}_\mathbb{R}(p)$. Then (1.35) follows, and the proof of Theorem 1.3.2 is complete. \square

Proof of Theorem 1.3.5. Recall that the conditions of Theorem 1.2.4 hold with \tilde{T} in place of T . Also, by (1.113), (1.44), and (1.42),

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

Thus, by Remark 1.2.5, (1.18), (1.112), and (1.17),

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

for all z as in (1.44), where $\tilde{\gamma}_z$ is as in (1.46). Recall the definitions (1.20) and (1.47) of τ and $\tilde{\tau}$, respectively, to see that $\tau \leq \tilde{\tau}$ follows from (1.118) and Lemmas 1.5.1 and 1.5.2. Then (1.45) is proved. \square

The following lemma provides two bounds on $\tilde{\gamma}_z$ in (1.46) which will be used in the proofs of Corollary 1.3.7 and Theorem 1.A.2.

Lemma 1.5.3. *Assume that the conditions of Theorem 1.3.5 hold. Take any real numbers $\kappa_2 > 0$ and $\kappa_3 > 0$, and let*

$$x_2 := \left(\frac{2\pi_1}{M_\epsilon} \sigma z \right)^{1/2}, \quad y_2 := \frac{x_2}{\kappa_2}, \quad \varepsilon_2 := \frac{s_p^p}{s_2^2 y_2^{p-2}} \wedge 1, \quad S_{y_2} := \sum_{i=1}^n X_i \mathbb{I}\{\|X_i\| \leq y_2\}, \quad (1.119)$$

$$x_3 := \pi_3 z, \quad y_3 := \frac{x_3}{\kappa_3}, \quad \varepsilon_3 := \frac{\sigma_p^p}{y_3^{p-2}} \wedge 1, \quad (1.120)$$

$$\text{PU}_2 := \text{PU}_{\text{tail}}(x_2, y_2, s_2^2, \mathbb{E}\|S_{y_2}\|, \varepsilon_2), \quad \text{and} \quad \text{PU}_3 := \text{PU}_{\text{tail}}(x_3, y_3, 1, 0, \varepsilon_3), \quad (1.121)$$

where PU_{tail} is as in (1.12). Then

$$\tilde{\gamma}_z \leq G_X(y_2) + \text{PU}_2 + G_\xi(\pi_2 z) + (G_\xi(y_3) + \text{PU}_3)G_\eta(w) \quad (1.122)$$

for all $z > 0$, where $\tilde{\gamma}_z$ is as in (1.46).

One consequence of (1.122) is that

$$\begin{aligned} \tilde{\gamma}_z \leq G_X \left(\left(\frac{\pi_1}{2p^2 \omega M_\epsilon} \right)^{1/2} \sigma z \right) &+ \left(\frac{2epM_\epsilon D^2}{\pi_1} \frac{s_2^2}{\sigma z} \right)^p \\ &+ G_\xi(\pi_2 z) + \left(G_\xi \left(\frac{2\pi_3}{p} z \right) + \frac{(ep/2)^{p/2}}{(\pi_3 z)^p} \right) G_\eta(w) \end{aligned} \quad (1.123)$$

for all z as in (1.44).

In the proof of Corollary 1.3.7, let us write $a \leq b$ if $|a| \leq \mathfrak{C}b$ for some \mathfrak{C} as in Corollary 1.3.7. Let us then write $a \asymp b$ if $a \leq b$ and $b \leq a$.

Proof of Corollary 1.3.7. Set $c_* = \frac{1}{2}$, $w = \delta_0 = 1$, $\pi_1 = (M_\epsilon \epsilon^2 / (2\omega)) \wedge \frac{1}{3}$, $\pi_2 = \pi_3 = \frac{1}{2}(1 - \pi_1)$, and $\theta = \tilde{\theta}(1 - \pi_1)$ in the statements of Theorems 1.3.2 and 1.3.5, so that (1.16) and (1.42) be satisfied. Further let $X_i = \frac{1}{n}V_i$. Then $S = \sum_{i=1}^n X_i = \bar{V}$ and, by the definitions (1.34), (1.27), (1.8), and (1.40),

$$\sigma = \frac{\tilde{\sigma}}{n^{1/2}}, \quad s_\alpha = \frac{\|V\|_\alpha}{n^{1-1/\alpha}}, \quad \sigma_\alpha = \frac{\|L(V)\|_\alpha}{\tilde{\sigma} n^{1/2-1/\alpha}}, \quad \text{and} \quad \lambda_\alpha = \frac{\|L\| \|V\|_\alpha}{\tilde{\sigma} n^{1/2-1/\alpha}} \quad (1.124)$$

for any $\alpha \geq 1$. Recalling also the definitions (1.38), (1.39), (1.36), and (1.47), as well as Remark 1.2.7, one has

$$\mathfrak{u} \asymp n^{-1/2}, \quad \sigma_p \mathfrak{v} \asymp n^{-1/2}, \quad \sigma_p^{\tilde{q}} \asymp n^{-1/2}, \quad \text{and hence} \quad \tilde{\tau} \leq n^{-1/2} \quad (1.125)$$

for all $p \in (2, 3]$; moreover, it is clear that the above expressions depend on the distribution of V only through $\tilde{\sigma}$, $\|L(V)\|_p$, $\|V\|_q$, $\|V\|_2$, and $\|V\|_p$. Also, for any $t > 0$, (1.37) and (1.43) imply

$$G_X(t) = n \mathbb{P}(\|V\| > nt) \leq \frac{\|V\|_p^p}{n^{p-1} t^p} \quad (1.126)$$

$$\text{and } G_\xi(t) \leq G_\eta(t) \leq n \mathbb{P}(\|L\| \|V\| > \sqrt{n} \tilde{\sigma} t) \leq \frac{1}{n^{p/2-1} t^p}. \quad (1.127)$$

By (1.41), $\mathbb{P}(\|S\| > \epsilon) \leq \|V\|_2^2 / (\epsilon^2 n)$. Next, there exists a positive absolute constant A such that

$$\sup_{z \in \mathbb{R}} |P(\sqrt{n}L(\bar{V})/\tilde{\sigma} \leq z) - \Phi(z)| \leq A \frac{\|L(V)\|_p^p}{n^{p/2-1}},$$

which follows from, say, Theorem 6 of [97, Chapter V]. Then (1.35), (1.125), and (1.127) yield (1.48).

Using (1.125) and recalling that $\theta = (1 - \pi_1)\tilde{\theta}$, one has $\tilde{\tau} e^{-(1-\pi_1)z/\theta} \leq 1/(\sqrt{n} e^{z/\tilde{\theta}})$. In view of (1.45), (1.126), (1.127), and (1.123), one obtains (1.50) with $\Phi(z)$ there replaced by

$P(\sqrt{n}L(\bar{V})/\tilde{\sigma} \leq z)$. To obtain (1.50) as stated, note that [109, Corollary 1.3] implies

$$\left| P\left(\frac{L(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{\sigma_p^p}{e^{z/\tilde{\theta}}} + G_\xi\left(\frac{z}{1+p/2}\right) + \frac{G_\xi(1)}{z^p} \quad (1.128)$$

for all $z > 0$. Combining (1.128) with (1.124) and (1.127), one completes the proof. \square

Proof of Lemma 1.5.1. Suppose first that $p = 3$, so that, in accordance with (1.115), $\bar{\Delta} = \frac{M_\epsilon}{2\sigma} \|S\|^2$. Then, by the Rosenthal-type inequality (1.31) and the definitions (1.40) and (1.38) of λ_α and \mathfrak{u} , respectively,

$$\|\bar{\Delta}\|_q = \frac{M_\epsilon}{2\sigma} \|S\|_{2q}^2 \leq \frac{M_\epsilon}{2\sigma} (\mathfrak{A}_{\mathfrak{X}}^2(2q)s_{2q}^2 + \mathfrak{B}_{\mathfrak{X}}^2(2q)s_2^2) = \frac{M_\epsilon\sigma}{2\|L\|^2} (\mathfrak{A}_{\mathfrak{X}}^2(2q)\lambda_{2q}^2 + \mathfrak{B}_{\mathfrak{X}}^2(2q)\lambda_2^2) = \mathfrak{u},$$

which proves the lemma when $p = 3$.

Now suppose that $p \in (2, 3)$. By (1.114), (1.40) and (1.37),

$$\|\mathbb{E} \tilde{S}\| = \left\| \sum_i \mathbb{E} X_i \mathbf{1}_{\{\eta_i > w\}} \right\| \leq \sum_i \mathbb{E} \|X_i\| \mathbf{1}_{\{\eta_i > w\}} \leq w^{-(p-1)} \sum_i \mathbb{E} \|X_i\| \eta_i^{p-1} = \frac{\sigma w}{\|L\|} \lambda_p^p. \quad (1.129)$$

Let

$$\hat{X}_i := \tilde{X}_i - \mathbb{E} \tilde{X}_i \quad \text{and} \quad \hat{S} := \sum_i \hat{X}_i = \tilde{S} - \mathbb{E} \tilde{S},$$

so that

$$\|\hat{X}_i\|_\alpha \leq \|\tilde{X}_i\|_\alpha + \|\mathbb{E} \tilde{X}_i\| \leq 2\|\tilde{X}_i\|_\alpha \leq 2\|X_i\|_\alpha \quad (1.130)$$

for all $\alpha \geq 1$, and also

$$\|\hat{X}_i\|_\alpha \leq 2\|\tilde{X}_i\|_\alpha \leq 2\left(\frac{\sigma w}{\|L\|}\right)^{1-p/\alpha} \|X_i\|_\alpha^{p/\alpha} \quad (1.131)$$

for all $\alpha \geq p$. Then

$$\begin{aligned} \frac{2\sigma}{M_\epsilon} \|\bar{\Delta}\|_q &= \|\tilde{S}\|_{2q}^2 \leq (\|\hat{S}\|_{2q} + \|\mathbb{E} \tilde{S}\|)^2 \leq \frac{5}{4} \|\hat{S}\|_{2q}^2 + 5\|\mathbb{E} \tilde{S}\|^2 \\ &\leq \frac{5}{4} \left(\mathfrak{A}_{\mathfrak{X}}^2(2q) \left(\sum_i \|\hat{X}_i\|_{2q}^{2q} \right)^{1/q} + \mathfrak{B}_{\mathfrak{X}}^2(2q) \sum_i \|\hat{X}_i\|_2^2 \right) + 5 \left(\frac{\sigma w}{\|L\|} \lambda_p^p \right)^2 \\ &\leq 5 \left(\mathfrak{A}_{\mathfrak{X}}^2(2q) \left(\frac{\sigma w}{\|L\|} \right)^{2-p/q} s_p^{p/q} + \mathfrak{B}_{\mathfrak{X}}^2(2q) s_2^2 + \left(\frac{\sigma w}{\|L\|} \lambda_p^p \right)^2 \right) \\ &= 5 \left(\frac{\sigma w}{\|L\|} \right)^2 (\mathfrak{A}_{\mathfrak{X}}^2(2q) \lambda_p^{p-1} + \mathfrak{B}_{\mathfrak{X}}^2(2q) \lambda_2^2 + \lambda_p^{2p}) = \frac{2\sigma}{M_\epsilon} \mathfrak{u}, \end{aligned}$$

where the easily verified inequality $(x+y)^2 \leq \frac{5}{4}x^2 + 5y^2$ is used in the first line above, the Rosenthal-type inequality (1.31) and (1.129) are used in the second line, (1.130) and (1.131) are used in the third line, and the definitions (1.40) and (1.38) of λ_α and \mathfrak{u} , respectively, are used in the last line. This completes the proof of the lemma. \square

Proof of Lemma 1.5.2. Suppose first that $p = 3$. Then, by (1.115) and (1.116), for each $i = 1, \dots, n$,

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

Also, by (1.30), $\|S - X_i\|_q \leq \|S - X_i\|_2 \leq Ds_2$. It follows that

$$\|\bar{\Delta} - \Delta_i\|_q \leq \frac{M_\epsilon}{2\sigma} (\|X_i\|_{2q}^2 + 2\|X_i\|_q \|S - X_i\|_q) \leq \frac{M_\epsilon}{2\sigma} (\|X_i\|_{2q}^2 + 2Ds_2\|X_i\|_q), \quad (1.132)$$

So,

$$\begin{aligned} \sum_{i=1}^n \|\xi_i\|_p \|\bar{\Delta} - \Delta_i\|_q &\leq \frac{M_\epsilon}{2\sigma} \sum_{i=1}^n \|\xi_i\|_p (\|X_i\|_{2q}^2 + 2Ds_2\|X_i\|_q) \\ &\leq \frac{M_\epsilon}{2\sigma} \sigma_p (s_{2q}^2 + 2Ds_2s_q) = \sigma_p \mathbf{v}, \end{aligned}$$

where Hölder's inequality is used for the last inequality, and the definitions (1.40) and (1.39) are used for the equality. This proves the lemma when $p = 3$.

Suppose now that $p \in (2, 3)$. Similarly to (1.132) and using the truncation in the definition (1.114),

$$\|\bar{\Delta} - \Delta_i\|_q \leq \frac{M_\epsilon}{2\sigma} (\|\tilde{X}_i\|_{2q}^2 + 2\|\tilde{X}_i\|_q \|\tilde{S} - \tilde{X}_i\|_q) \leq \frac{M_\epsilon}{2\sigma} \left(\left(\frac{\sigma w}{\|L\|} \right)^{2-p/q} \|X_i\|_p^{p/q} + 2\|X_i\|_q \|\tilde{S} - \tilde{X}_i\|_2 \right);$$

also using (1.30) and (1.129), and reasoning as in (1.130), one has

$$\|\tilde{S} - \tilde{X}_i\|_2 \leq \|\hat{S} - \hat{X}_i\|_2 + \|\mathbf{E} \tilde{S} - \mathbf{E} \tilde{X}_i\| \leq 2Ds_2 + \frac{\sigma w}{\|L\|} \lambda_p^p = \frac{\sigma w}{\|L\|} (2D\lambda_2 + \lambda_p^p).$$

So,

$$\begin{aligned} \sum_{i=1}^n \mathbf{E} \|\xi_i\|_p \|\bar{\Delta} - \Delta_i\|_q &\leq \frac{M_\epsilon}{2\sigma} \sum_{i=1}^n \|\xi_i\|_p \left(\left(\frac{\sigma w}{\|L\|} \right)^{2-p/q} \|X_i\|_p^{p/q} + 2 \frac{\sigma w}{\|L\|} \|X_i\|_q (2D\lambda_2 + \lambda_p^p) \right) \\ &\leq \frac{M_\epsilon \sigma w^2}{2\|L\|^2} \sigma_p \left(\lambda_p^{p/q} + 2\lambda_q (2D\lambda_2 + \lambda_p^p) \right) = \sigma_p \mathbf{v}. \end{aligned}$$

Thus, the lemma is proved for $p \in (2, 3)$ as well. \square

Proof of Lemma 1.5.3. By (1.120), for each $i = 1, \dots, n$

$$\mathbf{P}(W - \xi_i \geq \pi_3 z) \leq \mathbf{P}\left(\max_{j \neq i} \xi_j > y_3\right) + \mathbf{P}\left(\sum_{j \neq i} \xi_j \mathbf{1}_{\{\xi_j \leq y_3\}} \geq x_3\right) \leq G_\xi(y_3) + \mathbf{PU}_3,$$

with the last inequality following from (1.12), (1.14), and the definition of \mathbf{PU}_3 in (1.121). A similar use of truncation, together with (1.29), (1.119), and (1.121), yields

$$\mathbf{P}\left(\|S\| > \left(\frac{2\pi_1 \sigma z}{M_\epsilon}\right)^{1/2}\right) = \mathbf{P}(\|S\| > x_2) \leq G_X(y_2) + \mathbf{P}(\|S_{y_2}\| > x_2) \leq G_X(y_2) + \mathbf{PU}_2.$$

Then (1.122) follows from the definitions (1.19) and (1.46) of γ_z and $\tilde{\gamma}_z$.

By (1.13) and the definition of $\mathbf{BH}_{\text{tail}}$ right after (1.14),

$$\begin{aligned} \mathbf{PU}_{\text{tail}}(x, y, B, m, \varepsilon) &\leq \mathbf{BH}_{\text{tail}}(x, y, B, m) \\ &= \exp\left\{\frac{(x-m)_+}{y} \left(1 - \left(1 + \frac{B^2}{(x-m)_+ y}\right) \ln\left(1 + \frac{(x-m)_+ y}{B^2}\right)\right)\right\} \\ &\leq \left(\frac{eB^2}{(x-m)_+ y}\right)^{(x-m)_+ / y} \wedge 1, \end{aligned} \quad (1.133)$$

where the equality is implied by [51, (2.9)]. Now let $\kappa_2 = 2p$ and $\kappa_3 = p/2$. Since $G_X(y_2) \leq G_X(y_2(\sigma z/\omega)^{1/2})$ whenever (1.44) is satisfied, (1.123) follows from (1.122) and (1.133) once it is demonstrated that

$$\text{PU}_2 \leq \left(\Lambda_1 \frac{s_2^2}{\sigma z} \right)^p, \quad \text{where } \Lambda_1 := \frac{2peM_\epsilon D^2}{\pi_1}. \quad (1.134)$$

Assume now that $\Lambda_1 s_2^2 \leq \sigma z$, since otherwise (1.134) trivially holds. Then

$$\begin{aligned} \mathbb{E}\|S_{y_2}\| &\leq \mathbb{E}\|S\| + \mathbb{E}\|S - S_{y_2}\| \leq \|S\|_2 + \mathbb{E}\left\|\sum_i X_i \mathbf{1}\{\|X_i\| > y_2\}\right\| \leq Ds_2 + \frac{s_2^2}{y_2} \\ &= \frac{x_2}{4} \left(\left(\frac{16D^2 s_2^2}{2\pi_1 \sigma z / M_\epsilon} \right)^{1/2} + \frac{8ps_2^2}{2\pi_1 \sigma z / M_\epsilon} \right) < \frac{x_2}{4} \left(\left(\Lambda_1 \frac{s_2^2}{\sigma z} \right)^{1/2} + \Lambda_1 \frac{s_2^2}{\sigma z} \right) \leq \frac{x_2}{2}, \end{aligned}$$

where (1.30) is used in the first line above, the definitions in (1.119) are used for the equality, and the inequalities $8 < 2pe$ and $4 < 2eD^2$ (which follow since $p \geq 2$ and $D \geq 1$) are used for the penultimate inequality. Thus, $\text{PU}_2 \leq \text{PU}_{\text{tail}}(x_2, y_2, s_2^2, x_2/2, \varepsilon_2)$ follows – cf. (1.14); (1.134) is then seen to hold after an application of (1.133). \square

1.5.3 Proofs of results from Section 1.4

Proof of Theorem 1.4.1. Note that the conditions of Theorem 1.3.2 hold when we set

$$X_i = V_i/n$$

and take any real $w > 0$. Then, recalling also that $p = 3$, one has

$$q = \frac{3}{2}, \quad G_\eta(w) = 0 \text{ (cf. (1.37) and (1.9))}, \quad \tilde{q} = 3, \quad \text{and} \quad c_4 = \frac{1}{4(1-c_*)} \text{ (cf. (1.36))}.$$

By (1.124), and in accordance with the notation (1.51),

$$\sigma = \frac{\tilde{\sigma}}{\sqrt{n}}, \quad \sigma_p = \frac{\varsigma_3}{n^{1/6}}, \quad s_\alpha = \frac{v_\alpha}{n^{1-1/\alpha}}, \quad \text{and} \quad \lambda_\alpha = \frac{\|L\|v_\alpha}{\tilde{\sigma}n^{1/2-1/\alpha}} \quad (1.135)$$

for any $\alpha \geq 1$. Further, use (1.32) and the assumption that \mathfrak{X} is a Hilbert space to let $D = 1$, $\mathfrak{A}_\mathbb{R}(p) = \mathfrak{A}_\mathfrak{X}(2q) = 1$, and $\mathfrak{B}_\mathbb{R}(p) = \mathfrak{B}_\mathfrak{X}(2q) = 3^{1/3}$. Then, in view of (1.38), (1.39), and the inequality $v_{3/2} \leq v_2$,

$$\mathfrak{u} = \frac{1}{\sqrt{n}} \frac{M_\epsilon}{2\tilde{\sigma}} \left(\frac{v_3^2}{n^{1/3}} + 3^{2/3} v_2^2 \right) \quad \text{and} \quad \sigma_p \mathfrak{v} \leq \frac{1}{\sqrt{n}} \frac{M_\epsilon}{2\tilde{\sigma}} \varsigma_3 \left(\frac{v_3^2}{n^{1/2}} + 2v_2^2 \right). \quad (1.136)$$

One also has $\mathbb{P}(\|S\| > \epsilon) \leq \mathfrak{K}_{\mathfrak{u}\epsilon}$ by Remark 1.3.3 and (1.32). Then (1.35), combined with (1.26) and the above substitutions and inequalities, yields (1.53). Using now Young's inequality

$$\varsigma_3^i v_\alpha^2 \leq \frac{1}{3} \frac{\varsigma_3^{3i}}{\kappa_{\alpha,i}^3} + \frac{2}{3} \kappa_{\alpha,i}^{3/2} v_\alpha^3 \text{ for } (\alpha, i) \in \{2, 3\} \times \{0, 1\}, \quad (1.137)$$

one deduces (1.54) from (1.53). \square

Proof of Corollary 1.4.4. Let $n \rightarrow \infty$. In accordance with (1.118) and (1.36), let $\delta = \frac{\sigma_3^3}{4(1-c_*)} = \frac{\varsigma_3^3}{4(1-c_*)\sqrt{n}}$, so that $\delta \rightarrow 0$. Following the lines of the proof of (1.53), one can see that the bound there equals

$$0.13925 + 0.33554 \frac{\varsigma_3^3}{\sqrt{n}} + \frac{4\delta}{2c_*} + \frac{\mathfrak{C}}{\sqrt{n}}, \quad (1.138)$$

where $\mathfrak{C} := (\mathfrak{K}_{u20} + \mathfrak{K}_{u21}\varsigma_3)v_2^2 + (\mathfrak{K}_{u30} + \mathfrak{K}_{u31}\varsigma_3)v_3^2 + \mathfrak{K}_{u\epsilon}$ is an upper bound on $\frac{1}{2c_*}(\mathbb{E}|W\overline{\Delta}| + \sum_i \mathbb{E}|\xi_i(\overline{\Delta} - \Delta_i)|) - \text{cf. (1.6)}$. Hence, by Remark 1.2.2, the term 4δ in the bound (1.138) 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 $\mathfrak{K}_{u1} = 0.33554 + \frac{1}{2c_*(1-c_*)}$ in (1.53) can be replaced by one asymptotic to $0.33554 + \frac{1}{4c_*(1-c_*)}$.

Let now $\epsilon = \epsilon_n = n^{-1/8}$; the assumed continuity of f'' implies $M_\epsilon \downarrow \|f''(0)\|$, and from (1.56) we see that $\mathfrak{K}_{u\epsilon} \downarrow 0$. Moreover, then $(\mathfrak{K}_{u20}, \mathfrak{K}_{u21}, \mathfrak{K}_{u30}, \mathfrak{K}_{u31}) \rightarrow \frac{M_\epsilon}{4c_*\tilde{\sigma}}(3, 2, 0, 0)$.

Thus,

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

Since $\min_{c_* \in (0,1)} (\frac{a}{c_*(1-c_*)} + \frac{b}{c_*}) = 2a + b + 2\sqrt{a(a+b)}$ whenever $a > 0$ and $b > 0$, (1.61) follows.

To prove (1.63), 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 (1.60) holds for all $z \in [g(n), \sqrt{n}/g(n)]$. Then, for $z \geq z_0$ and large enough n we have $z^3 e^{-z/\tilde{\theta}} \leq z_0^3 e^{-z_0/\tilde{\theta}} \rightarrow 0$. Concerning the pre-constants in Theorem 1.A.2 in Appendix 1.A, 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 in the second inequality (and, if so desired, in the other two inequalities) in (1.167) by the asymptotically much smaller expression $\mathbb{E}(\frac{L(V)}{\tilde{\sigma}})^3 \mathbb{I}\{L(V) > \pi_2 \tilde{\sigma} z \sqrt{n}\} = o(\varsigma_3^3)$. Then the limit of the corresponding improved expression for \mathfrak{K}_{n1} becomes just 1, instead of $1 + \pi_2^{-3}$. Now (1.63) follows by Theorem 1.A.2. \square

Proof of Corollary 1.4.10. Take any natural number $N_0 \geq 1$ and any real numbers $\epsilon \in (0, 1)$, $c_* \in (0, 1)$, $\kappa_1 > 0$, and $\kappa_2 > 0$, and let $X_i := Y_i$, $\xi_i := Y_i/\sqrt{n}$, $W := \sum_i \xi_i$, and $\overline{Y^2} := \frac{1}{n} \sum_i Y_i^2 = \sum_i \xi_i^2$. Further let $\tilde{T} := T_1 \mathbb{I}\{|\overline{Y^2} - 1| \leq \epsilon\} + W \mathbb{I}\{|\overline{Y^2} - 1| > \epsilon\}$, where $T_1 = W/\sqrt{\overline{Y^2}}$ is the self-normalized sum as defined in (1.72). Then

$$\begin{aligned} |\tilde{T} - W| &= \left| W \left(\frac{1}{\sqrt{\overline{Y^2}}} - 1 \right) \mathbb{I}\{|\overline{Y^2} - 1| \leq \epsilon\} \right| = |W(\overline{Y^2} - 1)| \frac{\mathbb{I}\{|\overline{Y^2} - 1| \leq \epsilon\}}{\sqrt{\overline{Y^2}}(1 + \sqrt{\overline{Y^2}})} \\ &\leq \check{M}_\epsilon |W(\overline{Y^2} - 1)|, \end{aligned}$$

where

$$\check{M}_\epsilon := \frac{1}{\sqrt{1-\epsilon}(1+\sqrt{1-\epsilon})}.$$

Accordingly, let

$$\overline{\Delta} := \Delta := \check{M}_\epsilon W(\overline{Y^2} - 1) \quad \text{and} \quad \Delta_i := \check{M}_\epsilon W_{(i)}(\overline{Y^2}_{(i)} - 1),$$

where $W_{(i)} = W - \xi_i$ and $\overline{Y^2}_{(i)} = \overline{Y^2} - \xi_i^2$. Then the conditions of Theorem 1.2.1 hold with \tilde{T} in place of T if we let $\eta_i = 0$ for $i = 1, \dots, n$ (and then allow w to take any positive value).

Recall that $\|Y\|_2 = 1$ is being assumed, whence $\|Y^2 - 1\|_2^2 = \|Y\|_4^4 - 1$; also, Young's inequality implies

$$\|Y\|_4^2 \|Y^2 - 1\|_2 \leq \frac{1}{2} \left(\frac{\|Y\|_4^4}{\kappa_1^2} + \kappa_1^2 \|Y^2 - 1\|_2^2 \right) = \frac{\kappa_1^2 + 1/\kappa_1^2}{2} \|Y\|_4^4 - \frac{\kappa_1^2}{2}$$

and

$$\|Y^2 - 1\|_2 \leq \frac{1}{2} \left(\frac{1}{\kappa_2^2} + \kappa_2^2 \|Y^2 - 1\|_2^2 \right) = \frac{\kappa_2^2}{2} \|Y\|_4^4 - \frac{\kappa_2^2 - 1/\kappa_2^2}{2}.$$

From $\|\overline{Y^2} - 1\|_2 = \|Y^2 - 1\|_2/\sqrt{n}$ and $\|W\|_4^4 = n \mathbb{E} \xi_1^4 + 3n(n-1)(\mathbb{E} \xi_1^2)^2 \leq 3 + \|Y\|_4^4/n$, it follows that

$$\mathbb{E}|W\overline{\Delta}| = \check{M}_\epsilon \mathbb{E} W^2 |\overline{Y^2} - 1| \leq \check{M}_\epsilon \|W\|_4^2 \|\overline{Y^2} - 1\|_2 \leq \frac{\check{M}_\epsilon}{\sqrt{n}} \|Y^2 - 1\|_2 (\sqrt{3} + \frac{1}{\sqrt{n}} \|Y\|_4^2).$$

Also, $\overline{\Delta} - \Delta_1 = \check{M}_\epsilon (\xi_1^2 W_{(1)} + \xi_1 (\overline{Y^2} - 1))$, whence

$$\begin{aligned} \mathbb{E}|\xi_1(\overline{\Delta} - \Delta_1)| &\leq \check{M}_\epsilon (\mathbb{E}|\xi_1|^3 \mathbb{E}|W_{(1)}| + \mathbb{E} \xi_1^2 |\overline{Y^2} - 1|) \leq \check{M}_\epsilon (\|\xi_1\|_3^3 \|W_{(1)}\|_2 + \|\xi_1\|_4^2 \|\overline{Y^2} - 1\|_2) \\ &\leq \frac{\check{M}_\epsilon}{n\sqrt{n}} (\|Y\|_3^3 + \|Y\|_4^2 \|Y^2 - 1\|_2). \end{aligned}$$

In the case where $n \geq N_0$, combine (1.26) and (1.6) (use also (1.118) with $p = 3$) to obtain

$$\begin{aligned} |\mathbb{P}(T_1 \leq z) - \Phi(z)| &\leq \mathbb{P}(|\overline{Y^2} - 1| > \epsilon) + |\mathbb{P}(\tilde{T} \leq z) - \mathbb{P}(W \leq z)| + |\mathbb{P}(W \leq z) - \Phi(z)| \\ &\leq \frac{\|Y\|_4^4 - 1}{\epsilon^2 n} + \frac{1}{2c_*} \left(\frac{\|Y\|_3^3}{(1-c_*)\sqrt{n}} + \mathbb{E}|W\overline{\Delta}| + n \mathbb{E}|\xi_1(\overline{\Delta} - \Delta_1)| \right) + \frac{0.13925 + 0.33554\|Y\|_3^3}{\sqrt{n}} \\ &\leq \frac{1}{\sqrt{n}} (A_3 \|Y\|_3^3 + A_4 \|Y\|_4^4 - A_0), \end{aligned}$$

where

$$\begin{aligned} A_3 &:= 0.33554 + \frac{1}{2c_*(1-c_*)} + \frac{\check{M}_\epsilon}{2c_*}, \\ A_4 &:= \frac{\check{M}_\epsilon}{4c_*} \left(\left(1 + \frac{1}{\sqrt{N_0}}\right) \left(\kappa_1^2 + \frac{1}{\kappa_1^2}\right) + \sqrt{3}\kappa_2^2 \right) + \frac{1}{\epsilon^2 \sqrt{N_0}}, \\ A_0 &:= \frac{\check{M}_\epsilon}{4c_*} \left(\left(1 + \frac{1}{\sqrt{N_0}}\right) \kappa_1^2 + \sqrt{3} \left(\kappa_2^2 - \frac{1}{\kappa_2^2}\right) \right) + \frac{1}{\epsilon^2 \sqrt{N_0}} - 0.13925. \end{aligned}$$

Then the inequality (1.77) holds for any of the triples in (1.78), in the case where $n \geq N_0$, when the parameter values in the table below (to be interpreted as rational numbers) are substituted into the expressions for A_3 , A_4 , and A_0 above:

A_3	A_4	A_0	N_0	ϵ	c_*	κ_1	κ_2
3.01	5.16	4.75	12	0.357	0.606	1.821	1.778
3.20	2.20	1.14	19	0.494	0.664	1.147	0.930
3.65	1.31	-1.45	41	0.565	0.755	1.020	0.479

In the case where $n < N_0$ (or hence $n \leq N_0 - 1$), it suffices to use the trivial bound $|\mathbf{P}(T_1 \leq z) - \Phi(z)| \leq \sqrt{N_0 - 1}/\sqrt{n}$ and then note that $\sqrt{N_0 - 1} \leq A_3 + A_4 - A_0 \leq A_3\|Y\|_3^3 + A_4\|Y\|_4^4 - A_0$ for any of the three triples (A_3, A_4, A_0) in the table above. \square

1.A An explicit nonuniform bound

In this appendix, we state and prove Theorem 1.A.2, which presents an explicit nonuniform BE-type bound for the normalized statistic $\sqrt{n}f(\bar{V})/\tilde{\sigma}$ when the summands V_i are i.i.d. The following lemma quotes expressions found in [120, 114] for the exponential bound PU_{tail} on the tail probability defined in (1.12). These expressions will be needed in applications of Theorem 1.A.2, wherein PU_{tail} enters the expressions for several pre-constants.

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

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

Then

$$\text{PU}_{\text{tail}}(x, y, B, m, \varepsilon) = \text{PU}_{\text{tail}}(u, \kappa, \varepsilon) := \begin{cases} 1 & \text{if } u = 0, \\ \text{PU}_{\text{alt}}(u, \kappa, \varepsilon) & \text{if } u > 0 \text{ and } \varepsilon < 1, \\ \text{BH}_{\text{alt}}(u, \kappa) & \text{if } u > 0 \text{ and } \varepsilon = 1, \end{cases} \quad (1.139)$$

where

$$\begin{aligned} \text{BH}_{\text{alt}}(u, \kappa) &:= \exp\left\{\kappa\left(1 - \left(1 + \frac{1}{u}\right)\ln(1 + u)\right)\right\}, \\ \text{PU}_{\text{alt}}(u, \kappa, \varepsilon) &:= \exp\left\{\frac{\kappa}{2(1 - \varepsilon)u} \left((1 - \varepsilon)^2 \left[1 + \text{W}\left(\frac{\varepsilon}{1 - \varepsilon} \exp \frac{\varepsilon + u}{1 - \varepsilon}\right)\right]^2 - (\varepsilon + u)^2 - (1 - \varepsilon^2) \right)\right\}, \end{aligned} \quad (1.140)$$

and 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 $z = we^w$); in (1.139), we allowed ourselves the slight abuse of notation, by using the same symbol, PU_{tail} , to denote two different functions, represented by two expressions, which take the same values but expressed using two different sequences of arguments: $(x, y, B, m, \varepsilon)$ and (u, κ, ε) .

One also has the alternative identity

$$\text{PU}_{\text{tail}}(u, \kappa, \varepsilon) = \inf_{0 < \alpha < 1} \exp\{L_1 \vee L_2\}, \quad (1.141)$$

where

$$L_1 := L_1(\alpha, u, \kappa, \varepsilon) := \kappa \left(1 - \alpha - \alpha \frac{\varepsilon}{1 - \varepsilon} - \frac{\alpha(2 - \alpha)}{2(1 - \varepsilon)} u\right), \quad \text{with} \quad L_1(\alpha, u, \kappa, 1) := -\infty, \quad (1.142)$$

and

$$L_2 := L_2(\alpha, u, \kappa, \varepsilon) := \kappa \left(1 - \alpha - \left(1 - \frac{\alpha}{2} + \frac{\varepsilon}{u}\right) \ln\left(1 + (1 - \alpha) \frac{u}{\varepsilon}\right)\right), \quad \text{with} \quad L_2(\alpha, 0, \kappa, \varepsilon) := 0. \quad (1.143)$$

Indeed, (1.139) is essentially [114, Proposition 3.1], with the “boundary” case $\varepsilon = 1$ resulting in the Bennett–Hoeffding bound $\text{BH}_{\text{alt}}(u, \kappa)$. Next, (1.141) (for $\varepsilon < 1$) is established in [120, Corollary 1] and, again, immediately follows for $\varepsilon = 1$ using $\text{BH}_{\text{alt}}(u, \kappa)$.

Theorem 1.A.2. Assume that the conditions of Theorem 1.4.1 hold, and let

$$c_*, \theta, w, \delta_0, \pi_1, \pi_2, \pi_3, z_0, \omega, \kappa_{2,0}, \kappa_{3,0}, \kappa_{2,1}, \kappa_{3,1}, \kappa_2, \kappa_3, \alpha, \varepsilon_*, K_1, K_2, \text{ and } K_3 \quad (1.144)$$

all be positive real numbers satisfying the constraints

$$c_* < 1, \delta_0 \leq w, \pi_1 + \pi_2 + \pi_3 = 1, \omega \leq \frac{M_\epsilon \epsilon^2}{2\pi_1}, \kappa_3 \geq \frac{3}{2}, \alpha < 1, \varepsilon_* < 1, \hat{\kappa}_2 \geq 2, \text{ and } \hat{\gamma} < 1, \quad (1.145)$$

where

$$\hat{\gamma} := \left(\frac{M_\epsilon^2 \omega}{4\pi_1^2 K_2} \right)^{1/4} + \frac{\kappa_2^2}{K_3} \left(\frac{M_\epsilon \omega}{2\pi_1} \right)^{3/2} \quad (1.146)$$

and

$$\hat{\kappa}_2 := (1 - \hat{\gamma})\kappa_2. \quad (1.147)$$

Also introduce

$$t_2 := \frac{\pi_1 \alpha (2 - \alpha) (1 - \hat{\gamma})^2}{M_\epsilon (1 - \varepsilon_*)} \left(\frac{K_2}{\omega} \right)^{1/2}, \quad (1.148)$$

$$t_3 := \frac{\kappa_2^2}{(1 - \hat{\gamma})K_3} \left(\frac{M_\epsilon \omega}{2\pi_1} \right)^{3/2}, \quad (1.149)$$

$$u_0 := \frac{2\pi_1 (1 - \hat{\gamma})}{M_\epsilon \kappa_2} \left(\frac{K_2}{\omega} \right)^{1/2}, \quad (1.150)$$

$$\tilde{\varepsilon}_1 := \frac{1}{K_1 w}, \quad \text{and} \quad \tilde{a}_1 := 1/L_{W;3w/\theta, \tilde{\varepsilon}_1}, \quad (1.151)$$

where $L_{W;c,B}$ is as in (1.15); further let \tilde{c}_1, \tilde{c}_2 , and \tilde{c}_3 be obtained from c_1, c_2 , and c_3 in (1.21)–(1.23) 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 (1.139). Then for all $z \in \mathbb{R}$ and $n \in \mathbb{N}$ such that

$$z_0 \leq z \leq \frac{\omega}{\tilde{\sigma}} \sqrt{n}, \quad (1.152)$$

$$\frac{K_1 \varsigma_3^3}{\sqrt{n}} \leq 1, \quad \frac{K_2 v_2^4}{\tilde{\sigma}^3 z^3 \sqrt{n}} \leq 1, \quad \text{and} \quad \frac{K_3 v_3^3}{\tilde{\sigma}^3 z^3 \sqrt{n}} \leq 1 \quad (1.153)$$

one has

$$\begin{aligned} & \left| \mathbb{P} \left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \\ & \leq \frac{\mathfrak{K}_{n1} \varsigma_3^3 + ((\mathfrak{K}_{n21} \vee \mathfrak{K}_{n22}) v_2^4) \vee (\mathfrak{K}_{n31} v_3^3) + \mathfrak{K}_{n32} v_3^3}{z^3 \sqrt{n}} + \frac{\mathfrak{K}_{e0} + \mathfrak{K}_{e1} \varsigma_3^3 + \mathfrak{K}_{e2} v_2^3 + \mathfrak{K}_{e3} v_3^3}{e^{(1-\pi_1)z/\theta} \sqrt{n}}, \end{aligned} \quad (1.154)$$

where

$$\mathfrak{K}_{n1} := 30.2211 + \frac{1}{\pi_2^3} + \frac{\kappa_3^{3/2}}{(w\pi_3)^3} \left(\frac{\kappa_3^{3/2}}{K_1} + \sup_{u \geq \pi_3^2 z_0^2 / \kappa_3} u^{3/2} \text{PU}_{\text{tail}} \left(u, \kappa_3, \frac{\kappa_3}{K_1 \pi_3 z_0} \wedge 1 \right) \right), \quad (1.155)$$

$$\mathfrak{K}_{n21} := \frac{\omega \exp\{\hat{\kappa}_2(1 - \alpha - \frac{\alpha\varepsilon_*}{1-\varepsilon_*})\}}{\tilde{\sigma}^3} \left(\frac{M_\epsilon(1 - \varepsilon_*)}{\pi_1 \alpha(2 - \alpha)(1 - \hat{\gamma})^2} \right)^2 \sup_{t \geq t_2} t^2 e^{-t}, \quad (1.156)$$

$$\mathfrak{K}_{n22} := \frac{\omega}{\tilde{\sigma}^3} \left(\frac{M_\epsilon \kappa_2}{2\pi_1(1 - \hat{\gamma})} \right)^2 \sup_{u \geq u_0} u^2 \text{PU}_{\text{tail}}(u, \hat{\kappa}_2, \varepsilon_*), \quad (1.157)$$

$$\mathfrak{K}_{n31} := \frac{\kappa_2^2 e^{\hat{\kappa}_2(1-\alpha)}}{\tilde{\sigma}^3(1 - \hat{\gamma})} \left(\frac{M_\epsilon \omega}{2\pi_1} \right)^{3/2} \sup_{t \in (0, t_3]} \frac{1}{t} \exp \left\{ -\hat{\kappa}_2 \left(1 - \frac{\alpha}{2} + t \right) \ln \left(1 + \frac{1 - \alpha}{t} \right) \right\}, \quad (1.158)$$

$$\mathfrak{K}_{n32} := \left(\frac{\kappa_2}{\tilde{\sigma}} \right)^3 \left(\frac{M_\epsilon \omega}{2\pi_1} \right)^{3/2}, \quad (1.159)$$

$$\mathfrak{K}_{e0} := \frac{M_\epsilon \tilde{c}_2}{6\tilde{\sigma}} \left(\frac{1}{\kappa_{3,0}^3 K_1^{2/3}} + \frac{3^{2/3}}{\kappa_{2,0}^3} \right), \quad (1.160)$$

$$\mathfrak{K}_{e1} := \frac{\tilde{c}_3}{4(1 - c_*)} + \frac{M_\epsilon \tilde{c}_1}{6\tilde{\sigma}} \left(\frac{1}{\kappa_{3,1}^3 K_1} + \frac{2}{\kappa_{2,1}^3} \right), \quad (1.161)$$

$$\mathfrak{K}_{e2} := \frac{M_\epsilon}{3\tilde{\sigma}} (2\tilde{c}_1 \kappa_{2,1}^{3/2} + 3^{2/3} \tilde{c}_2 \kappa_{2,0}^{3/2}), \quad (1.162)$$

$$\mathfrak{K}_{e3} := \frac{M_\epsilon}{3\tilde{\sigma}} \left(\frac{\tilde{c}_1 \kappa_{3,1}^{3/2}}{K_1} + \frac{\tilde{c}_2 \kappa_{3,0}^{3/2}}{K_1^{2/3}} \right); \quad (1.163)$$

moreover, each of the expressions in (1.155)–(1.163) is finite.

Remark 1.A.3. Suppose here that $L(V)$ is symmetric. Then the statement of Theorem 1.A.2 holds when the replacement mentioned in Remark 1.2.8 is made in the expression (1.22) for the pre-constant c_2 and, accordingly, in the expression for \tilde{c}_2 defined right after (1.151). Also, one can take \mathfrak{K}_{n1} in (1.154) to be defined as

$$\mathfrak{K}_{n1} := 30.2211 + \frac{1}{2\pi_2^3} + \frac{\kappa_3^{3/2}}{2(w\pi_3)^3} \left(\frac{\kappa_3^{3/2}}{2K_1} + \sup_{u \geq \pi_3^2 z_0^2 / \kappa_3} u^{3/2} \text{PU}_{\text{tail}} \left(u, \kappa_3, \frac{\kappa_3}{K_1 \pi_3 z_0} \wedge 1 \right) \right), \quad (1.164)$$

because one can then use $G_\eta(t) = G_\xi(t) \leq \varsigma_3^3 / (2t^3 \sqrt{n})$ in place of $G_\eta(t) = G_\xi(t) \leq \varsigma_3^3 / (t^3 \sqrt{n})$ to improve the bounds in (1.167) (in the proof of Theorem 1.A.2).

Remark 1.A.4. That all the pre-constants in Theorem 1.A.2 are finite is easily verifiable by inspection, except perhaps for the pre-constants \mathfrak{K}_{n1} , \mathfrak{K}_{n22} , and \mathfrak{K}_{n31} , whose expressions in (1.155), (1.157), and (1.158) involve comparatively complicated suprema. However, Lemma 1.C.1 in Appendix 1.C provides the sufficient conditions $\kappa_3 \geq \frac{3}{2}$ and $\hat{\kappa}_2 \geq 2$ in (1.145) for these three suprema, and hence for the pre-constants \mathfrak{K}_{n1} , \mathfrak{K}_{n22} , and \mathfrak{K}_{n31} , to be finite.

One can substantially improve the bound on $\tilde{\gamma}_z$ in (1.123). The following lemma is key to that, and its proof will be given after the proof of Theorem 1.A.2.

Lemma 1.A.5. *Assume that the conditions of Theorem 1.A.2 hold. Then, for all $z \in \mathbb{R}$*

and $n \in \mathbb{N}$ satisfying the inequalities in (1.152) and (1.153),

$$\text{PU}_2 \leq \frac{((\mathfrak{K}_{n21} \vee \mathfrak{K}_{n22})v_2^4) \vee (\mathfrak{K}_{n31}v_3^3)}{z^3\sqrt{n}}, \quad (1.165)$$

where PU_2 is as defined in (1.121) and \mathfrak{K}_{n21} , \mathfrak{K}_{n22} , and \mathfrak{K}_{n31} are as defined in (1.156), (1.157), and (1.158), respectively.

Proof of Theorem 1.A.2. Take any $z \in \mathbb{R}$ and $n \in \mathbb{N}$ such that (1.152) and (1.153) hold. The conditions of Theorem 1.3.5 are met when we let $p = 3$ and $X_i = V_i/n$, so that (1.26), (1.152), and (1.45) imply

$$\left| \mathbb{P}\left(\frac{f(\bar{V})}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{30.2211\varsigma_3^3}{z^3\sqrt{n}} + \tilde{\gamma}_z + \tilde{\tau}e^{-(1-\pi_1)z/\theta}. \quad (1.166)$$

Recall (1.43) to see that $\eta_i = \xi_i = L(V_i)/(\tilde{\sigma}\sqrt{n})$; then, for any $t > 0$, $G_\eta(t) = G_\xi(t) \leq \varsigma_3^3/(t^3\sqrt{n})$ and $G_X(t) \leq v_3^3/(t^3n^2)$ (cf. the inequalities (1.126) and (1.127)). Using these inequalities and also the first inequality of (1.153), (1.152), (1.13), and Lemma 1.A.1, one has

$$\begin{aligned} G_\eta(w) &\leq \frac{1}{w^3} \frac{\varsigma_3^3}{\sqrt{n}}, \quad G_\xi(\pi_2 z) \leq \frac{1}{\pi_2^3} \frac{\varsigma_3^3}{z^3\sqrt{n}}, \quad G_\xi(y_3) \leq \frac{\kappa_3^3}{\pi_3^3} \frac{\varsigma_3^3}{z^3\sqrt{n}} \leq \frac{\kappa_3^3}{K_1\pi_3^3} \frac{1}{z^3}, \\ G_X(y_2) &\leq \frac{v_3^3}{y_2^3 n^2} \frac{\omega^{3/2} n^{3/4}}{(\tilde{\sigma} z)^{3/2}} = \frac{\kappa_2^3 \omega^{3/2}}{\tilde{\sigma}^3 (2\pi_1/M_\epsilon)^{3/2}} \frac{v_3^3}{z^3\sqrt{n}}, \\ \text{PU}_3 &= \text{PU}_{\text{tail}}(x_3 y_3, \kappa_3, \varepsilon_3) \leq \frac{\kappa_3^3}{\pi_3^3 z^3} \sup_{u \geq \pi_2^2 z_0^2 / \kappa_3} u^{3/2} \text{PU}_{\text{tail}}\left(u, \kappa_3, \frac{\kappa_3}{K_1 \pi_3 z_0} \wedge 1\right). \end{aligned} \quad (1.167)$$

Then (1.122) and Lemma 1.A.5 yield

$$\frac{30.2211\varsigma_3^3}{z^3\sqrt{n}} + \tilde{\gamma}_z \leq \frac{1}{z^3\sqrt{n}} \left(\mathfrak{K}_{n1}\varsigma_3^3 + ((\mathfrak{K}_{n21} \vee \mathfrak{K}_{n22})v_2^4) \vee (\mathfrak{K}_{n31}v_3^3) + \mathfrak{K}_{n32}v_3^3 \right); \quad (1.168)$$

where $\mathfrak{K}_{n1}, \dots, \mathfrak{K}_{n32}$ are as in (1.155)–(1.159).

Next, in the definitions (1.21)–(1.23) and (1.25), set $p = 3$, $\mathfrak{A}_\mathbb{R}(p) = 1$, and $\mathfrak{B}_\mathbb{R}(p) = 3^{1/3}$ – recall here (1.32). Also, by the first inequality of (1.153) and (1.24), $\sigma_p = \varsigma_3 n^{-1/6} \leq K_1^{-1/3}$, $\varepsilon_1 \leq 1 \wedge (K_1 w)^{-1}$, and $\|\xi_i\|_2/w = 1/(w\sqrt{n}) \leq \varsigma_3^3/(w\sqrt{n}) \leq 1/(K_1 w)$. Then, referring to (1.151), we see that $a_1 \leq \tilde{a}_1$ (as $L_{W;c,\sigma}$ is nonincreasing with respect to σ) and $c_j \leq \tilde{c}_j$ for $j = 1, 2, 3$. Using the definition (1.47) of $\tilde{\tau}$, as well as (1.136) and (1.137), one obtains the inequalities

$$\begin{aligned} \tilde{\tau} &\leq \frac{M_\epsilon}{2\tilde{\sigma}\sqrt{n}} \left(\tilde{c}_1 \varsigma_3 \left(\frac{v_3^2}{\sqrt{n}} + 2v_2^2 \right) + \tilde{c}_2 \left(\frac{v_3^2}{n^{1/3}} + 3^{2/3} v_2^2 \right) \right) + \frac{\tilde{c}_3}{4(1-c_*)} \frac{\varsigma_3^3}{\sqrt{n}} \\ &\leq \frac{1}{\sqrt{n}} \left(\mathfrak{K}_{e0} + \mathfrak{K}_{e1}\varsigma_3^3 + \mathfrak{K}_{e2}v_2^3 + \mathfrak{K}_{e3}v_3^3 \right), \end{aligned} \quad (1.169)$$

where $\mathfrak{K}_{e0}, \dots, \mathfrak{K}_{e3}$ are as in (1.160)–(1.163); here, the first inequality of (1.153) is again used to see that $n \geq K_1^2 \varsigma_3^6 \geq K_1^2$.

Combine now the inequalities (1.166), (1.168), and (1.169); then (1.154) follows. \square

Proof of Lemma 1.A.5. As we have let $X_i = V_i/n$ and $p = 3$ in Theorem 1.A.2, (1.135) holds. Let now

$$c_x := \left(\frac{2\pi_1}{M_\epsilon} \right)^{1/2}, \text{ so that } x_2 = \frac{c_x(\tilde{\sigma}z)^{1/2}}{n^{1/4}} \text{ and } y_2 = \frac{c_x(\tilde{\sigma}z)^{1/2}}{\kappa_2 n^{1/4}},$$

by (1.119). Then

$$\begin{aligned} \mathbb{E}\|S_{y_2}\| &\leq \mathbb{E}\|S\| + \mathbb{E}\|S_{y_2} - S\| \leq \|S\|_2 + \sum_i \mathbb{E}\|X_i\| \mathbb{1}\{\|X_i\| > y_2\} \leq s_2 + \frac{s_3^3}{y_2^2} = \frac{v_2}{\sqrt{n}} + \frac{v_3^3}{y_2^2 n^2} \\ &= x_2 \left(\frac{v_2}{c_x(\tilde{\sigma}z)^{1/2} n^{1/4}} + \frac{\kappa_2^2 v_3^3}{c_x^3(\tilde{\sigma}z)^{3/2} n^{5/4}} \right) \\ &= x_2 \left(\frac{1}{c_x} \left(\frac{v_2^4}{\tilde{\sigma}^3 z^3 \sqrt{n}} \right)^{1/4} \left(\frac{\tilde{\sigma}z}{\sqrt{n}} \right)^{1/4} + \frac{\kappa_2^2}{c_x^3} \frac{v_3^3}{\tilde{\sigma}^3 z^3 \sqrt{n}} \left(\frac{\tilde{\sigma}z}{\sqrt{n}} \right)^{3/2} \right) \\ &\leq x_2 \left(\frac{1}{c_x} \left(\frac{\omega}{K_2} \right)^{1/4} + \frac{\kappa_2^2 \omega^{3/2}}{c_x^3 K_3} \right) = \hat{\gamma} x_2, \end{aligned}$$

where (1.152) and (1.153) are used to obtain the last inequality above, and the definition (1.146) of $\hat{\gamma}$ is used for the last equality. Then, since $\hat{\gamma} < 1$ is assumed in (1.145), Lemma 1.A.1 yields

$$\text{PU}_2 \leq \text{PU}_{\text{tail}}(\hat{u}, \hat{\kappa}_2, \varepsilon_2), \text{ where } \hat{u} := \frac{(1-\hat{\gamma})x_2 y_2}{s_2^2} = \frac{c_x^2(1-\hat{\gamma})}{\kappa_2} \frac{\tilde{\sigma}z}{v_2^2} \sqrt{n} \text{ and } \hat{\kappa}_2 := (1-\hat{\gamma})\kappa_2 \quad (1.170)$$

(recall (1.14)). Also, in accordance with (1.119), $\varepsilon_2 = \frac{s_3^3}{s_2^2 y_2} \wedge 1 = \frac{\kappa_2 v_3^3}{c_x v_2^2 (\tilde{\sigma}z)^{1/2} n^{3/4}} \wedge 1$.

The inequality in (1.165) is proved by taking any $\varepsilon_* \in (0, 1)$, as in Theorem 1.A.2, and 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 (1.170) and (1.141),

$$\text{PU}_2 \leq \text{PU}_{\text{tail}}(\hat{u}, \hat{\kappa}_2, \varepsilon_2) \leq \exp\{L_1(\alpha, \hat{u}, \hat{\kappa}_2, \varepsilon_2)\} \vee \exp\{L_2(\alpha, \hat{u}, \hat{\kappa}_2, \varepsilon_2)\} \quad (1.171)$$

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

$$\begin{aligned} r_2^2 &:= \frac{1}{\hat{u}} = \frac{\kappa_2}{c_x^2(1-\hat{\gamma})} \frac{v_2^2}{\tilde{\sigma}z\sqrt{n}} = \frac{\kappa_2}{c_x^2(1-\hat{\gamma})} \left(\frac{v_2^4}{\tilde{\sigma}^3 z^3 \sqrt{n}} \right)^{1/2} \left(\frac{\tilde{\sigma}z}{\sqrt{n}} \right)^{1/2} \\ &\leq \frac{\kappa_2 \omega^{1/2}}{c_x^2(1-\hat{\gamma})} \left(\frac{v_2^4}{\tilde{\sigma}^3 z^3 \sqrt{n}} \right)^{1/2} \quad (1.172) \end{aligned}$$

$$\leq \frac{\kappa_2 \omega^{1/2}}{K_2^{1/2} c_x^2(1-\hat{\gamma})} = \frac{1}{u_0} \quad (1.173)$$

and

$$\begin{aligned} r_3^3 &:= \frac{\varepsilon_2}{\hat{u}} \leq \frac{\kappa_2 v_3^3}{c_x v_2^2 (\tilde{\sigma} z)^{1/2} n^{3/4}} \frac{\kappa_2 v_2^2}{(1 - \hat{\gamma}) c_x^2 \tilde{\sigma} z \sqrt{n}} = \frac{\kappa_2^2}{(1 - \hat{\gamma}) c_x^3} \frac{v_3^3}{\tilde{\sigma}^3 z^3 \sqrt{n}} \left(\frac{\tilde{\sigma} z}{\sqrt{n}} \right)^{3/2} \\ &\leq \frac{\kappa_2^2 \omega^{3/2}}{(1 - \hat{\gamma}) c_x^3} \frac{v_3^3}{\tilde{\sigma}^3 z^3 \sqrt{n}} \end{aligned} \quad (1.174)$$

$$\leq \frac{\kappa_2^2 \omega^{3/2}}{(1 - \hat{\gamma}) c_x^3 K_3} = t_3, \quad (1.175)$$

where (1.152) is used to establish the inequalities in (1.172) and (1.174), and (1.153), (1.150), and (1.149) are used for (1.173) and (1.175).

Next, in view of (1.173), (1.147), and (1.148), one has

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

So, the case condition $\varepsilon_2 \in (\varepsilon_*, 1]$ together with the definitions of (1.142) and (1.172) of L_1 and r_2^2 imply

$$e^{L_1} \leq e^{\hat{\kappa}_2 (1 - \alpha - \alpha \varepsilon_*/(1 - \varepsilon_*))} \left(\frac{2(1 - \varepsilon_*)}{\hat{\kappa}_2 \alpha (2 - \alpha)} \right)^2 \left(\sup_{t \geq t_2} t^2 e^{-t} \right) r_2^4 \leq \mathfrak{R}_{n21} \frac{v_2^4}{z^3 \sqrt{n}}, \quad (1.176)$$

where the last inequality follows by the definition (1.156) of \mathfrak{R}_{n21} and (1.172) (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 (1.176) trivial (using the convention $\exp\{-\infty\} := 0$).

Again by the case condition $\varepsilon_2 \in (\varepsilon_*, 1]$, now together with (1.143) and (1.175),

$$e^{L_2} \leq e^{\hat{\kappa}_2 (1 - \alpha)} \left(\sup_{t \in (0, t_3]} \frac{1}{t} \exp \left\{ -\hat{\kappa}_2 \left(1 - \frac{\alpha}{2} + t \right) \ln \left(1 + \frac{1 - \alpha}{t} \right) \right\} \right) r_3^3 \leq \mathfrak{R}_{n31} \frac{v_3^3}{z^3 \sqrt{n}}, \quad (1.177)$$

where the last inequality follows by the definition (1.158) of \mathfrak{R}_{n31} and (1.174). Now, upon combining (1.171), (1.176), and (1.177), we obtain the result (1.165) in the case $\varepsilon_2 \in (\varepsilon_*, 1]$.

Consider the remaining case, when $\varepsilon_2 \in (0, \varepsilon_*]$. Then, by (1.13), (1.170), (1.172), (1.173), and the definition (1.157) of \mathfrak{R}_{n22} ,

$$\text{PU}_2 \leq \text{PU}_{\text{tail}}(\hat{u}, \hat{\kappa}_2, \varepsilon_2) \leq \text{PU}_{\text{tail}}(\hat{u}, \hat{\kappa}_2, \varepsilon_*) \leq r_2^4 \left(\sup_{u \geq u_0} u^2 \text{PU}_{\text{tail}}(u, \hat{\kappa}_2, \varepsilon_*) \right) \leq \mathfrak{R}_{n22} \frac{v_2^4}{z^3 \sqrt{n}}. \quad (1.178)$$

Thus, (1.178) yields (1.165) in the case $\varepsilon_2 \in (0, \varepsilon_*]$ as well, and the lemma is proved. \square

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

The following proposition shows that the upper bound on z in (1.49), and hence in (1.44), is in general optimal, up to the choice of the constant factor ω .

Proposition 1.B.1. *Let $\mathfrak{X} = \mathbb{R}$ and $f(x) \equiv x + x^2$, so that (1.33) is satisfied when $L(x) \equiv x$, $M_\epsilon = 2$, and $\epsilon = 1$. For any $p \in (2, 3]$, let V, V_1, \dots, V_n 's be real-valued symmetric i.i.d. r.v.'s with density $|v|^{-p-1} \ln^{-2} |v|$ 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 $\|V\|_p < \infty$. 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 (1.50) 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 (1.49) holds for all pairs $(n, z) \in \mathbf{NZ}$.*

Remark 1.B.2. Let $r \in (0, p)$. Then an application of Chebyshev's inequality to the first two terms in the bound of (1.50) yields

$$\begin{aligned} & \left| \mathbb{P} \left(\frac{f(\bar{V})}{\bar{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \\ & \leq \mathfrak{C} \left(\frac{\mathbb{E} \|V\|^r \mathbb{I} \{ \|V\| > \mathfrak{C} \sqrt{n} z \}}{n^{r/2-1} z^r} + \frac{\mathbb{E} \|V\|^r \mathbb{I} \{ \|V\| > \mathfrak{C} \sqrt{n} \}}{n^{r/2-1} z^r} + \frac{1}{(\sqrt{n} z)^p} + \frac{1}{n^{p/2-1} e^{z/\bar{\theta}}} \right) \end{aligned} \quad (1.179)$$

for any z satisfying (1.49). The arguments of the proof of Proposition 1.B.1 can be used to demonstrate that the bound of (1.179) (larger than that in (1.50)) generally fails to hold if $z/\sqrt{n} \rightarrow \infty$. Using Chebyshev's inequality when $r = p$ yields

$$\left| \mathbb{P} \left(\frac{f(\bar{V})}{\bar{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \leq \frac{\mathfrak{C}}{n^{p/2-1} z^p}. \quad (1.180)$$

One might hope that a bound of the form in (1.180) could hold for all f satisfying the smoothness condition (1.33) and for all $z > 0$. However, another modification of the proof of Proposition 1.B.1 (which will be given in Section 1.5) demonstrates that (1.180) fails to be true whenever

$$\frac{z}{\sqrt{n} \ln^\alpha n} \rightarrow \infty, \quad \text{where } \alpha \text{ is any fixed number such that } \alpha p > 1; \quad (1.181)$$

the extra log factor above is needed because the bound in (1.180) is worse than that in (1.179).

Proof of Proposition 1.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 1.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 $z > \frac{\omega}{\sigma} \sqrt{n}$. Now, for the rest of the proof of Proposition 1.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 $v > v_0$

$$\mathbf{P}(V > v) = \int_v^\infty \frac{du}{u^{p+1} \ln^2 u} \asymp \frac{1}{v^p \ln^2 v}$$

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

So,

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

and

$$1 - \Phi(z) \asymp \frac{1}{ze^{z^2/2}} = \frac{\zeta^{p/2-1} n^{p-3/2} \ln^2(\zeta^{1/2}n)}{e^{\zeta^2 n/2}} \frac{n}{\vartheta^p \ln^2 \vartheta} = o(n \mathbf{P}(V > \vartheta)). \quad (1.182)$$

Then (1.50) and (1.128) imply that $|\mathbf{P}(T \leq z) - \Phi(z)|$ and $|\mathbf{P}(W \leq z) - \Phi(z)|$ are both $o(n \mathbf{P}(V > \vartheta))$. Now let $\Delta = T - W = \sqrt{n}S^2$, so that

$$\mathbf{P}(\Delta > 2z) \leq \mathbf{P}(T > z) + \mathbf{P}(-W > z) = \mathbf{P}(T > z) + \mathbf{P}(W > z) = o(n \mathbf{P}(V > \vartheta)), \quad (1.183)$$

by (1.182).

On the other hand, by [28, Lemma 2.3],

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

for large enough n , where

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

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

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

for large enough n , which contradicts (1.183).

The statements of Remark 1.B.2 are proved with only a few modifications to the above arguments, using the relation

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

as $v \rightarrow \infty$, for any $r \in (0, p)$. In order to show that (1.180) fails to hold simultaneously with (1.181), let V have density $1/(|v|^{p+1} \ln^{\alpha p} |v|)$ 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

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

from which (1.183) follows and the contradiction is derived as done previously. \square

1.C Proofs of bounds with explicit numerical constants, using a computer algebra system (CAS)

This appendix contains proofs of Corollaries 1.4.11, 1.4.12, and 1.4.20. The numerical computations that arise in these proofs are easily performed with a CAS; of course the calculations could, in principle, be done without the aid of a computer, but the amount of time required for such a task makes the use of a CAS practically indispensable.

Proof of Corollary 1.4.11. Consider the i.i.d. r.v.'s $V := (Y, Y^2 - 1)$ and $V_i := (Y_i, Y_i^2 - 1)$, taking values in $\mathfrak{X} = \mathbb{R}^2$ with the standard Euclidean norm, and let $f(\mathbf{x}) := x_1/\sqrt{1+x_2}$ for $\mathbf{x} = (x_1, x_2) \in \mathfrak{X}$ with $x_2 > -1$ (also let $f(\mathbf{x})$ take an arbitrary value for all other $\mathbf{x} \in \mathfrak{X}$). Further let $L = f'(0)$, so that $\|L\| = 1$, $L(V) = Y$, and $\tilde{\sigma} = \|L(V)\|_2 = 1$. Then $\sqrt{n}f(\bar{V})/\tilde{\sigma} = T_1$ a.s., by (1.72). On recalling (1.64), it is clear that f satisfies the smoothness condition (1.33) whenever $\epsilon < 1$, whence the conditions of Theorem 1.4.1 hold.

For any $\mathbf{x} \in \mathfrak{X}$ such that $\|\mathbf{x}\| \leq \epsilon < 1$, the spectral norm of the Hessian matrix $f''(\mathbf{x})$ is

$$\|f''(\mathbf{x})\| = \left| \frac{3x_1 + \sqrt{9x_1^2 + 16(1+x_2)^2}}{8(1+x_2)^{5/2}} \right| \vee \left| \frac{3x_1 - \sqrt{9x_1^2 + 16(1+x_2)^2}}{8(1+x_2)^{5/2}} \right|.$$

It is easy to see that $\|f''(\mathbf{x})\|$ is symmetric with respect to x_1 ; moreover, $\|f''(\mathbf{x})\|$ is increasing in $x_1 \geq 0$ and decreasing in x_2 . Hence,

$$M_\epsilon = \sup_{\|\mathbf{x}\| \leq \epsilon} \|f''(\mathbf{x})\| = \sup_{\|\mathbf{x}\| = \epsilon} \|f''(\mathbf{x})\| = \sup_{-\epsilon \leq x_2 \leq 0} \frac{3\sqrt{\epsilon^2 - x_2^2} + \sqrt{9(\epsilon^2 - x_2^2) + 16(1+x_2)^2}}{8(1+x_2)^{5/2}}; \quad (1.184)$$

given some specific rational ϵ , a CAS can be used to obtain an algebraic expression for M_ϵ .

Next, introduce

$$y_3 := \|Y\|_3, \quad y_4 := \|Y\|_4, \quad \text{and} \quad y_6 := \|Y^2 - 1\|_3^{1/2}; \quad (1.185)$$

then (1.51) yields

$$\varsigma_3 = y_3, \quad v_2 = y_4^2, \quad \text{and} \quad v_3 = \|Y^2 + (Y^2 - 1)^2\|_{3/2}^{1/2}. \quad (1.186)$$

For any nonnegative numbers \tilde{w}_0 , \tilde{w}_3 , and \tilde{w}_4 , let

$$\nu_3 := \nu_3(\tilde{w}_0, \tilde{w}_3, \tilde{w}_4) := \sup_{y \in \mathbb{R}} \frac{(y^2 + (y^2 - 1)^2)^{3/2}}{\tilde{w}_0(1 - y^2) + \tilde{w}_3|y|^3 + \tilde{w}_4y^4 + |y^2 - 1|^3}, \quad (1.187)$$

so that (1.185) and (1.186) imply

$$v_3^3 \leq \nu_3 \cdot (\tilde{w}_3y_3^3 + \tilde{w}_4y_4^4 + y_6^6); \quad (1.188)$$

note that, whenever the numbers \tilde{w}_0 , \tilde{w}_3 , and \tilde{w}_4 happen to be such that the denominator in (1.187) is negative for some $y \in \mathbb{R}$, then necessarily $\nu_3(\tilde{w}_0, \tilde{w}_3, \tilde{w}_4) = \infty$ and the inequality in (1.188) is trivially satisfied.

Introduce arbitrary positive parameters $N_0 \in \mathbb{N}$, w_4 , and w_6 . Consider two cases: (i) $n \leq N_0 - 1$ and (ii) $n \geq N_0$. In the first case, when $n \leq N_0 - 1$, use the inequalities

$y_4 \geq y_3 \geq 1$ to see that

$$|\mathbf{P}(T_1 \leq z) - \Phi(z)| \leq 1 \leq \frac{\sqrt{N_0 - 1}}{\sqrt{n}} \leq \frac{1}{\sqrt{n}} (A_{3,1}y_3^3 + A_{4,1}y_4^6 + A_{6,1}y_6^6), \quad (1.189)$$

where

$$A_{3,1} := \frac{\sqrt{N_0 - 1}}{1 + w_4}, \quad A_{4,1} := \frac{w_4 \sqrt{N_0 - 1}}{1 + w_4}, \quad \text{and} \quad A_{6,1} := 0.$$

Consider then the case when $n \geq N_0$, and let c_* , $\kappa_{2,0}$, $\kappa_{3,0}$, $\kappa_{2,1}$, and $\kappa_{3,1}$ be as in (1.52). Further let $w_{6,2} := 1$, take any nonnegative numbers $w_{0,2}$, $w_{3,2}$, and $w_{4,2}$ (to be specified later), and let

$$\nu_{j,2} := \nu_3(w_{0,2}, w_{3,2}, w_{4,2})w_{j,2} \quad \text{for } j \in \{3, 4, 6\}, \quad \text{so that} \quad v_3^3 \leq \nu_{3,2}y_3^3 + \nu_{4,2}y_4^6 + \nu_{6,2}y_6^6, \quad (1.190)$$

by (1.188). Then (1.54) and (1.190) imply

$$|\mathbf{P}(T_1 \leq z) - \Phi(z)| \leq \frac{1}{\sqrt{n}} (\tilde{\mathfrak{K}}_{u0} + \tilde{\mathfrak{K}}_{u1}y_3^3 + \tilde{\mathfrak{K}}_{u2}y_4^6 + \tilde{\mathfrak{K}}_{u3}v_3^3) \leq \frac{1}{\sqrt{n}} (A_{3,2}y_3^3 + A_{4,2}y_4^6 + A_{6,2}y_6^6) \quad (1.191)$$

where $\tilde{\mathfrak{K}}_{u0}, \dots, \tilde{\mathfrak{K}}_{u3}$ are as in (1.57), but with N_0 replacing each instance of n in those expressions,

$$A_{3,2} := \pi \tilde{\mathfrak{K}}_{u0} + \tilde{\mathfrak{K}}_{u1} + \nu_{3,2} \tilde{\mathfrak{K}}_{u3}, \quad A_{4,2} := (1 - \pi) \tilde{\mathfrak{K}}_{u0} + \tilde{\mathfrak{K}}_{u2} + \nu_{4,2} \tilde{\mathfrak{K}}_{u3}, \quad A_{6,2} := \nu_{6,2} \tilde{\mathfrak{K}}_{u3},$$

and π is any number in the interval $[0, 1]$. Now choose π to minimize $A_{3,2} \vee (A_{4,2}/w_4)$ subject to the constraint that $\pi \in [0, 1]$; that is, let

$$\pi := 1 \wedge \left(\frac{\tilde{\mathfrak{K}}_{u0} + \tilde{\mathfrak{K}}_{u2} + \nu_{4,2} \tilde{\mathfrak{K}}_{u3} - w_4(\tilde{\mathfrak{K}}_{u1} + \nu_{3,2} \tilde{\mathfrak{K}}_{u3})}{\tilde{\mathfrak{K}}_{u0}(1 + w_4)} \right)_+.$$

Referring now to (1.189) and (1.191), we see that (1.79) holds when

$$A_j := A_{j,1} \vee A_{j,2} \quad \text{for } j \in \{3, 4, 6\}.$$

As mentioned before, the two triples (A_3, A_4, A_6) in (1.80) are the result of trying to approximately minimize $A_3 \vee (A_4/w_4) \vee (A_6/w_6)$, with weights $(w_4, w_6) \in \{1, 0.25\} \times \{0.05\}$. Using a CAS to find exact expressions for M_ϵ in (1.184) and ν_3 in (1.187), and substituting the parameter values given in Table 1.3 below (which should be interpreted as exact, rational numbers), one can verify that (1.79) indeed holds with the specific values of the triples (A_3, A_4, A_6) listed in (1.80).

To prove that the set in (1.86) can replace that in (1.80) when the symmetry of Y is assumed, one need only redefine $\mathfrak{K}_{u20}, \dots, \mathfrak{K}_{u31}$ according to Remark 1.4.2, and then use the set of parameter values listed in Table 1.3. \square

Proof of Corollary 1.4.12. Adopt the notation used in the proof of Corollary 1.4.11; particularly recall (1.185) and (1.186)). Recall also the positive parameters in (1.144) satisfying the constraints in (1.145); we shall specify their values later in the proof. In

Table 1.3: Parameters associated with triples (A_3, A_4, A_6) in (1.80) and (1.86)

	(1.80)		(1.86)	
A_3	3.33	5.79	3.09	5.32
A_4	3.33	1.45	3.09	1.33
A_6	0.17	0.26	0.16	0.27
w_4	1	0.25	1	0.25
w_6	0.05	0.05	0.05	0.05
ϵ	0.346	0.352	0.370	0.377
N_0	44	53	39	45
c_*	0.684	0.744	0.669	0.726
$\kappa_{2,0}$	1	0.762	1	0.771
$\kappa_{3,0}$	0.870	0.919	0.905	0.886
$\kappa_{2,1}$	1	0.768	1	0.781
$\kappa_{3,1}$	0.874	0.939	0.904	0.931
$w_{0,2}$	0.392	0	0.392	0
$w_{3,2}$	0	0.629	0	0.498
$w_{4,2}$	1	0	1	0

addition, take any

$$\epsilon \in (0, 1), \quad c \in (0, 1), \quad \pi_4 \in [0, 1],$$

$$w_{j,k} \geq 0 \quad ((j, k) \in \{0, 3, 4\} \times \{2, 3\}), \quad \text{and} \quad w_{g,k} \geq 0 \quad (k \in \{1, 2, 3\}).$$

Then let (cf. (1.187) and (1.188))

$$\nu_{j,k} := w_{j,k} \nu_3(w_{0,k}, w_{3,k}, w_{4,k}) \quad \text{for } (j, k) \in \{3, 4, 6\} \times \{2, 3\}, \quad (1.192)$$

so that $v_3^3 \leq \nu_{3,k} y_3^3 + \nu_{4,k} y_4^8 + \nu_{6,k} y_6^6$, where $w_{6,k} := 1$ for $k \in \{2, 3\}$, and also let (cf. (1.81))

$$g_k(z) := \frac{1}{z^3} + \frac{w_{g,k}}{e^{z/\tilde{\theta}}} \quad \text{for } k \in \{1, 2, 3\}, \quad \text{where } \tilde{\theta} := \frac{\theta}{1 - \pi_1}. \quad (1.193)$$

Similarly to the proof of [111, Theorem 1.1], consider three cases.

Case 1 (“small z ”): $0 < z < z_0$.

Let (A_3, A_4, A_6) be any triple of constants such that (1.76) holds; we shall provide specific values for the triple (A_3, A_4, A_6) at the end of the proof, using general expressions obtained in [111, Theorem 1.2]. Since g_1 in (1.193) is decreasing on $(0, \infty)$, (1.76) and the case condition $0 < z < z_0$ then imply

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq \frac{1}{\sqrt{n}} (A_3 y_3^3 + A_4 y_4^8 + A_6 y_6^6) \leq \frac{g_1(z)}{\sqrt{n}} (A_{3,1} y_3^3 + A_{4,1} y_4^8 + A_{6,1} y_6^6), \quad (1.194)$$

where

$$A_{j,1} := \frac{A_j}{g_1(z_0)} \quad \text{for } j \in \{3, 4, 6\}. \quad (1.195)$$

Case 2 (“large z , small n ”): $z \geq z_0$ and (1.153) fails to hold.

Recall the definition (1.72) of T_1 and also that $c \in (0, 1)$, and then note that

$$\mathbb{P}(T_1 > z) \leq \mathbb{P}(\sqrt{n}\bar{Y} > \sqrt{c}z) + \mathbb{P}(\bar{Y}^2 < c).$$

By (1.26),

$$\mathbb{P}(\sqrt{n}\bar{Y} > \sqrt{c}z) \leq 1 - \Phi(\sqrt{c}z) + \frac{30.2211}{c^{3/2}} \frac{y_3^3}{z^3\sqrt{n}}.$$

Next, by [120, Theorem 7] with $\xi_i := -Y_i^2$,

$$\begin{aligned} \mathbb{P}(\bar{Y}^2 \leq c) &\leq \exp\left\{-\frac{n}{y_4^4}(1 - c + c \ln c)\right\} \leq \left(\frac{2}{e(1 - c + c \ln c)}\right)^2 \frac{y_4^8}{n^2} \\ &\leq \omega^3\left(\frac{2}{e(1 - c + c \ln c)}\right)^2 \frac{y_4^8}{z^3\sqrt{n}}, \end{aligned}$$

where $\sup_{x>0} x^2 e^{-x} = (2/e)^2$ is used for the penultimate inequality above, and the restriction on z (1.82) is used for the last inequality. Thus, since $1 - \Phi(z) < 1 - \Phi(z\sqrt{c})$ and $1/z^3 \leq g_2(z)$,

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq h(z) + \frac{\tilde{A}_{3,2}y_3^3}{z^3\sqrt{n}} + \frac{\tilde{A}_{4,2}y_4^8}{z^3\sqrt{n}} \leq h(z) + \frac{g_2(z)}{\sqrt{n}}(\tilde{A}_{3,2}y_3^3 + \tilde{A}_{4,2}y_4^8), \quad (1.196)$$

where

$$h(z) := 1 - \Phi(\sqrt{c}z), \quad \tilde{A}_{3,2} := \frac{30.2211}{c^{3/2}}, \quad \text{and} \quad \tilde{A}_{4,2} := \omega^3\left(\frac{2}{e(1 - c + c \ln c)}\right)^2. \quad (1.197)$$

By the assumed conditions of Case 2, at least one of the inequalities in (1.153) fails to hold. Therefore and in view of (1.192),

$$\begin{aligned} h(z) &\leq h(z)\left(\frac{K_1y_3^3}{\sqrt{n}} \vee \frac{K_2y_4^8}{z^3\sqrt{n}} \vee \frac{K_3v_3^3}{z^3\sqrt{n}}\right) \\ &\leq \frac{g_2(z)}{\sqrt{n}} \max(K_1S_{3,2}y_3^3, K_2S_{4,2}y_4^8, K_3S_{4,2}(\nu_{3,2}y_3^3 + \nu_{4,2}y_4^8 + \nu_{6,2}y_6^6)), \end{aligned} \quad (1.198)$$

where

$$S_{3,2} := \sup_{z \geq z_0} \frac{h(z)}{g_2(z)} \quad \text{and} \quad S_{4,2} := \sup_{z \geq z_0} \frac{h(z)}{z^3g_2(z)}. \quad (1.199)$$

Thus, by (1.196) and (1.198),

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq \frac{g_2(z)}{\sqrt{n}}(A_{3,2}y_3^3 + A_{4,2}y_4^8 + A_{6,2}y_6^6), \quad (1.200)$$

where

$$\begin{aligned} A_{3,2} &:= \tilde{A}_{3,2} + \max(K_1S_{3,2}, K_3S_{4,2}\nu_{3,2}), \\ A_{4,2} &:= \tilde{A}_{4,2} + S_{4,2} \max(K_2, K_3\nu_{4,2}), \\ \text{and } A_{6,2} &:= K_3S_{4,2}\nu_{6,2}. \end{aligned}$$

Case 3 (“large z , large n ”): $z \geq z_0$ and (1.153) is true.

In this final case, the assumptions of Theorem 1.A.2 all hold when M_ϵ is as in (1.184). Recall now the definition (1.193) of $\tilde{\theta}$, the inequality in (1.192), and also note that $\mathfrak{K}_{\mathbf{e}0} \leq \mathfrak{K}_{\mathbf{e}0}(\pi_4 y_3^3 + (1 - \pi_4) y_4^8)$ (which follows because $1 \leq y_3 \leq y_4$). Then (1.154) yields

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq \frac{z^{-3}(\beta_3 y_3^3 + \beta_4 y_4^8 + \beta_6 y_6^6) + e^{-z/\tilde{\theta}}(\beta_{3,e} y_3^3 + \beta_{4,e} y_4^8 + \beta_{6,e} y_6^6)}{\sqrt{n}}, \quad (1.201)$$

where

$$\begin{aligned} \beta_3 &:= \mathfrak{K}_{\mathbf{n}1} + \nu_{3,3}(\mathfrak{K}_{\mathbf{n}31} + \mathfrak{K}_{\mathbf{n}32}), & \beta_{3,e} &:= \pi_4 \mathfrak{K}_{\mathbf{e}0} + \mathfrak{K}_{\mathbf{e}1} + \nu_{3,3} \mathfrak{K}_{\mathbf{e}3}, \\ \beta_4 &:= \mathfrak{K}_{\mathbf{n}21} \vee \mathfrak{K}_{\mathbf{n}22} \vee (\nu_{4,3} \mathfrak{K}_{\mathbf{n}31}) + \nu_{4,3} \mathfrak{K}_{\mathbf{n}32}, & \beta_{4,e} &:= (1 - \pi_4) \mathfrak{K}_{\mathbf{e}0} + \mathfrak{K}_{\mathbf{e}2} + \nu_{4,3} \mathfrak{K}_{\mathbf{e}3}, \\ \beta_6 &:= \nu_{6,3}(\mathfrak{K}_{\mathbf{n}31} + \mathfrak{K}_{\mathbf{n}32}), & \beta_{6,e} &:= \nu_{6,3} \mathfrak{K}_{\mathbf{e}3}. \end{aligned}$$

Next, let

$$e_z := \sup_{z \geq z_0} z^3 e^{-z/\tilde{\theta}}. \quad (1.202)$$

Then, by the definition (1.193) of $g_3(z)$, for any $j \in \{3, 4, 6\}$

$$\frac{\beta_j}{z^3} + \frac{\beta_{j,e}}{e^{z/\tilde{\theta}}} \leq g_3(z) \sup_{z \geq z_0} \frac{\beta_j z^{-3} + \beta_{j,e} e^{-z/\tilde{\theta}}}{z^{-3} + w_{g,3} e^{-z/\tilde{\theta}}} = g_3(z) \sup_{r \in (0, e_z]} \frac{\beta_j + \beta_{j,e} r}{1 + w_{g,3} r} = A_{j,3} g_3(z), \quad (1.203)$$

where

$$A_{j,3} := \beta_j \vee \frac{\beta_j + \beta_{j,e} e_z}{1 + w_{g,3} e_z} \quad \text{for } j \in \{3, 4, 6\}.$$

Now, by (1.201) and (1.203),

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq \frac{g_3(z)}{\sqrt{n}} (A_{3,3} y_3^3 + A_{4,3} y_4^8 + A_{6,3} y_6^6). \quad (1.204)$$

Now combine the inequalities in (1.194), (1.200), and (1.204), and recall also the definitions (1.193) of the functions g_k , to see that

$$|\mathbb{P}(T_1 \leq z) - \Phi(z)| \leq \frac{z^{-3} + w_g e^{-z/\tilde{\theta}}}{\sqrt{n}} (\hat{A}_3 y_3^3 + \hat{A}_4 y_4^8 + \hat{A}_6 y_6^6), \quad (1.205)$$

where

$$\hat{A}_j := \max_{k \in \{1,2,3\}} A_{j,k} \text{ for } j \in \{3, 4, 6\}, \quad \text{and} \quad w_g := \max_{(j,k) \in \{3,4,6\} \times \{1,2,3\}} \frac{A_{j,k}}{\hat{A}_j} w_{g,k}. \quad (1.206)$$

In view of (1.205) and (1.83), the proof will be complete upon demonstrating the existence of a set of parameters such that the constants listed in Table 1.1 are in accordance with the definitions in (1.206).

Similarly to the proof of [111, Theorem 1.1], those constants are obtained by trying to minimize the value of $\hat{A}_3 \vee (\hat{A}_4/w_4) \vee (\hat{A}_6/w_6)$ for each of the points $(\omega, w_g, (w_4, w_6)) \in \{0.1, 0.5\} \times \{0, 1\} \times \{(1, 1), (0.5, 0.2)\}$. Note that treating w_g in (1.206) as an arbitrarily

fixed constant introduces the restriction that $w_{g,k} \leq w_g \min_j \hat{A}_j / A_{j,k}$ for each $k \in \{1, 2, 3\}$, and so $w_{g,k} = 0$ when $w_g = 0$; further, on recalling the definitions (1.81) and (1.193) of g and $\tilde{\theta}$ along with the bound in (1.205), one has the additional restriction that $\tilde{\theta} = 2$, whence $\theta = 2(1 - \pi_1)$.

The parameters used to obtain the constants \hat{A}_j are tabulated in Tables 1.4 and 1.5 below. There are a few remarks to be made concerning the verification that the values listed in those tables indeed prove the statement of Corollary 1.4.12. First, it is a practical necessity to use a sufficiently powerful CAS; we performed the calculations with the Mathematica software. In order to skirt any issue of rounding error in intermediate calculations, the values in Tables 1.4 and 1.5 should be interpreted as being exact rational numbers; in this way, \hat{A}_j (and the expressions upon which the \hat{A}_j 's depend) can be calculated to within any prescribed precision.

Some care must be taken in order to implement the expressions for the \hat{A}_j 's. Note that ν_3 in (1.187) (used in the definition of $\nu_{j,k}$ in (1.192)) and M_ϵ in (1.184) are algebraic expressions and therefore can be calculated exactly in a CAS. Concerning the numbers $A_{j,1}$ in (1.195), the triples (A_3, A_4, A_6) are obtained by similar calculations (with exact rational numbers) as directed by the proof of [111, Theorem 1.2]; one should also replace the absolute constant 0.4785 in the proof there (due to Tyurin [141]) with the smaller constant 0.4748 (due to Shevtsova [136]). For each of the two pairs $(w_4, w_6) \in \{(1, 1), (0.5, 0.2)\}$ considered here, the parameters used to obtain the triples (A_3, A_4, A_6) are listed below (using the notation of [111]):

w_4	w_6	A_3	A_4	A_6	α	ε_2	ε_3	ε_4	θ_3	θ_4	κ
1	1	1.5175	1.4852	1.4814	0.080	0.206	3.187	0.135	0.415	2.898	0.173
0.5	0.2	1.9946	0.9996	0.1897	0.216	0.369	0.761	0.278	0.408	3.532	0.275

Also note that (1.155)–(1.158), (1.199), and (1.202) contain expressions of the general form $\sup_{x \geq x_0} k(x)$ or $\sup_{0 < x \leq x_0} k(x)$ for some function k and positive number x_0 . For the specific values of the parameters listed in Tables 1.4 and 1.5, one can use Lemma 1.C.1 below to see that these suprema are all attained at the boundary point x_0 . Finally, bounding \tilde{a}_1 in (1.151) involves estimating the root of the equation in [108, (2.3)]; as noted at the end of the paragraph containing formula (1.15), $L_{W;c,\sigma}$ is nonincreasing in $\sigma > 0$, and hence any upper bound on the mentioned root results in an upper bound on \tilde{a}_1 . Implementation of the expressions \hat{A}_j in accordance with the above remarks and the parameter values listed in Tables 1.4 and 1.5 will then demonstrate that (1.83) holds.

To prove that the statement of Corollary 1.4.12 holds when Y is assumed to be symmetric and Table 1.2 is used in place of Table 1.1, one need only amend the definitions of \tilde{c}_2 and \mathfrak{R}_{n1} as prescribed by Remark 1.A.3, and then use the parameter values given below in Tables 1.4 and 1.5. \square

Lemma 1.C.1. *Say that a function k is $\nearrow \searrow$ on $(0, \infty)$ whenever there exists a point $x_* \in (0, \infty)$ such that k is increasing on $(0, x_*)$ and decreasing on (x_*, ∞) . Also say that “the supremum of a function k is attained at the finite (or positive) boundary point” if $\sup_{x \geq x_0} k(x) = k(x_0)$ (or $\sup_{0 < x \leq x_0} k(x) = k(x_0)$). Then the following statements are all true:*

Table 1.4: Parameters associated with Corollary 1.4.12, for $w_g = 0$. For all columns below, set $w_{g,1} = w_{g,2} = w_{g,3} = 0$, $\kappa_3 = 1.5$, $\pi_2 = 1 - \pi_1 - \pi_3$, $\varepsilon_* = 0.001$, $\kappa_{2,0} = \kappa_{3,0} = \kappa_{3,1} = 1$, and $\pi_4 = 0$.

ω (w_4, w_6)	0.5		0.1	
	(1,1)	(0.5,0.2)	(1,1)	(0.5,0.2)
\hat{A}_3	166	229	151	170
\hat{A}_4	166	115	148	85
\hat{A}_6	165	45	147	29
ϵ	0.232	0.301	0.054	0.073
z_0	4.782	4.855	4.629	4.390
c	0.757	0.759	0.900	0.891
K_1	6.9×10^4	1.3×10^5	2.0×10^5	9.2×10^4
K_2	6.3×10^6	4.0×10^6	2.2×10^7	4.1×10^6
K_3	6.9×10^6	3.4×10^6	2.3×10^7	1.6×10^6
$w_{0,2}$	0.156	0.380	0.206	0.147
$w_{3,2}$	0.400	0.036	0.600	1.000
$w_{4,2}$	0.380	1.000	0.742	0.778
c_*	0.536	0.621	0.500	0.514
θ	0.861	0.880	0.875	0.978
w	0.360	0.316	0.376	0.398
δ_0	0.007	0.009	0.007	0.010
π_1	0.042	0.083	0.008	0.015
π_3	0.645	0.635	0.660	0.660
κ_2	2.108	2.093	2.102	2.116
$\kappa_{2,1}$	1.570	0.800	6.050	1.612
α	0.070	0.050	0.075	0.080
$w_{0,3}$	0.278	0.275	0.216	0.392
$w_{3,3}$	0	0.365	0	0
$w_{4,3}$	0.595	0.980	0.45	1

- (i) For any $c \in (0, 1)$, the function h as defined in (1.197) is decreasing on $(0, \infty)$.
- (ii) For any $p > 0$ and $\kappa > 0$, the function $x \mapsto x^p e^{-\kappa x}$ is $\nearrow \searrow$ on $(0, \infty)$.
- (iii) For any $0 < p \leq \kappa$ and $\varepsilon \in (0, 1)$, the function $x \mapsto x^p \text{PU}_{\text{tail}}(x, \kappa, \varepsilon)$ is $\nearrow \searrow$ on $(0, \infty)$.
- (iv) For any $\kappa \geq 2$ and $\alpha \in (0, 1)$, the function $x \mapsto \frac{1}{x} \exp\{-\kappa(1-\alpha/2+x) \ln(1+(1-\alpha)/x)\}$ is $\nearrow \searrow$ on $(0, \infty)$.
- (v) For any $c \in (0, 1)$, the function $x \mapsto x^3 h(x)$ is $\nearrow \searrow$ on $(0, \infty)$, where h is as in (1.197).

The suprema in the expressions (1.155)–(1.158), (1.199), and (1.202) are all attained at the respective finite (or positive) boundary points whenever the values in Tables 1.4 and 1.5 are substituted in those expressions.

Proof of Lemma 1.C.1. Statements (i) and (ii) are trivial to verify by differentiation.

Table 1.5: Parameters associated with Corollary 1.4.12, for $w_g = 1$. For all columns below, set $w_{g,1} = w_{g,3} = 1$, $w_{g,2} = 0$, $\kappa_3 = 1.5$, $\pi_2 = 1 - \pi_1 - \pi_3$, $\theta = 2(1 - \pi_1)$, $\alpha = 0.05$, $\varepsilon_* = 0.001$, and $\kappa_{3,0} = \kappa_{3,1} = 1$.

ω (w_4, w_6)	0.5		0.1	
	(1,1)	(0.5,0.2)	(1,1)	(0.5,0.2)
\hat{A}_3	48	66	38	39
\hat{A}_4	48	33	36	20
\hat{A}_6	42	13	36	7
ϵ	0.363	0.438	0.066	0.112
z_0	6.800	7.175	6.550	6.074
c	0.738	0.708	0.885	0.874
K_1	4.0×10^5	5.0×10^7	7.9×10^6	1.2×10^6
K_2	9.0×10^7	3.5×10^9	6.6×10^{10}	1.6×10^9
K_3	1.0×10^8	5.5×10^9	1.3×10^{10}	3.5×10^8
$w_{0,2}$	0.040	0.133	0.263	0.142
$w_{3,2}$	0	1	0.100	0.590
$w_{4,2}$	0.080	0.600	0.588	0.396
c_*	0.490	0.741	0.500	0.552
w	1.160	0.940	1.655	1.530
δ_0	0.039	0.027	0.016	0.018
π_1	0.108	0.257	0.012	0.038
π_3	0.422	0.409	0.415	0.423
κ_2	2.095	2.012	2.011	2.017
$\kappa_{2,0}$	1	0.799	1	1.046
$\kappa_{2,1}$	0.983	1.496	4.750	1.104
π_4	0	0.467	0	0.220
$w_{0,3}$	0.392	0.318	0.392	0.392
$w_{3,3}$	0	0.224	0	0
$w_{4,3}$	1	1	1	1

By (1.140), to prove statement (iii), it suffices to show that

$$x \mapsto p \ln x + \frac{\kappa}{2(1-\varepsilon)x} \left((1-\varepsilon)^2 \left(1 + W\left(\frac{\varepsilon}{1-\varepsilon} \exp\left\{\frac{\varepsilon+x}{1-\varepsilon}\right\}\right) \right)^2 - (\varepsilon+x)^2 - (1-\varepsilon^2) \right)$$

is $\nearrow \searrow$ on $(0, \infty)$. Now let $w := \frac{1-\varepsilon}{\varepsilon} W\left(\frac{\varepsilon}{1-\varepsilon} \exp\left\{\frac{\varepsilon+x}{1-\varepsilon}\right\}\right)$, whence $x = (1-\varepsilon)\left(\frac{\varepsilon}{1-\varepsilon} w + \ln w\right) - \varepsilon$, and note that w continuously increases from 1 to ∞ as x increases from 0 to ∞ . Thus, it suffices to show that

$$\begin{aligned} k(w) &:= p \ln \left((1-\varepsilon) \left(\frac{w\varepsilon}{1-\varepsilon} + \ln w \right) - \varepsilon \right) \\ &\quad + \frac{\kappa \left((1-\varepsilon)^2 \left(1 + \frac{w\varepsilon}{1-\varepsilon} \right)^2 - (1-\varepsilon)^2 \left(\frac{w\varepsilon}{1-\varepsilon} + \ln w \right)^2 - (1-\varepsilon^2) \right)}{2(1-\varepsilon) \left((1-\varepsilon) \left(\frac{w\varepsilon}{1-\varepsilon} + \ln w \right) - \varepsilon \right)} \end{aligned}$$

Table 1.6: Parameters associated with Remark 1.4.16, for $w_g = 0$. For all columns below, set $w_{g,1} = w_{g,2} = w_{g,3} = 0$, $\kappa_3 = 1.5$, $\pi_2 = 1 - \pi_1 - \pi_3$, $\varepsilon_* = 10^{-4}$, and $\kappa_{3,0} = \kappa_{3,1} = 1$.

ω (w_4, w_6)	0.5		0.1	
	(1,1)	(0.5,0.2)	(1,1)	(0.5,0.2)
\hat{A}_3	141	206	124	145
\hat{A}_4	138	103	124	73
\hat{A}_6	138	42	121	22
ϵ	0.264	0.310	0.072	0.082
z_0	4.527	4.679	4.328	4.170
c	0.750	0.762	0.900	0.918
K_1	2.3×10^4	3.0×10^4	5.4×10^4	3.4×10^4
K_2	1.8×10^6	1.4×10^6	5.1×10^6	8.0×10^5
K_3	2.0×10^6	1.5×10^6	2.3×10^6	6.0×10^5
$w_{0,2}$	0.274	0.173	0.144	0.153
$w_{3,2}$	0.214	0.852	0.100	1
$w_{4,2}$	0.688	0.916	0.300	1
c_*	0.565	0.643	0.510	0.581
θ	0.849	0.894	0.890	1.060
w	0.320	0.381	0.430	0.344
δ_0	0.010	0.048	0.009	0.038
π_1	0.054	0.090	0.009	0.019
π_3	0.655	0.601	0.664	0.655
κ_2	2.137	2.119	2.143	2.159
$\kappa_{2,0}$	1	1.127	0.848	1
$\kappa_{2,1}$	1.310	0.868	3.819	1.142
α	0.200	0.150	0.120	0.150
$w_{0,3}$	0.276	0.220	0.280	0.392
$w_{3,3}$	0	0.595	0	0
$w_{4,3}$	0.590	1	0.600	1

is $\nearrow \searrow$ on $(1, \infty)$. Next, introduce

$$\begin{aligned}
k_1(w) &:= \frac{2w((w-1)\varepsilon + (1-\varepsilon)\ln w)^2}{1-\varepsilon+\varepsilon w} k'(w) \\
&= 2(p + \varepsilon(\kappa - p)) \ln w - 2(w-1)(\kappa - p)\varepsilon - (1-\varepsilon)\kappa \ln^2 w,
\end{aligned}$$

and note that k_1 and k' have the same sign on $(1, \infty)$. Also introduce

$$k_2(w) := \frac{w}{2} k'_1(w) = p - \varepsilon(w-1)(\kappa - p) - (1-\varepsilon)\kappa \ln w.$$

Then k_2 and k'_1 share the same sign on $(1, \infty)$ and k_2 is decreasing on $(1, \infty)$. Further, since $k_2(1) = p > 0$ and $k_2(\infty) = -\infty$, we see that k_2 and hence k'_1 change sign once from $+$ to $-$ on $(1, \infty)$; that is, k_1 is $\nearrow \searrow$ on $(1, \infty)$. As $k_1(1) = 0$ and $k_1(\infty) = -\infty$, it follows that k_1 and hence k' change sign once from $+$ to $-$ on $(1, \infty)$. That is, k is $\nearrow \searrow$ on $(1, \infty)$, and

Table 1.7: Parameters associated with Remark 1.4.16, for $w_g = 1$. For all columns below, set $w_{g,1} = w_{g,3} = 1$, $w_{g,2} = 0$, $\kappa_3 = 1.5$, $\pi_2 = 1 - \pi_1 - \pi_3$, $\theta = 2(1 - \pi_1)$, $\varepsilon_* = 10^{-4}$, and $\kappa_{3,0} = \kappa_{3,1} = 1$.

ω (w_4, w_6)	0.5		0.1	
	(1,1)	(0.5,0.2)	(1,1)	(0.5,0.2)
\hat{A}_3	48	57	35	37
\hat{A}_4	48	29	32	19
\hat{A}_6	41	12	31	5
ϵ	0.365	0.456	0.153	0.131
z_0	6.800	6.885	6.200	6.015
c	0.738	0.677	0.910	0.894
K_1	1.0×10^5	8.2×10^4	4.0×10^5	4.0×10^5
K_2	3.0×10^6	1.0×10^9	5.0×10^7	2.0×10^8
K_3	1.0×10^7	5.0×10^8	9.0×10^7	1.0×10^8
$w_{0,2}$	0	0.392	0.224	0.018
$w_{3,2}$	0.030	0	0.481	0.514
$w_{4,2}$	0	1	0.704	0.041
c_*	0.760	0.703	0.470	0.625
w	0.692	0.913	1.612	1.163
δ_0	0.124	0.078	0.055	0.282
π_1	0.144	0.291	0.023	0.052
π_3	0.453	0.393	0.432	0.461
κ_2	2.082	2.015	2.053	2.024
$\kappa_{2,0}$	1.588	1.101	1.476	1.313
$\kappa_{2,1}$	0.838	0.796	2.474	3.073
π_4	0.487	0.950	0	0.368
α	0.067	0.150	0.103	0.137
$w_{0,3}$	0.363	0.251	0.239	0.383
$w_{3,3}$	0	0.461	0	0.026
$w_{4,3}$	0.856	1	0.500	1

thus statement (iii) is proved.

To prove (iv), let

$$\begin{aligned}
k(x) &:= -\kappa \left(1 - \frac{\alpha}{2} + x\right) \ln \left(1 + \frac{1-\alpha}{x}\right) - \ln x, \\
k_1(x) &:= k'(x) = \frac{\kappa(1-\alpha)(2+2x-\alpha)}{2x(1-\alpha+x)} - \frac{1}{x} - \kappa \ln \left(1 + \frac{1-\alpha}{x}\right), \\
k_2(x) &:= 2x^2(1-\alpha+x)^2 k_1'(x) = 2x^2 - 2x(1-\alpha)(\kappa-2) - (1-\alpha)^2(\kappa(2-\alpha)-2).
\end{aligned}$$

Then k_2 is decreasing on $(0, x_*)$ and increasing on (x_*, ∞) , where $x_* := \frac{1}{2}(\kappa-2)(1-\alpha)$. Since $k_2(0) = -(1-\alpha)^2(\kappa(2-\alpha)-2) < 0$ and $k_2(\infty) = \infty$, it follows that k_2 and hence k_1' change sign once from $-$ to $+$ on $(0, \infty)$. So, k_1 is $\searrow \nearrow$ on $(0, \infty)$; as $k_1(0+) = \infty$ and $k_1(\infty) = 0$, we see that k_1 changes sign once from $+$ to $-$, and hence k is $\nearrow \searrow$ on $(0, \infty)+$.

Thus, $x \mapsto \exp\{k(x)\}$ is $\nearrow \searrow$ on $(0, \infty)$, proving statement (iv).

The proof of part (v) is easily done by using the l'Hospital-type rule for monotonicity, as in the proof of Lemma 3 in [104].

To finish the proof, make the various substitutions from Tables 1.4 and 1.5 into the respective expressions of (1.155)–(1.158), (1.199), and (1.202); note that, since $w_{g,2} = 0$ in all of the parameter sets, $g_2(z) = z^{-3}$ and hence $S_{4,2} = h(z_0)$ follows from statement (i) and (1.199). Next estimate the unique positive critical point x_* of each of the functions in statements (ii)–(v) by finding rational numbers $x_1 < x_2 < x_3$ such that $k(x_1) < k(x_2)$ and $k(x_2) > k(x_3)$; then we shall know that $x_* \in (x_1, x_3)$. So, it will follow that $\sup_{x \geq x_0} k(x)$ is attained at the boundary point x_0 by checking that $x_0 \geq x_3$, and that $\sup_{0 < x \leq x_0} k(x)$ is attained at x_0 by checking that $x_0 < x_1$. Thus, one completes the proof. \square

Proof of Corollary 1.4.20. For $\alpha \geq 1$, let

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

Also adopt the notation of Theorem 1.4.17, with $\rho = 0$, so that

$$V = (Y, Z, Y^2 - 1, Z^2 - 1, YZ), \quad L(V) = YZ, \quad \text{and} \quad \tilde{\sigma} = \|YZ\|_2.$$

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

$$|P(\sqrt{n}R/\tilde{\sigma} \leq z) - \Phi(z)| \leq 1 \leq \frac{\sqrt{N_0 - 1}}{\sqrt{n}} \leq \frac{y_6^6 + z_6^6}{\sqrt{n}} \left(B_{0,1} + \frac{B_{3,1}}{\tilde{\sigma}^3} \right), \quad (1.207)$$

where

$$(B_{0,1}, B_{3,1}) := \frac{\sqrt{N_0 - 1}}{2(1 + b_3)} (1, b_3). \quad (1.208)$$

Suppose then that $n \geq N_0$. Take any $\epsilon \in (0, \sqrt{3}/2)$ and $c_* \in (0, 1)$, so that the conditions of Theorem 1.4.1 are satisfied (cf. the discussion following (1.88)); also introduce the parameters $\kappa > 0$ and $\pi \in [0, 1]$, so that $\mathfrak{R}_{u0} \leq \pi \mathfrak{R}_{u0} + (1 - \pi) \mathfrak{R}_{u0} \varsigma_3^3$. Recall the notation in (1.51), 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 (1.53) implies

$$\left| \mathbb{P}\left(\frac{R}{\tilde{\sigma}/\sqrt{n}} \leq z\right) - \Phi(z) \right| \leq \frac{y_6^6 + z_6^6}{\sqrt{n}} \left(A_0 + \frac{A_1}{\tilde{\sigma}} + \frac{A_2}{\tilde{\sigma}^2} + \frac{A_3}{\tilde{\sigma}^3} \right) \leq \frac{y_6^6 + z_6^6}{\sqrt{n}} \left(B_{0,2} + \frac{B_{3,2}}{\tilde{\sigma}^3} \right), \quad (1.209)$$

where

$$\begin{aligned} A_0 &:= \frac{\pi}{2} \mathfrak{K}_{u0} + \frac{3}{2\epsilon^2\sqrt{N_0}} \wedge \frac{3^{3/2}(3+1/\sqrt{N_0})}{2\epsilon^3 N_0}, & A_1 &:= \frac{3}{2}(\mathfrak{K}_{u20} + \mathfrak{K}_{u30})\tilde{\sigma}, \\ A_2 &:= \frac{3}{2}(\mathfrak{K}_{u21} + \mathfrak{K}_{u31})\tilde{\sigma}, & A_3 &:= \frac{1-\pi}{2} \mathfrak{K}_{u0} + \frac{1}{2} \mathfrak{K}_{u1}, \end{aligned} \quad (1.210)$$

with N_0 replacing n in the expressions $\mathfrak{K}_{u1}, \dots, \mathfrak{K}_{u3,1}$,

$$B_{0,2} := A_0 + \frac{2}{3} \kappa^{-3/2} A_1 + \frac{1}{3} \kappa^{-3} A_2, \quad \text{and} \quad B_{3,2} := A_3 + \frac{1}{3} \kappa^3 A_1 + \frac{2}{3} \kappa^{3/2} A_2. \quad (1.211)$$

Then (1.207) and (1.209) yield the desired inequality (1.91) if we let

$$B_0 := B_{0,1} \vee B_{0,2} \quad \text{and} \quad B_3 := B_{3,1} \vee B_{3,2}. \quad (1.212)$$

We shall show that, for f as in (1.88), (1.33) holds for any pair

$$(\epsilon, M_\epsilon) \in \{(0.06, 1.094), (0.17, 1.365), (0.25, 1.688), (0.30, 1.962)\}. \quad (1.213)$$

Then, substituting the values of the parameters b_3 , N_0 , ϵ , M_ϵ , c_* , and κ given in Table 1.8 below into the expressions for B_0 and B_3 in (1.212) (which depend on the expressions in (1.208), (1.211), (1.210), and (1.55)), one will see that (1.91) holds for any of the pairs (B_0, B_3) listed in (1.92).

Table 1.8: Parameters associated with pairs (B_0, B_3) in (1.92) and (1.93)

	(1.92)					(1.93)				
B_0	0.71	1.38	4.08	14.32	37.82	0.68	1.31	3.74	13.19	36.40
B_3	19.16	11.02	4.08	1.79	1.41	18.27	10.47	3.74	1.65	1.35
b_3	27	8	1	1/8	1/27	27	8	1	1/8	1/27
N_0	1580	614	267	1038	6154	1436	555	224	881	5699
ϵ	0.3	0.3	0.25	0.17	0.06	0.3	0.3	0.25	0.17	0.06
c_*	0.922	0.893	0.794	0.647	0.5724	0.922	0.892	0.792	0.627	0.559
κ	2.324	1.745	0.978	0.460	0.2954	2.374	1.807	1.0148	0.452	0.285
π	0	0	0	1	1	0	0	0	1	1

To prove Remark 1.4.22, one only needs to redefine the pre-constants $\mathfrak{K}_{u20}, \dots, \mathfrak{K}_{u31}$ as directed by Remark 1.4.2 and use the parameter values given in Table 1.8.

To complete the proof of Corollary 1.4.20 (and Remark 1.4.22), it now remains to verify (1.213). Toward that end, take any $\epsilon \in (0, \sqrt{3}/2)$, and recall the definition (1.88) of f (with

$\rho = 0$) to see that

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

and

$$f(x_1, x_2, x_3, x_4, x_5) \equiv f(x_2, x_1, x_4, x_3, x_5)$$

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

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

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

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

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

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

$$\begin{aligned} M_\epsilon^* &\leq \max_{\mathbf{u} \in U} \sup_{\mathbf{x} \in C_{\mathbf{u}}} \|f''(\mathbf{x})\| \leq \max_{\mathbf{u} \in U} \left(\|f''(\mathbf{c}_{\mathbf{u}})\| + \sup_{\mathbf{x} \in C_{\mathbf{u}}} \|f''(\mathbf{x}) - f''(\mathbf{c}_{\mathbf{u}})\|_F \right) \\ &\leq \max_{\mathbf{u} \in U} \left(\|f''(\mathbf{c}_{\mathbf{u}})\| + \sqrt{5} \frac{\delta_\epsilon}{2} \sup_{\mathbf{x} \in C_{\mathbf{u}}} \|f'''(\mathbf{x})\|_F \right), \end{aligned} \tag{1.214}$$

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. $\min_{\mathbf{u} \in U} \inf_{\mathbf{x} \in C_{\mathbf{u}}} (1 + x_3 - x_1^2)(1 + x_4 - x_2^2) > 0$).

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 CAS that

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

$$\text{where } \tilde{\mathbf{x}} := (\tilde{x}_1, \dots, \tilde{x}_5) := \left(x_1, x_2, \frac{1}{1+x_3-x_1^2}, \frac{1}{1+x_4-x_2^2}, x_5\right), \quad (1.215)$$

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

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

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 $s := \text{Sgn}(a) \text{Sgn}(u_1)^{d_1} \text{Sgn}(u_2)^{d_2} \text{Sgn}(u_5)^{d_5}$,

$$a\tilde{x}_1^{d_1}\tilde{x}_2^{d_2}\tilde{x}_3^{d_3}\tilde{x}_4^{d_4}\tilde{x}_5^{d_5} = s|a||\tilde{x}_1|^{d_1} \dots |\tilde{x}_5|^{d_5} \leq s|a||\tilde{x}_{1,s}|^{d_1} \dots |\tilde{x}_{5,s}|^{d_5} = a\tilde{x}_{1,s}^{d_1}\tilde{x}_{2,s}^{d_2}\tilde{x}_{3,s}^{d_3}\tilde{x}_{4,s}^{d_4}\tilde{x}_{5,s}^{d_5}, \quad (1.216)$$

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 (1.216), we see from (1.215) that

$$\|f'''(\mathbf{x})\|_F \leq \frac{\sqrt{3\tilde{x}_{3,1}\tilde{x}_{4,1}}}{8} \sqrt{p_{\text{Sgn}(u_1)}(\tilde{x}_{1,1}, \dots, \tilde{x}_{5,1}, \tilde{x}_{1,-1}, \dots, \tilde{x}_{5,-1})}, \quad (1.217)$$

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 (1.214) and (1.217), 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 $\epsilon \in \{\frac{6}{100}, \frac{17}{100}, \frac{25}{100}, \frac{30}{100}\}$, (1.213) follows. \square

1.D On Fisher's z transform

A statistic closely related to Pearson's R is commonly known as the Fisher z transform, defined by the formula $R_z := \tanh^{-1}(R) = \frac{1}{2} \ln\left(\frac{1+R}{1-R}\right)$. An advantage to using R_z (as opposed to R) in making statistical inferences about ρ follows from its variance-stabilizing property in normal populations; that is, $n \operatorname{Var}(R_z) \rightarrow 1$ for all $\rho \in (-1, 1)$ as $n \rightarrow \infty$, as opposed to $n \operatorname{Var}(R) \rightarrow (1 - \rho^2)^2$, whenever (Y, Z) has a bivariate normal distribution. Moreover, the distribution of R_z converges to normality more rapidly than does the distribution of R (especially for non-zero values of ρ) when the pair (Y, Z) comes from a normal population; see e.g. Fisher [34], David [26], and Hotelling [55]. In his discussion of Hotelling's paper, Kendall provides heuristics suggesting that such variance stabilization of the distribution of a statistic may often result in it being closer to normality. Namely, if an approximate constancy of the variance of a statistic were the same as an approximate constancy of its distribution itself, and if the distribution is close to normality at least for one value of the parameter (say, ρ , as in the present case), then it would be close to normality for all values of ρ .

However, it is well known that the closeness of the distribution of a statistic to normality is usually mainly determined, not by the variance, but by the third moments of the underlying distribution. It is therefore natural to wonder whether or to what extent the nice properties of the z transform hold for non-normal populations. For moderate sample sizes n , Gayen [38] observed that the convergence to normality for both R and R_z is lessened for non-normal populations with $\rho \neq 0$, and Monte Carlo sampling performed by Berry and Mielke [12] suggests that the presence of skewness or heavy tails in the population of (Y, Z) significantly reduces the accuracy of a normal approximation to R_z when $\rho \neq 0$. In [98], explicit expressions for $\Delta_R = \lim_{n \rightarrow \infty} \sqrt{n} |F_R - \Phi|_K$ and $\Delta_{R_z} = \lim_{n \rightarrow \infty} \sqrt{n} |F_{R_z} - \Phi|_K$ are derived, where F_R and F_{R_z} are the d.f.'s of R and R_z and $|\cdot|_K$ denotes the Kolmogorov distance. These "asymptotic distances" generally depend on up to the sixth moments of Y and Z when $\rho \neq 0$, and it is demonstrated in [98] that, if the distribution of (Y, Z) is not bivariate normal, Δ_{R_z} can be just as easily greater than Δ_R as less.

In light of the above considerations, we now briefly investigate how any of the BE-type bounds of Section 1.3, when applied to the statistic R_z , would fare in a comparison with corresponding bounds associated with R . Aside from the choice of parameter values, the only differences between the applications of our bounds to R and R_z are those arising from the choice of f ; namely, upon letting $g(\mathbf{x}) := \tanh^{-1}(f(\mathbf{x}) + \rho) - \tanh^{-1} \rho$ for all \mathbf{x} with f as in (1.88), one has $g(\bar{V}) = R_z - \tanh^{-1} \rho$. In the case when $\rho = 0$, we see that $f'(0) = g'(0)$ and $g''(0) = f''(0)$; moreover, in view of results in [14], one can see that an asymptotic expansion up to $\mathcal{O}(1/\sqrt{n})$ of the d.f. of R is identical to that of R_z , whether or not the population of (Y, Z) is Gaussian.

Despite these similarities between R and R_z , it appears that $M_g := \sup_{\|\mathbf{x}\| \leq \epsilon} \|g''(\mathbf{x})\| > M_f := \sup_{\|\mathbf{x}\| \leq \epsilon} \|f''(\mathbf{x})\|$ for $\epsilon > 0$, at least when $\rho = 0$. In particular, we showed (in the proof of Corollary 1.4.20) that $M_f \leq 1.962$ when $\epsilon = \frac{3}{10}$; on the other hand, one can see that

$$\|g''(\mathbf{x})\| > 2.104 \text{ and } \|\mathbf{x}\| < \frac{3}{10} \text{ when } \mathbf{x} = -\left(\frac{28269}{200000}, \frac{28269}{200000}, \frac{45081}{500000}, \frac{45081}{500000}, \frac{183801}{1000000}\right),$$

so that $M_g > 2.104 > 1.962 \geq M_f$, which will result, at least using the method presented in this paper, in a worse BE-type bound for R_z as compared with that for R . In view of these points, one can conclude that, at least for $\rho = 0$, the use of Fisher's z transform R_z in place of Pearson's R will hardly yield better BE-bounds.

Chapter 2

A Berry-Esseen type bound for the null distribution of the F -statistic from fixed effects general linear models

2.1 Introduction

We were led to the subject of the present paper by studying rates of convergence to normality for a general class of nonlinear statistics. In [116], Berry-Esseen (BE) bounds are developed for statistics of the form $f(S)$, where f is a smooth nonlinear real-valued function and S is a sum of independent zero-mean random vectors (r.v.'s). Under the assumption that $f(0) = 0$ (which can always be assumed by replacing f with $f - f(0)$ if needed), the approximation $f(S) \approx L(S) + c|S|^2$ holds near the origin, with the constant c of course depending on f and the size of the neighborhood about the origin. By adapting methods employed by Chen and Shao [19], we were able to bound $|\mathbf{P}(f(S)/\sigma > z) - \mathbf{P}(L(S)/\sigma > z)|$, where $\sigma = \sqrt{\mathbf{E} L(S)^2}$, by terms of the appropriate order; namely, if the summands of S are also assumed to be identically distributed and have a finite p^{th} moment for some $p \in (2, 3]$, a bound on the order $\mathcal{O}(1/n^{(p-2)/2})$ is obtained in [116]. Then, since $L(S)$ is a sum of independent random variables (r.v.'s), the distance $|\mathbf{P}(L(S)/\sigma > z) - \Phi(z)|$ (where Φ is the standard normal distribution function (d.f.)) is bounded on a similar order by utilizing the classical BE bound for linear statistics; cf. [97, Chapter V] and the references there for several instances of a BE bound for linear statistics.

One needn't search far for specific statistics, of the aforementioned form $f(S)$, which are commonly used in practical applications. Indeed, in [116], BE bounds on the distance to normality of the distributions of the (central or non-central) Student statistic, the Pearson correlation statistic, and the non-central Hotelling statistic are shown to exist; bounds with a simple dependence on n , certain moments of relevant r.v.'s, and explicit absolute constants are also provided for the central Student and Pearson statistics. We note that BE bounds for the central Student statistic, and the closely related self-normalized sum, have been studied by many authors previous to our work in [116]; of particular note are the results by Slavova [139], Bentkus and Götze [10], Novak [93, 94], Shao [134], and Pinelis [111].

Inspired by the relative ease with which BE bounds for such a large class of statistics could be obtained, we turned our attention to another class of statistics: the F -statistic used

The material contained in this chapter is in preparation for submission to the pre-print server <http://arxiv.org>.

to test hypotheses in a general linear model (GLM). We adopt the coordinate-free approach to GLM's. In particular, let \mathcal{H} be an n -dimensional real Hilbert space, referred to as the *sample space*, and let \mathcal{V} be a nonzero proper linear subspace of \mathcal{H} ; \mathcal{V} is called the *model space*. The model assumptions are that

$$Y = \mu + \sigma\varepsilon \quad \text{for some (unknown) } \mu \in \mathcal{V} \text{ and } \sigma > 0, \quad (2.1)$$

where ε is an \mathcal{H} -valued r.v. satisfying the moment conditions

$$\mathbb{E} \varepsilon = 0 \quad \text{and} \quad \text{Cov } \varepsilon = 1_{\mathcal{H}}; \quad (2.2)$$

here, $1_{\mathcal{H}}$ denotes the identity operator on \mathcal{H} . We refer to ε as the *error vector* and Y as the *response vector*.

We believe the coordinate-free approach to GLM's is advantageous for several reasons. In the first place, several seemingly distinct estimation and testing problems are unified under a common notation; particularly, simple linear regression, multiple linear regression, ANOVA, ANCOVA, MANOVA, and MANCOVA models are all easily represented by (2.1) after a suitable choice of the model space \mathcal{V} is made. Moreover, the use of a common notation for all of these different models has the added advantage of simplifying the notation. The matrix and column vector notation traditionally taught to students (and commonly found in the research literature) is often cumbersome or even opaque; the statements in (2.1) and (2.2) strip away any potentially confusing choice of notation and lay bare the fundamentals of the GLM. In conjunction with the generality and simplicity of the notation, the coordinate-free approach to GLM's proffers a more explicit geometric (and thus likely more intuitive) perspective to estimation and testing problems. For instance, Kruskal [76, Theorem 1] characterizes the situation when the Gauss-Markov and least-squares estimators of μ are identical (under the even more general condition that $\text{Cov } \varepsilon$ is positive definite) in just a single paragraph, using only rudimentary facts concerning operators on Hilbert spaces. The reader is referred to [48] for a discussion on the development of the coordinate-free approach to GLM's; a more detailed treatment is provided in the recent text by Wichura [144], and Eaton [31] takes this geometric perspective in his text on multivariate statistics.

Hypothesis testing is also greatly simplified by the coordinate-free GLM. Particularly, most of the testing problems can be stated as testing the null hypothesis

$$H_0: \mu \in \mathcal{V}_0 \quad \text{against the alternative} \quad H_1: \mu \in \mathcal{V} \setminus \mathcal{V}_0, \quad (2.3)$$

where \mathcal{V}_0 is some proper linear subspace of \mathcal{V} . If, in addition to (2.2), we make the distributional assumption that the error ε is normally distributed, the likelihood-ratio test rejects H_0 for large values of the test statistic F (defined later in (2.5)); this statistic has the central \mathcal{F} distribution under the null hypothesis, and otherwise has the noncentral \mathcal{F} distribution whenever (2.1) is assumed.

The assumption that the error vector ε is normally distributed is convenient for theoretical purposes, though it should always be suspect to the practicing statistician. It thus becomes necessary to assess the robustness of the F statistic to violations of normality. Non-normality of a r.v. is generally measured in terms of its skewness and kurtosis, defined to be the third and fourth moments, respectively, after a normalizing (i.e. centering and

scaling to have a mean of 0 and a variance of 1) transformation is applied. The general consensus appears to be that a skewness far from 0 or a kurtosis far from 3 (the kurtosis of the standard normal distribution) can negatively affect inferences made from the F -test by increasing either Type I or II errors. However, these effects can be ameliorated to the point where F may be claimed to be robust to non-normality by having a large number of observations and/or appropriately designing the experiment. In a linear regression model, this amounts to choosing the regressors to be not overly non-normal (cf. [16, 1], and in an ANOVA model robustness can be achieved by choosing equal sample sizes for the various treatments (cf. [130, 3]).

With this robustness of the F statistic in mind, it seems (and indeed is) reasonable to assume that a BE-type bound exists for the null distribution of F under some fairly mild conditions on ε and the model space \mathcal{V} . In Section 2.2, we present such a bound in Theorem 2.2.1 and discuss relevant results in the literature. In Section 2.3, the mentioned BE-type bound is restated in Corollary 2.3.1 under the additional (and common) model assumption that the coordinates of the error ε are independent and identically distributed (i.i.d.); this corollary is then used to demonstrate that the BE-type bound is on the order $\mathcal{O}(1/\sqrt{n})$ for several types of GLM's. The proof of Theorem 2.2.1 is deferred to Section 2.4.

2.1.1 Notation and basic results

Before progressing to the main result, let us set down notation and conventions that will be used throughout the remainder of the paper.

For any natural number k , each instance of an expression of the form \mathfrak{K}_k shall be used to denote a finite positive real number whose value depends solely on the value of k ; that is, (\mathfrak{K}_k) is viewed as a sequence of positive real numbers. Different instances of \mathfrak{K}_k will generally represent different numbers.

Let \mathcal{K} be any finite-dimensional real Hilbert space with scalar product $\langle \cdot, \cdot \rangle$ and norm $|\cdot|$. For any subsets \mathcal{A} and \mathcal{B} of \mathcal{K} , let $\mathcal{A}^\perp := \{y \in \mathcal{K} : \langle x, y \rangle = 0 \text{ for all } x \in \mathcal{A}\}$ denote the orthogonal complement of \mathcal{A} , and let $\mathcal{A} + \mathcal{B} := \{x + y : x \in \mathcal{A}, y \in \mathcal{B}\}$ denote the sum of \mathcal{A} and \mathcal{B} . Further, write $\mathcal{A} \perp \mathcal{B}$ whenever $\langle x, y \rangle = 0$ for all $x \in \mathcal{A}$ and $y \in \mathcal{B}$. We allow for a slight abuse of the above notation when \mathcal{A} or \mathcal{B} is a singleton, in which case “ $\{x\}$ ” is replaced with “ x ”.

The identity operator on \mathcal{K} is denoted by $1_{\mathcal{K}}$. The adjoint of a bounded linear operator A on \mathcal{K} is denoted by A^\top , so that $\langle Ax, y \rangle = \langle x, A^\top y \rangle$ for all x and y in \mathcal{K} .

Let \mathcal{U} be any subspace of \mathcal{K} , and suppose that Π is a linear operator on \mathcal{K} with range \mathcal{U} . The following are then equivalent: (i) Π is idempotent ($\Pi = \Pi^2$) and self-adjoint ($\Pi = \Pi^\top$); (ii) Π is idempotent and $\mathcal{U} \perp \ker(\Pi)$; (iii) $\mathcal{K} = \mathcal{U} \oplus \ker(\Pi)$ and $\mathcal{U} \perp \ker(\Pi)$; and (iv) $|x - \Pi x| = \inf_{y \in \mathcal{U}} |x - y|$. Whenever any of the above conditions are satisfied, Π is called an orthoprojector; the notation $\Pi_{\mathcal{U}}$ will be used to denote the unique orthoprojector whose range is \mathcal{U} . Some other well-known facts that will be used without further mention include:

for any linear subspaces \mathcal{U} and \mathcal{V} of \mathcal{K} ,

$$\begin{aligned}\Pi_{\mathcal{U}} + \Pi_{\mathcal{V}} &= \Pi_{\mathcal{U}+\mathcal{V}} \iff \mathcal{U} \perp \mathcal{V}, \\ \Pi_{\mathcal{U}} - \Pi_{\mathcal{V}} &= \Pi_{\mathcal{V}^\perp \cap \mathcal{U}} \iff \mathcal{V} \subseteq \mathcal{U}, \\ \Pi_{\mathcal{U}}\Pi_{\mathcal{V}} &= \Pi_{\mathcal{U} \cap \mathcal{V}} \iff \Pi_{\mathcal{U}}\Pi_{\mathcal{V}} = \Pi_{\mathcal{V}}\Pi_{\mathcal{U}}, \\ \text{and } \Pi_{\mathcal{U}+\mathcal{V}} &= \Pi_{\mathcal{U}} + \Pi_{\mathcal{V}} - \Pi_{\mathcal{U} \cap \mathcal{V}};\end{aligned}$$

see e.g. Halmos [46] for proofs of the above statements.

For a zero-mean \mathcal{K} -valued r.v. X , the covariance operator $R_X = \text{Cov } X$ is the unique linear operator that satisfies $\mathbb{E}\langle X, x \rangle \langle X, y \rangle = \langle R_X x, y \rangle$ for all x and y in \mathcal{K} ; such expectations are assumed to exist. The quality that R_X is positive definite (i.e. $\langle R_X x, x \rangle > 0$ for all nonzero $x \in \mathcal{K}$) is denoted by $R_X > 0$.

The normal (or Gaussian) distribution with mean μ and covariance operator R is denoted by $\mathcal{N}_{\mathcal{K}}(\mu, R)$; that is, $X \sim \mathcal{N}(\mu, R)$ if and only if $\langle X - \mu, x \rangle$ is normally distributed on the real line (with mean 0 and variance $\langle Rx, x \rangle$) for all $x \in \mathcal{K}$. Wherever it appears in the paper, the notation $Z_{\mathcal{K}}$ will denote a \mathcal{K} -valued standard normal r.v.; that is, $Z_{\mathcal{K}} \sim \mathcal{N}_{\mathcal{K}}(0, 1_{\mathcal{K}})$. We note that $\mathbb{P}(Z_{\mathcal{K}} \in \mathcal{U}) = 0$ whenever \mathcal{U} is a proper linear (or affine) subspace of \mathcal{K} .

For $d = \dim \mathcal{K} > 0$ and any $\mu \in \mathcal{K}$, we say that $|Z_{\mathcal{K}} + \mu|^2 \sim \chi_d^2(|\mu|^2)$, with this distribution being called the noncentral chi-squared distribution with d degrees of freedom and noncentrality parameter $|\mu|^2$; the distribution is called central when $\mu = 0$, and then is denoted by χ_d^2 . If $W_1 \sim \chi_{\nu_1}^2(\lambda)$ and $W_2 \sim \chi_{\nu_2}^2$ are independent, we write $(W_1/\nu_1)/(W_2/\nu_2) \sim \mathcal{F}_{\nu_1, \nu_2}(\lambda)$, with this latter distribution being called the noncentral \mathcal{F} distribution (with numerator and denominator degrees of freedom ν_1 and ν_2 , respectively, and noncentrality parameter λ); if $\lambda = 0$, the distribution is called central, and denoted $\mathcal{F}_{\nu_1, \nu_2}$. We shall occasionally use χ_d^2 or $\mathcal{F}_{\nu_1, \nu_2}$ to denote a r.v. with the appropriate distribution; taken in context, there should be no confusion as to when this notation represents a distribution or a r.v.

2.2 Bounding the convergence rate of the F -statistic

Recall the model assumptions (2.1) and (2.2), as well as the hypotheses (2.3), and let

$$0 < k_0 := \dim \mathcal{V}_0 < k := \dim \mathcal{V} < n := \dim \mathcal{H}. \quad (2.4)$$

Under the additional assumption that ε is normally distributed, it is a straightforward exercise to demonstrate that the likelihood-ratio test rejects the null hypothesis for large values of the test statistic

$$F := \frac{|\Pi_{\mathcal{V}_1} Y|^2 / k_1}{|\Pi_{\mathcal{V}^\perp} Y|^2 / \mathbf{n}} = \frac{|\Pi_{\mathcal{V}_1} (\varepsilon + \mu / \sigma)|^2 / k_1}{|\Pi_{\mathcal{V}^\perp} \varepsilon|^2 / \mathbf{n}}, \quad (2.5)$$

where

$$\mathcal{V}_1 := \mathcal{V}_0^\perp \cap \mathcal{V}, \quad (2.6)$$

$$k_1 := \dim \mathcal{V}_1 = k - k_0, \quad \text{and} \quad \mathbf{n} := \dim \mathcal{V}^\perp = n - k. \quad (2.7)$$

Let us agree to say that $F = 0$ on the event $\{Y \in \mathcal{V}_0\}$ and $F = \infty$ on the event $\{Y \in \mathcal{V} \setminus \mathcal{V}_0\}$; these events have 0 probability when ε is Gaussian.

Since $\mathcal{V}_1 \perp \mathcal{V}^\perp$, the statistics in the numerator and denominator of (2.5) are independent (when ε is Gaussian) whence $F \sim \mathcal{F}_{k_1, n}(\lambda)$, where $\lambda = |\Pi_{\mathcal{V}_1} \mu|^2 / \sigma^2$; moreover, $\Pi_{\mathcal{V}_1}(\varepsilon + \mu / \sigma) = \Pi_{\mathcal{V}_1} \varepsilon$ when $\mu \in \mathcal{V}_0$, so that $F \sim \mathcal{F}_{k_1, n}$ under the null hypothesis. Since $k_1 \mathcal{F}_{k_1, n}$ converges in distribution to $\chi_{k_1}^2$, we expect the null distribution of $k_1 F$ to be well-approximated by $\chi_{k_1}^2$ for large n , even if we drop the assumption that ε is normally distributed. With this observation in mind, we endeavor to bound

$$\Delta := \sup_{z \in \mathbb{R}, \mu \in \mathcal{V}_0, \sigma > 0} |\mathbb{P}(k_1 F \leq z) - \mathbb{P}(\chi_{k_1}^2 \leq z)| \quad (2.8)$$

by an expression which decreases to 0 as $n \rightarrow \infty$; of course, \mathcal{H} changes with n (and hence n), so this bound will necessarily depend on the manner in which \mathcal{H} and \mathcal{V} on n .

Let

$$\mathfrak{X} := \mathcal{V} \times \mathbb{R} = \{(u, \alpha) : u \in \mathcal{V}, \alpha \in \mathbb{R}\} \quad (2.9)$$

be a Hilbert space endowed with the usual product topology, so that

$$\langle (u, \alpha), (v, \beta) \rangle = \langle u, v \rangle + \alpha \beta \quad \text{for any } (u, \alpha), (v, \beta) \in \mathfrak{X}.$$

Further let

$$S := (\Pi_{\mathcal{V}} \varepsilon, |\varepsilon|^2 - n) \quad (2.10)$$

take values in \mathfrak{X} . By (2.2), $\mathbb{E} S = 0$ and $\Sigma := \text{Cov } S$ is defined by

$$\Sigma(u, \alpha) = (u + \alpha \varsigma, \langle \varsigma, u \rangle + \gamma^2 \alpha) \quad \text{for all } (u, \alpha) \in \mathfrak{X}, \quad (2.11)$$

where

$$\varsigma := \mathbb{E}(|\varepsilon|^2 - n) \Pi_{\mathcal{V}} \varepsilon \quad \text{and} \quad \gamma := \sqrt{\mathbb{E}(|\varepsilon|^2 - n)^2}. \quad (2.12)$$

With the above notation in mind, we are ready to state the main result of the paper:

Theorem 2.2.1. *Assume that (2.2) and (2.1) hold, and also that $\Sigma > 0$. Then $|\varsigma| < \gamma$, and*

$$\Delta \leq \text{BE}_{\mathfrak{X}}(S) + \Lambda_{\mathcal{V}, \mathcal{V}_0}(S), \quad (2.13)$$

where

$$\text{BE}_{\mathfrak{X}}(S) := \sup_{A \in \mathfrak{A}} |\mathbb{P}(S \in A) - \mathbb{P}(\Sigma^{1/2} Z_{\mathfrak{X}} \in A)|, \quad (2.14)$$

\mathfrak{A} is the collection of all convex subsets of \mathfrak{X} ,

$$\Lambda_{\mathcal{V}, \mathcal{V}_0}(S) := \mathfrak{K}_k \left(\frac{1}{\gamma \sqrt{1 - \rho^2}} + \frac{\gamma}{n} \left(1 + \sqrt{\frac{1 + \rho_1}{1 - \rho_1}} \right) \right) \leq \frac{\mathfrak{K}_k}{\sqrt{n}} \frac{(\gamma / \sqrt{n}) \vee (\sqrt{n} / \gamma)}{\sqrt{1 - \rho^2}}, \quad (2.15)$$

$$\rho := \frac{|\varsigma|}{\gamma} \quad \text{and} \quad \rho_1 := \frac{|\Pi_{\mathcal{V}_1} \varsigma|}{\gamma}. \quad (2.16)$$

Also, if $\{e_1, \dots, e_n\}$ is an orthonormal basis of \mathcal{H} and

$$\varepsilon_1, \dots, \varepsilon_n \text{ are independent, where } \varepsilon_j := \langle \varepsilon, e_j \rangle \text{ for } j = 1, \dots, n, \quad (2.17)$$

then

$$\begin{aligned} \text{BE}_{\mathfrak{X}}(S) &\leq \mathfrak{K}_k \sum_{j=1}^n \mathbb{E} |\Sigma^{-1/2}(\varepsilon_j \Pi_{\mathcal{V}} e_j, \varepsilon_j^2 - 1)|^3 \\ &\leq \frac{\mathfrak{K}_k}{(1 - \rho^2)^{3/2}} \sum_{j=1}^n \left(\mathbb{E} |\varepsilon_j|^3 |\Pi_{\mathcal{V}} e_j|^3 + \frac{\mathbb{E} |\varepsilon_j^2 - 1|^3}{\gamma^3} \right). \end{aligned} \quad (2.18)$$

The bound in (2.13) is, as suggested by its appearance, the result of two approximations. From (2.5) and (2.10), it is easy to see that $k_1 F = f(S)$ for a nonlinear function $f: \mathfrak{X} \rightarrow [0, \infty]$. The distribution of $f(S)$ is first approximated by the distribution of $f(\Sigma^{1/2} Z_{\mathfrak{X}})$, giving rise to the term $\text{BE}_{\mathfrak{X}}(S)$, and then distribution of $f(\Sigma^{1/2} Z_{\mathfrak{X}})$ is approximated by the $\chi_{k_1}^2$ distribution, yielding the term $\Lambda_{\mathcal{V}, \nu_0}(S)$.

Let us investigate the two aforementioned approximations a bit further. It is no coincidence that a coordinate system for \mathcal{H} is not introduced until we bound the term $\text{BE}_{\mathfrak{X}}(S)$ in (2.18). Indeed, by (2.17), $\varepsilon = \sum_j \varepsilon_j e_j$ and $|\varepsilon|^2 = \sum_j \varepsilon_j^2$, so that S in (2.10) is seen to be the sum of the n independent zero-mean r.v.'s $(\varepsilon_j \Pi_{\mathcal{V}} e_j, \varepsilon_j^2 - 1)$; that is, S is a linear statistic under the independence assumption in (2.17). We may then apply any published BE-type bounds for the multivariate normal distribution (as in e.g. [44, 13, 18]) to estimate the distribution of S by that of $\Sigma^{1/2} Z_{\mathfrak{X}}$.

While $\Lambda_{\mathcal{V}, \nu_0}(S)$ is bounded in (2.15) independently of a coordinate system, note that the assumption in (2.17) implies

$$\varsigma = \sum_{j=1}^n \mathbb{E}(\varepsilon_j^2 - 1) \varepsilon_j \Pi_{\mathcal{V}} e_j = \sum_{j=1}^n \mathbb{E} \varepsilon_j^3 \Pi_{\mathcal{V}} e_j \quad \text{and} \quad \gamma^2 = \sum_{j=1}^n \mathbb{E}(\varepsilon_j^2 - 1)^2. \quad (2.19)$$

If each of the ε_j 's have no skew (i.e. $\mathbb{E} \varepsilon_j^3 = 0$), then $\varsigma = 0$ and hence $\rho_1 = \rho = 0$ by (2.16). The bounds in both (2.15) and (2.18) are seen to be decreasing with respect to ρ and ρ_1 , and so are minimized only if $\varsigma = 0$.

Remark 2.2.2. The non-degeneracy condition $\Sigma > 0$ is equivalent to $|\varsigma| < \gamma$. Indeed, by (2.11), $\langle \Sigma x, x \rangle = \gamma^2 \alpha^2 + 2\langle \varsigma, u \rangle \alpha + |u|^2$ for all $x = (u, \alpha) \in \mathfrak{X}$. Viewing $\langle \Sigma x, x \rangle$ as a quadratic polynomial in α , $\langle \Sigma x, x \rangle > 0$ for all nonzero $x \in \mathfrak{X}$ is equivalent to the discriminant $4(\langle \varsigma, u \rangle^2 - \gamma^2 |u|^2)$ being negative for all nonzero $u \in \mathcal{V}$; note that this latter condition also implies $\gamma > 0$, since $\dim \mathcal{V} \geq 2$ implies the existence of a nonzero $u \in \varsigma^\perp$. Since $\langle \varsigma, u \rangle^2 < \gamma^2 |u|^2$ for all nonzero $u \in \mathcal{V}$ is equivalent to $|\langle \varsigma, u \rangle| < \gamma$ for all unit vectors $u \in \mathcal{V}$, the Cauchy-Schwarz inequality yields the equivalence of $\Sigma > 0$ and $|\varsigma| < \gamma$.

On the other hand, the degeneracy condition $\Sigma \not> 0$ is equivalent to the existence of $\xi \in \mathcal{V}$ such that $\langle \varepsilon, \xi \rangle = |\varepsilon|^2 - n$ almost surely (a.s.). Indeed, if $|\varsigma| = \gamma$ (equivalent to $\Sigma \not> 0$, by the previous paragraph), then $0 = \langle \Sigma(\varsigma, -1), (\varsigma, -1) \rangle = \mathbb{E}(\langle \Pi_{\mathcal{V}} \varepsilon, \varsigma \rangle - (|\varepsilon|^2 - n))^2 = \mathbb{E}(\langle \varepsilon, \varsigma \rangle - (|\varepsilon|^2 - n))^2$, so that $\langle \varepsilon, \varsigma \rangle = |\varepsilon|^2 - n$ a.s. Conversely, if $\langle \varepsilon, \xi \rangle = |\varepsilon|^2 - n$ a.s. for some $\xi \in \mathcal{V}$, then $\gamma^2 = \mathbb{E} \langle \varepsilon, \xi \rangle^2 = |\xi|^2$ and $\langle \varsigma, u \rangle = \mathbb{E} \langle \varepsilon, \xi \rangle \langle \Pi_{\mathcal{V}}, u \rangle = \langle \xi, u \rangle$ for all $u \in \mathcal{V}$, so that $|\varsigma|^2 = \langle \xi, \varsigma \rangle = |\xi|^2 = \gamma^2$. When (2.17) is also assumed, we see that $\Sigma \not> 0$ if and only if $\varepsilon_j^2 - 1 = \varsigma_j \varepsilon_j$ a.s. for $j = 1, \dots, n$; cf. (2.19). This latter condition is equivalent to each ε_j being supported on the set $\{\varsigma_j/2 \pm ((\varsigma_j/2)^2 - 1)^{1/2}\}$ for some $\varsigma \in \mathcal{V}$.

Theorem 2.2.1 provides a bound on the Kolmogorov distance between $k_1 F$ and $\chi_{k_1}^2$.

There is no essential loss in accuracy if we approximate the null distribution of F by $\mathcal{F}_{k_1, \mathbf{n}}$. Let $\chi_{k_1}^2$ and $\chi_{\mathbf{n}}^2$ be independent r.v.'s, so that $\mathcal{F}_{k_1, \mathbf{n}} = (\chi_{k_1}^2/k_1)/(\chi_{\mathbf{n}}^2/\mathbf{n})$, and let p denote the density of $\chi_{k_1}^2$. Then, for any $z > 0$,

$$\begin{aligned} \left| \mathbb{P}(k_1 \mathcal{F}_{k_1, \mathbf{n}} \leq z) - \mathbb{P}(\chi_{k_1}^2 \leq z) \right| &= \left| \mathbb{P}(\chi_{k_1}^2 \leq z \chi_{\mathbf{n}}^2/\mathbf{n}) - \mathbb{P}(\chi_{k_1}^2 \leq z) \right| \\ &\leq \mathbb{P}(|\chi_{\mathbf{n}}^2 - \mathbf{n}| > \frac{\mathbf{n}}{2}) + \left| \mathbb{E} \mathbb{I} \left\{ |\chi_{\mathbf{n}}^2 - \mathbf{n}| \leq \frac{\mathbf{n}}{2} \right\} \int_0^{(\chi_{\mathbf{n}}^2 - \mathbf{n})/\mathbf{n}} zp(z(1+\theta)) d\theta \right| \\ &\leq \frac{\mathbb{E}|\chi_{\mathbf{n}}^2 - \mathbf{n}|}{\mathbf{n}/2} + \sup_{\frac{1}{2}z < w < \frac{3}{2}z} zp(w) \frac{\mathbb{E}|\chi_{\mathbf{n}}^2 - \mathbf{n}|}{\mathbf{n}}. \end{aligned}$$

Since $\sup_{z/2 < w < 3z/2} zp(w) \leq 2 \sup_{w>0} wp(w) = \mathcal{O}(\sqrt{k_1})$ and $\mathbb{E}|\chi_{\mathbf{n}}^2 - \mathbf{n}| \leq \sqrt{\mathbb{E}(\chi_{\mathbf{n}}^2 - \mathbf{n})^2} = \sqrt{2\mathbf{n}}$, Theorem 2.2.1, along with the above remarks, implies

Corollary 2.2.3. *The statement of Theorem 2.2.1 holds when Δ is replaced by*

$$\sup_{z \in \mathbb{R}, \mu \in \mathcal{V}_0, \sigma > 0} \left| \mathbb{P}(F \leq z) - \mathbb{P}(\mathcal{F}_{k_1, \mathbf{n}} \leq z) \right|.$$

Remark 2.2.4. The assumption that $\text{Cov } \varepsilon = 1_{\mathcal{H}}$ in (2.2) is not essential. Indeed, if $\text{Cov } \varepsilon = R$, for some linear operator $R > 0$, the statement of Theorem 2.2.1 holds when ε is replaced by $R^{-1/2}\varepsilon$. Alternatively, the scalar product $\langle \cdot, \cdot \rangle$ can always be chosen such that $\text{Cov } \varepsilon = 1_{\mathcal{H}}$. The largest potential problem with a misspecified covariance structure is that the coordinates of the error ε are not independent; in such a case, (2.18) will not hold. Rinott [125, Theorem 1] provides a BE-type bound for sums of *bounded* r.v.'s under a local dependence structure which could potentially be used when (2.17) is not assumed.

The bound in (2.13) appears to be new to the literature; we have found no mention of a BE-type bound for any F -statistic (used in GLM's or for other purposes). Several recent results have been published concerning the asymptotic expansion of the null distribution of the F -statistic under non-normality; these include Yanagihara [146], concerning the linear regression model, and Yanagihara [145] and Harrar [47], concerning the one-way ANOVA model.

2.3 Applications

In this section we consider the bound on Δ when it is applied to several commonly used hypothesis testing problems. For the sake of simplicity, we will make the (common in practice) assumption that the coordinates of ε are i.i.d. Then the following corollary to Theorem 2.2.1 is nearly immediate.

Corollary 2.3.1. *Let the conditions of Theorem 2.2.1 hold, and assume that*

$$\varepsilon_1, \dots, \varepsilon_n \text{ are i.i.d.} \tag{2.20}$$

Then

$$\begin{aligned}\Delta &\leq \frac{\mathfrak{K}_k}{\sqrt{n}} \left(\frac{\mathbb{E}|\varepsilon_1|^3}{(1-\rho^2)^{3/2}} \Gamma_{\mathcal{V}} + \frac{\mathbb{E}|\varepsilon_1^2 - 1|^3}{\tilde{\gamma}^3(1-\rho^2)^{3/2}} + \frac{1}{\tilde{\gamma}(1-\rho^2)^{1/2}} + \tilde{\gamma} \left(1 + \sqrt{\frac{1+\rho_1}{1-\rho_1}} \right) \right) \\ &\leq \frac{\mathfrak{K}_k}{(1-\rho^2)^{3/2}\sqrt{n}} \left(\mathbb{E}|\varepsilon_1|^3 \Gamma_{\mathcal{V}} + \frac{\mathbb{E}|\varepsilon_1^2 - 1|^3}{\tilde{\gamma}^3} + \tilde{\gamma} \vee \frac{1}{\tilde{\gamma}} \right),\end{aligned}\quad (2.21)$$

where

$$\Gamma_{\mathcal{V}} := \sqrt{n} \sum_{j=1}^n |\Pi_{\mathcal{V}} e_j|^3 \quad \text{and} \quad \tilde{\gamma} := \sqrt{\mathbb{E}(\varepsilon_1^2 - 1)^2}. \quad (2.22)$$

Also,

$$\rho = \frac{|\mathbb{E} \varepsilon_1^3|}{\tilde{\gamma}} \left| \Pi_{\mathcal{V}} \frac{\mathbf{j}_{\mathcal{H}}}{|\mathbf{j}_{\mathcal{H}}|} \right| \quad \text{and} \quad \rho_1 = \frac{|\mathbb{E} \varepsilon_1^3|}{\tilde{\gamma}} \left| \Pi_{\mathcal{V}_1} \frac{\mathbf{j}_{\mathcal{H}}}{|\mathbf{j}_{\mathcal{H}}|} \right|, \quad \text{where} \quad \mathbf{j}_{\mathcal{H}} := \sum_{j=1}^n e_j. \quad (2.23)$$

The proof of Corollary 2.3.1 consists of taking note of a few simple implications obtained from the i.i.d. assumption in (2.20). Namely, $\varsigma = \mathbb{E} \varepsilon_1^3 \Pi_{\mathcal{V}} \mathbf{j}_{\mathcal{H}}$ and $\gamma = \sqrt{n} \tilde{\gamma}$ follows from (2.19) and (2.20); since $|\mathbf{j}_{\mathcal{H}}| = \sqrt{n}$, (2.23) is implied by (2.16). Further, by (2.7), $\mathbf{n} \leq n \leq \mathfrak{K}_k \mathbf{n}$ and so $\gamma/\mathbf{n} \leq \mathfrak{K}_k \tilde{\gamma}/\sqrt{n}$. Then (2.21) is a result of applying the above observations to the inequalities (2.13), (2.15), and (2.18).

Let k , k_0 , and k_1 be fixed constants. Then we see from (2.21) that the bound on Δ is of the order $\mathcal{O}(1/\sqrt{n})$ as $n \rightarrow \infty$ whenever: the common model assumptions (2.1), (2.2) and (2.20) hold; the non-degeneracy condition $|\varsigma| < \gamma$ holds; $\mathbb{E} \varepsilon_1^6 < \infty$; and $\Gamma_{\mathcal{V}} = \mathcal{O}(1)$. By Remark 2.2.2, we see that, under the assumption of (2.20), $|\varsigma| = \gamma$ if and only if ε_1 has a two-point distribution and $\mathbf{j}_{\mathcal{H}} \in \mathcal{V}$. The assumption that ε_1 has a finite sixth moment follows naturally from the use of a BE-type bound, where the finiteness of the third absolute moment of the summands of S is required to obtain a bound of the desired order.

The only other requirement for Δ to be $\mathcal{O}(1/\sqrt{n})$ is that $\Gamma_{\mathcal{V}}$ in (2.22) be $\mathcal{O}(1)$. Note that $|\Pi_{\mathcal{V}} e_j|^2$ is the i^{th} diagonal entry in the matrix representation of $\Pi_{\mathcal{V}}$. This number is often called the *leverage* of the i^{th} observation, and it is often used in regression model diagnostics. Particularly, if the leverage for any observation is large, relative to the leverages of (most of) the remaining observations, the model is usually deemed to be poorly designed (as the F -statistic is then no longer robust to non-normality). Thus, if the experiment is designed well so that each of the leverages are roughly equal to one another (i.e. approximately $1/n$), we see that $\Gamma_{\mathcal{V}} = \mathcal{O}(1)$ and so $\Delta = \mathcal{O}(1/\sqrt{n})$ (assuming that the previously mentioned model and moment assumptions hold). In the applications that follow (including linear regression, and one- and two-way ANOVA models), sufficient conditions for $\Gamma_{\mathcal{V}} = \mathcal{O}(1)$ are provided; after investigating such specific applications, it should be clear to the reader that the ANCOVA and MANOVA models also fall under the auspices of Corollary 2.3.1.

2.3.1 Simple linear regression

Identify \mathcal{H} with \mathbb{R}^n , and take any $x \in \mathcal{H}$ that is linearly independent from $\mathbf{j}_{\mathcal{H}}$; that is, assume that x is not a “constant vector” $\alpha \mathbf{j}_{\mathcal{H}}$ for some $\alpha \in \mathbb{R}$. Let $\bar{x} = \langle x, \mathbf{j}_{\mathcal{H}} \rangle / |\mathbf{j}_{\mathcal{H}}|^2 = \sum_j x_j / n$ and $x_c = x - \bar{x} \mathbf{j}_{\mathcal{H}}$, so that $\mathbf{j}_{\mathcal{H}} \perp x_c$. Then the simple linear regression model is defined by the

2-dimensional subspace

$$\mathcal{V} = \text{span}\{\mathbf{j}_{\mathcal{H}}, x\} = \text{span}\{\mathbf{j}_{\mathcal{H}}, x_c\};$$

of course, the null subspace associated with this model is nearly always taken to be $\mathcal{V}_0 = \text{span}\{\mathbf{j}_{\mathcal{H}}\}$, so that $\mathcal{V}_1 = \text{span}\{x_c\}$. By (2.22),

$$\Gamma_{\mathcal{V}} = \sqrt{n} \sum_{j=1}^n \left(\frac{\langle e_j, \mathbf{j}_{\mathcal{H}} \rangle^2}{|\mathbf{j}_{\mathcal{H}}|^2} + \frac{\langle e_j, x_c \rangle^2}{|x_c|^2} \right)^{3/2} \leq n^{3/2} \left(\frac{1}{n} + \max_j \frac{\langle e_j, x_c \rangle^2}{|x_c|^2} \right)^{3/2} = \mathcal{O}(\delta^3),$$

where

$$\delta = \max_j \frac{|\langle e_j, x_c \rangle|}{|x_c|/\sqrt{n}} = \max_j \frac{|x_j - \bar{x}|}{s_x}, \quad \text{and} \quad s_x = \frac{|x_c|}{\sqrt{n}} = \sqrt{\frac{1}{n} \sum_j (x_j - \bar{x})^2}.$$

That is, $\Gamma_{\mathcal{V}} = \mathcal{O}(1)$ whenever x is chosen to depend on n such that $\delta = \mathcal{O}(1)$; this condition can be loosely interpreted as equivalent to the absence of any “outliers” in the regressor x . In particular, suppose that (a_n) and (b_n) are any two sequences of real numbers such that $a_n < b_n$, and that the components of x are equally spaced on the interval $[a_n, b_n]$ for each $n \in \mathbb{N}$. Then $\bar{x} = \frac{1}{2}(a_n + b_n)$, $\max_j |x_j - \bar{x}| = \frac{1}{2}(b_n - a_n)$, and $s_x = (b_n - a_n)\sqrt{(n+1)/(12(n-1))}$, so that $\delta \leq \sqrt{3} = \mathcal{O}(1)$.

We could also easily incorporate the model with no “intercept” term; that is, let $\mathcal{V} = \text{span}\{x\}$ and $\mathcal{V}_0 = \{0\}$. In this case,

$$\Gamma_{\mathcal{V}} = \sqrt{n} \sum_{j=1}^n \frac{|x_j|^3}{|x|^3} = \mathcal{O} \left(\frac{\max_j |x_j|}{|x|/\sqrt{n}} \right)^3.$$

2.3.2 One-way ANOVA

For any natural number $k \geq 2$, let $\mathcal{H} = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$, so that $n = \dim \mathcal{H} = \sum_j n_j$. For $i = 1, \dots, k$, assume that $\{e_{i,1}, \dots, e_{i,n_i}\} \subseteq \mathcal{H}$ forms an orthonormal basis for \mathbb{R}^{n_i} , and let $\mathbf{j}_i := \sum_{j=1}^{n_i} e_{i,j}$. Letting

$$\mathcal{V} := \{(\tau_1 \mathbf{j}_1, \dots, \tau_k \mathbf{j}_k) : (\tau_1, \dots, \tau_k) \in \mathbb{R}^k\} = \text{span}\{\mathbf{j}_1, \dots, \mathbf{j}_k\},$$

we call the model in (2.1) a one-way ANOVA; the model is called balanced when $n_1 = \cdots = n_k = m$. For any $y \in \mathcal{H}$, let $\bar{y}_i := \langle y, \mathbf{j}_i \rangle / |\mathbf{j}_i|^2 = \sum_{j=1}^{n_i} y_{i,j} / n_i$ for $i = 1, \dots, k$. Then

$$\Pi_{\mathcal{V}} y = (\bar{y}_1 \mathbf{j}_1, \dots, \bar{y}_k \mathbf{j}_k),$$

whence

$$\sum_{i=1}^k \sum_{j=1}^{n_i} |\Pi_{\mathcal{V}} e_{i,j}|^3 = \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{1}{n_i^{3/2}} = \sum_{i=1}^k \frac{1}{\sqrt{n_i}} = \frac{1}{\sqrt{n}} \sum_{i=1}^k \sqrt{\frac{n}{n_i}} \leq \sqrt{\frac{\max_i n_i}{\min_i n_i}} \frac{k^{3/2}}{\sqrt{n}}. \quad (2.24)$$

Suppose we let $\mathcal{V}_0 = \text{span}\{\mathbf{j}_{\mathcal{H}}\} = \text{span}\{(\mathbf{j}_1, \dots, \mathbf{j}_k)\}$, which is equivalent to testing the null hypothesis that the k “treatment means” are all equal. Let $Y = (Y_1, \dots, Y_k)$, where $Y_i = (Y_{i,1}, \dots, Y_{i,n_i})$ for $i = 1, \dots, k$, $\Pi_{\mathcal{V}} Y = (\bar{Y}_1 \mathbf{j}_1, \dots, \bar{Y}_k \mathbf{j}_k)$, and $\Pi_{\mathcal{V}_0} Y = (\bar{Y} \mathbf{j}_1, \dots, \bar{Y} \mathbf{j}_k)$.

Then the test statistic in (2.5) takes the well-known form

$$F = \frac{\sum_{i=1}^k |(\bar{Y}_i - \bar{Y})\mathbf{j}_i|^2 / (k-1)}{\sum_{i=1}^k |Y_i - \bar{Y}\mathbf{j}_i|^2 / (n-k)} = \frac{\sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2 / (k-1)}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i)^2 / (n-k)}.$$

Then Corollary 2.3.1 provides a bound of order $\mathcal{O}(1/\sqrt{n})$ on the distance between F and $\mathcal{F}_{k-1, n-k}$ provided $\max_i n_i / \min_i n_i$ is absolutely bounded for $n > k$ (and the stated conditions on ε_j are satisfied), which is certainly the case in the balanced model. Note that tests for orthogonal contrasts may also be performed by an appropriate choice of \mathcal{V}_0 ; while (2.21) may not be applicable (if \mathcal{V}_0 does not contain $\mathbf{j}_{\mathcal{H}}$), the bound on Δ_n is still of order $\mathcal{O}(1/\sqrt{n})$ in the balanced model.

2.3.3 Two-way ANOVA

Take any natural numbers $a \geq 2$ and $b \geq 2$, and identify \mathcal{H} with the set of all $a \times b$ matrices $y = [y_{i,j}]$ whose entries $y_{i,j}$ are vectors in \mathbb{R}^m for $i = 1, \dots, a$ and $j = 1, \dots, b$; that is, $n = \dim \mathcal{H} = abm$. We implement the balanced two-way ANOVA model here for the sake of simplicity, though of course an unbalanced model is easily obtained. Similarly to the one-way ANOVA model, assume that $\{e_{i,j,1}, \dots, e_{i,j,m}\} \subseteq \mathcal{H}$ forms an orthonormal basis for the (i,j) entry of an element of \mathcal{H} ; also let $\bar{y}_{i,j} = \langle y, \mathbf{j}_{i,j} \rangle / |\mathbf{j}_{i,j}|^2$, where $\mathbf{j}_{i,j} = \sum_{\ell=1}^m e_{i,j,\ell}$. Let

$$\mathcal{V} = \{[\mu_{i,j}\mathbf{j}_{i,j}] \in \mathcal{H} : \mu_{i,j} \in \mathbb{R}\}, \quad \text{so that} \quad \Pi_{\mathcal{V}} y = [\bar{y}_{i,j}\mathbf{j}_{i,j}]$$

for all $y \in \mathcal{H}$; it is then easy to see that $\sum_{i=1}^a \sum_{j=1}^b \sum_{\ell=1}^m |\Pi_{\mathcal{V}} e_{i,j,\ell}|^3 = ab/\sqrt{m} = k^{3/2}/\sqrt{n}$, where $k = \dim \mathcal{V} = ab$.

Now let

$$\mathcal{R} = \{[\mu_{i,j}\mathbf{j}_{i,j}] \in \mathcal{V} : \mu_{i,1} = \mu_{i,2} = \dots = \mu_{i,b} \text{ for } i = 1, \dots, a\}$$

and

$$\mathcal{C} = \{[\mu_{i,j}\mathbf{j}_{i,j}] \in \mathcal{V} : \mu_{1,j} = \mu_{2,j} = \dots = \mu_{a,j} \text{ for } j = 1, \dots, b\}$$

denote the subspaces of \mathcal{V} which consist of matrices with constant rows and constant columns, respectively; note that $\dim \mathcal{R} = a$ and $\dim \mathcal{C} = b$. Further let

$$\mathcal{V}_0 = \mathcal{R} + \mathcal{C} = \{[\mu_{i,j}\mathbf{j}_{i,j}] \in \mathcal{V} : \mu_{i,j} = \mu_{i\bullet} + \mu_{\bullet j} \text{ for all } i = 1, \dots, a, j = 1, \dots, b\};$$

that is, we view $\mu \in \mathcal{V}_0$ as the sum of an independent “row effect” and “column effect”, whereas $\mu \in \mathcal{V}$ is viewed as the sum of a row, column, and “interaction” effect. It is easy to see that $\Pi_{\mathcal{R}}\Pi_{\mathcal{C}} = \Pi_{\mathcal{C}}\Pi_{\mathcal{R}} = \Pi_{\mathbf{j}_{\mathcal{H}}}$, whence $\Pi_{\mathcal{V}_0} = \Pi_{\mathcal{R}} + \Pi_{\mathcal{C}} - \Pi_{\mathbf{j}_{\mathcal{H}}}$ (cf. [4, Theorems 22.2 and 22.7]). As $\mathbf{j}_{\mathcal{H}} \in \mathcal{V}_0$, Corollary 2.3.1 provides a bound of order $\mathcal{O}(1/\sqrt{n})$ on the distance between the distribution of

$$F = \frac{|(\Pi_{\mathcal{V}} - \Pi_{\mathcal{R}} - \Pi_{\mathcal{C}} + \Pi_{\mathbf{j}_{\mathcal{H}}})Y|^2 / ((a-1)(b-1))}{|Y - \Pi_{\mathcal{V}} Y|^2 / (n-ab)}$$

and $\mathcal{F}_{(a-1)(b-1), n-ab}$. That $k_1 = \dim \mathcal{V}_1 = \dim \mathcal{V}_0^\perp \cap \mathcal{V} = (a-1)(b-1)$ follows from noting that

$$\Pi_{\mathcal{V}} = \Pi_{\mathbf{j}_{\mathcal{H}}} + (\Pi_{\mathcal{R}} - \Pi_{\mathbf{j}_{\mathcal{H}}}) + (\Pi_{\mathcal{C}} - \Pi_{\mathbf{j}_{\mathcal{H}}}) + (\Pi_{\mathcal{V}} - \Pi_{\mathcal{R}} - \Pi_{\mathcal{C}} + \Pi_{\mathbf{j}_{\mathcal{H}}})$$

yields an orthogonal decomposition of the model space \mathcal{V} (cf. [4, Theorem 22.5 and 22.6]); note this decomposition may also be used to test for the presence of a “row effect” or “column effect” by appropriately choosing \mathcal{V}_0 .

2.4 Proof of Theorem 2.2.1

By Remark 2.2.2, $\Sigma > 0$ implies $|\varsigma| < \gamma$, which further implies $\rho \in [0, 1)$; cf. (2.16).

Then let us first prove (2.18). Refer to (2.10) and (2.17) to see that

$$\Sigma^{-1/2}S = \sum_{j=1}^n X_j, \quad \text{where} \quad X_j := \Sigma^{-1/2}(\varepsilon_j \Pi_{\mathcal{V}} e_j, \varepsilon_j^2 - 1) \text{ for } j = 1, \dots, n,$$

so that $\Sigma^{-1/2}S$ is a sum of zero-mean independent r.v.'s with $\text{Cov}(\Sigma^{-1/2}S) = 1_{\mathfrak{X}}$. Since \mathfrak{X} is isometric to \mathbb{R}^{k+1} (cf. (2.9) and (2.4)), [18, Theorem 3.5] implies the first inequality of (2.18). From (2.11), it is straightforward to verify that

$$\Sigma^{-1}(u, \alpha) = \left(u - \frac{\alpha - \langle u, \varsigma \rangle}{\gamma^2 - |\varsigma|^2} \varsigma, \frac{\alpha - \langle u, \varsigma \rangle}{\gamma^2 - |\varsigma|^2} \right) \quad (2.25)$$

for all $(u, \alpha) \in \mathfrak{X}$, so that

$$\begin{aligned} |\Sigma^{-1/2}(u, \alpha)|^3 &= \left(|u|^2 + \frac{(\alpha - \langle u, \varsigma \rangle)^2}{\gamma^2 - |\varsigma|^2} \right)^{3/2} \\ &\leq \left(|u|^2 + \frac{2(\alpha^2 + \langle u, \varsigma \rangle^2)}{\gamma^2 - |\varsigma|^2} \right)^{3/2} \leq \left(\frac{1+\rho^2}{1-\rho^2} |u|^2 + \frac{2\alpha^2}{\gamma^2(1-\rho^2)} \right)^{3/2} \\ &\leq \frac{4}{(1-\rho^2)^{3/2}} (|u|^3 + |\alpha|^3/\gamma^3); \end{aligned}$$

Jensen's inequality (implying $|a + b|^p \leq 2^{p-1}(|a|^p + |b|^p)$ for any $p \geq 1$ and real numbers a and b) is used in the first and last inequalities above, and the Cauchy-Schwarz inequality and definition (2.16) of ρ are used in the second inequality above. The second inequality of (2.18) then follows.

It remains to prove (2.13). Take any $\mu \in \mathcal{V}_0$ and $\sigma > 0$, so that $\Pi_{\mathcal{V}_1}(\varepsilon + \mu/\sigma) = \Pi_{\mathcal{V}_1}\varepsilon = \Pi_{\mathcal{V}_1}\Pi_{\mathcal{V}}\varepsilon$; also, $|\Pi_{\mathcal{V}^\perp}\varepsilon|^2 = |\varepsilon|^2 - |\Pi_{\mathcal{V}}\varepsilon|^2$. Then the definitions (2.5) and (2.10) of F and S imply $k_1 F = f(S)$, where $f: \mathfrak{X} \rightarrow [0, \infty]$ is defined by

$$f(u, \alpha) = \frac{n|\Pi_{\mathcal{V}_1}u|^2}{\alpha + n - |u|^2} \quad \text{when } \alpha + n > |u|^2 \text{ or when } u \in \mathcal{V}_0 \text{ and } \alpha + n \geq |u|^2,$$

and $f(u, \alpha) = \infty$ for all other $(u, \alpha) \in \mathfrak{X}$. Let $A_z := \{x \in \mathfrak{X}: f(x) \leq z\}$ for any $z \in \mathbb{R}$, so that A_z is empty when $z < 0$,

$$A_0 = \{x \in \mathfrak{X}: f(x) = 0\} = \{(u, \alpha) \in \mathfrak{X}: u \in \mathcal{V}_0, \alpha + n \geq |u|^2\},$$

and, for any $z > 0$,

$$\begin{aligned} A_z &= \{(u, \alpha) \in \mathfrak{X} : \mathbf{n}|\Pi_{\mathcal{V}_1} u|^2 + z|u|^2 \leq z(\alpha + n)\} \\ &= \left\{ (u, \alpha) \in \mathfrak{X} : \mathbf{n}\left(\frac{1}{z}|\Pi_{\mathcal{V}_1} u|^2 - 1\right) + |u|^2 - k \leq \alpha \right\}; \end{aligned} \quad (2.26)$$

the last equality above follows from $n = \mathbf{n} + k$ (cf. (2.7)). Since $|\cdot|^2$ is a convex function, A_z is seen to be convex for all $z \in \mathbb{R}$. Then the definitions (2.8) and (2.14) of Δ and $\text{BE}_{\mathfrak{X}}(S)$ imply

$$\Delta \leq \text{BE}_{\mathfrak{X}}(S) + \sup_{z>0} |\mathbb{P}(\Sigma^{1/2} Z_{\mathfrak{X}} \in A_z) - \mathbb{P}(\chi_k^2 \leq z)|; \quad (2.27)$$

note that the supremum above is taken only over the positive reals, since both of the events $\{\Sigma^{1/2} Z_{\mathfrak{X}} \in A_z\}$ and $\{\chi_k^2 \leq z\}$ have zero probability when $z \leq 0$.

We next derive the density of $\Sigma^{1/2} Z_{\mathfrak{X}}$, for which $\det(\Sigma)$ is needed. Let $\Sigma x = \lambda x$ for some nonzero eigenvector $x = (u, \alpha) \in \mathfrak{X}$, so that (cf. (2.11)) $(\lambda - 1)u = \alpha\varsigma$ and $(\lambda - \gamma^2)\alpha = \langle u, \varsigma \rangle$. In the case when $\varsigma = 0$, we see that $\lambda = 1$ if and only if $\alpha = 0$ and that $\lambda = \gamma^2$ if and only if $u = 0$; that is, $\det(\Sigma) = \gamma^2 = \gamma^2 - |\varsigma|^2$ when $\varsigma = 0$. Suppose then that $\varsigma \neq 0$. Then $\lambda = 1$ if and only if $\alpha = 0$ and $\langle u, \varsigma \rangle = 0$. Otherwise, if $\lambda \neq 1$, $\alpha|\varsigma|^2 = (\lambda - 1)\langle u, \varsigma \rangle = (\lambda - 1)(\lambda - \gamma^2)\alpha$, so that $(\lambda - 1)(\lambda - \gamma^2) = |\varsigma|^2$. That is, $\lambda \in \{\lambda_1, \lambda_{-1}\}$, where $\lambda_s = (1 + \gamma^2 + s\sqrt{(1 + \gamma^2)^2 - 4(\gamma^2 - |\varsigma|^2)})/2$ for $s \in \{1, -1\}$. Each of the eigenspaces associated with λ_1 and λ_{-1} are easily seen to be 1-dimensional, so that $\det(\Sigma) = \lambda_1 \lambda_{-1} = \gamma^2 - |\varsigma|^2$. Then, by (2.25), the density of $\Sigma^{1/2} Z_{\mathfrak{X}} \sim \mathcal{N}_{\mathfrak{X}}(0, \Sigma)$ at the point $(u, \alpha) \in \mathfrak{X}$ is seen to be

$$\frac{\exp\{-\frac{1}{2}\langle \Sigma^{-1}(u, \alpha), (u, \alpha) \rangle\}}{(2\pi)^{(k+1)/2}(\gamma^2 - |\varsigma|^2)^{1/2}} = \frac{\exp\{-\frac{1}{2}|u|^2\}}{(2\pi)^{k/2}} \frac{\exp\{-\frac{1}{2}(\alpha - \langle \varsigma, u \rangle)^2/(\gamma^2 - |\varsigma|^2)\}}{\sqrt{2\pi(\gamma^2 - |\varsigma|^2)}}. \quad (2.28)$$

Now let

$$Z_{\mathcal{V}} \sim \mathcal{N}_{\mathcal{V}}(0, 1_{\mathcal{V}}) \quad \text{and} \quad Z_{\mathbb{R}} \sim \mathcal{N}_{\mathbb{R}}(0, 1) \quad \text{be independent r.v.'s,}$$

so that $Z_{\mathfrak{X}} = (Z_{\mathcal{V}}, Z_{\mathbb{R}})$ may be asserted without any loss of generality. Also let

$$W := |\Pi_{\mathcal{V}_1} Z_{\mathcal{V}}|^2, \quad T := \mathbf{n}\left(\frac{W}{z} - 1\right), \quad \text{and} \quad \zeta := \sqrt{\gamma^2 - |\varsigma|^2} Z_{\mathbb{R}} + \langle Z_{\mathcal{V}}, \varsigma \rangle. \quad (2.29)$$

Then, for any $z > 0$, (2.28), (2.29), and (2.26) imply

$$\begin{aligned} \mathbb{P}(\Sigma^{1/2} Z_{\mathfrak{X}} \in A_z) &= \mathbb{E} \mathbb{P}(\Sigma^{1/2}(Z_{\mathcal{V}}, Z_{\mathbb{R}}) \in A_z | Z_{\mathcal{V}}) = \mathbb{P}((Z_{\mathcal{V}}, \zeta) \in A_z) \\ &= \mathbb{P}(T + |Z_{\mathcal{V}}|^2 - k \leq \zeta) \\ &= \mathbb{P}(T \leq \zeta) + I_1, \end{aligned} \quad (2.30)$$

where I_1 is implicitly defined above. By first conditioning on $Z_{\mathcal{V}}$, an application of the mean-value theorem yields

$$|I_1| \leq \frac{\mathbb{E}||Z_{\mathcal{V}}|^2 - k|}{\sqrt{2\pi(\gamma^2 - |\varsigma|^2)}} \leq \frac{\mathfrak{K}_k}{\gamma\sqrt{1 - \rho^2}}. \quad (2.31)$$

Let X and Y be any real-valued r.v.'s such that (T, X, Y) and $(T, -X, -Y)$ share the same distribution. Then, for any $p \in [0, 1]$,

$$|\mathbf{P}(T \leq X + Y) - \mathbf{P}(T \leq 0)| \leq \mathbf{P}(X > p|T|) + \mathbf{P}(Y \geq (1 - p)|T|),$$

which follows from the simple observations $\{t \leq x + y\} \subseteq \{t \leq 0\} \cup \{x > pt > 0\} \cup \{y \geq (1 - p)t > 0\}$ and $\{t \leq 0\} \subseteq \{t \leq x + y\} \cup \{x < pt \leq 0\} \cup \{y \leq (1 - p)t \leq 0\}$, valid for any real numbers t, x , and y . Now let

$$X := \langle Z_{\mathcal{V}}, \Pi_{\mathcal{V}_1} \varsigma \rangle \quad \text{and} \quad Y := \sqrt{\gamma^2 - |\varsigma|^2} Z_{\mathbb{R}} + \langle Z_{\mathcal{V}}, \Pi_{\mathcal{V}_0} \varsigma \rangle, \quad (2.32)$$

so that $\mathbf{P}(T \leq X + Y) = \mathbf{P}(T \leq \zeta)$ follows from (2.29); note also that $\mathbf{P}(T \leq 0) = \mathbf{P}(W \leq z) = \mathbf{P}(\chi_k^2 \leq z)$. Thus, upon letting $p = \rho_1$ (and noting from (2.16) that $0 \leq \rho_1 \leq \rho < 1$),

$$\sup_{z>0} |\mathbf{P}(T \leq \zeta) - \mathbf{P}(\chi_k^2 \leq z)| \leq \sup_{z>0} \mathbf{P}(X > \rho_1|T|) + \sup_{z>0} \mathbf{P}(Y \geq (1 - \rho_1)|T|). \quad (2.33)$$

Then (2.13) follows from (2.27), (2.30), (2.31), (2.33), and the results of the following two lemmas.

Lemma 2.4.1. *One has*

$$\sup_{z>0} \mathbf{P}(X > \rho_1|T|) \leq \mathfrak{K}_{k_1} \frac{\gamma}{\mathbf{n}}. \quad (2.34)$$

Lemma 2.4.2. *One has*

$$\sup_{z>0} \mathbf{P}(Y \geq (1 - \rho_1)|T|) \leq \mathfrak{K}_{k_1} \frac{\gamma}{\mathbf{n}} \sqrt{\frac{1 + \rho_1}{1 - \rho_1}}. \quad (2.35)$$

Proof of Lemma 2.4.1. Recall the definitions (2.16) and (2.32) of ρ_1 and X , and assume that $\rho_1 \neq 0$, since otherwise $X = 0$ a.s. and (2.34) holds trivially. Let $\varsigma_1 := \Pi_{\mathcal{V}_1} \varsigma$. By decomposing \mathcal{V}_1 into an orthogonal sum along the subspace spanned by ς_1 , we see that

$$W = Z^2 + V, \quad \text{where } Z := \frac{X}{|\varsigma_1|} \sim \mathcal{N}_{\mathbb{R}}(0, 1) \text{ and } V := |\Pi_{\varsigma_1^\perp \cap \mathcal{V}_1} Z_{\mathcal{V}}|^2 \sim \chi_{k_1-1}^2; \quad (2.36)$$

if $k_1 = 1$ (i.e. if \mathcal{V}_1 is generated by ς_1), then identify V with a degenerate r.v. which takes the value 0 with probability 1. Further, Z and V are independent.

Let

$$s := \frac{\mathbf{n}}{\gamma}. \quad (2.37)$$

Then (2.36), (2.29), (2.16) imply

$$\begin{aligned} \mathbf{P}(X > \rho_1|T|) &= \mathbf{P}\left(-Z < s\left(\frac{W}{z} - 1\right) < Z\right) \\ &= \mathbf{P}\left(Z^2 - \frac{z}{s}Z - (z - V) < 0 < Z^2 + \frac{z}{s}Z - (z - V)\right). \end{aligned}$$

Whenever $b > 0$, any x that satisfies the inequalities $x^2 - bx - c < 0 < x^2 + bx - c$ will also

satisfy the inequalities

$$x < \frac{1}{2}(b + \sqrt{b^2 + 4c}) \leq \frac{1}{2}(\sqrt{b^2 + 4c_+} + b) \text{ and } x > \frac{1}{2}|b - \sqrt{b^2 + 4c}| \geq \frac{1}{2}(\sqrt{b^2 + 4c_+} - b).$$

Then, since Z and V have light-tailed distributions,

$$\mathbb{P}(X > \rho_1 | T|) \leq \frac{\mathfrak{K}_{k_1}}{s} + \mathbb{P}\left(a_1 < Z < (a_2 \wedge \sqrt{2 \ln s}), V \leq (z \wedge 2(k_1 - 1) \ln s)\right), \quad (2.38)$$

where

$$a_1 := \sqrt{\left(\frac{z}{2s}\right)^2 + (z - V)_+} - \frac{z}{2s}, \quad \text{and} \quad a_2 := a_1 + \frac{z}{s}. \quad (2.39)$$

From (2.39) it follows that

$$a_1 = \frac{2(z - V)_+}{\sqrt{\left(\frac{z}{s}\right)^2 + 4(z - V)_+} + \frac{z}{s}} \geq \tilde{a}_1 := \frac{\nu(z - V)_+}{\sqrt{(z - V)_+} \vee \frac{z}{s}} \quad (2.40)$$

where $\nu = 2/(1 + \sqrt{5})$. Then in bounding the probability on the RHS of the inequality in (2.38), we may assume that

$$z \leq \frac{s}{2}. \quad (2.41)$$

Indeed, if $z > \frac{s}{2}$ then $z - V > \frac{s}{2} - 2(k_1 - 1) \ln s = \frac{s}{2}(1 - 4(k_1 - 1)\frac{\ln s}{s})$ and $\frac{z - V}{z/s} > s - 4(k_1 - 1) \ln s$; then, by (2.40), $Z \geq a_1 \geq \tilde{a}_1 \geq c\sqrt{s}$ for large enough s and some positive number c , which contradicts $Z \leq \sqrt{2 \ln s}$ (note that we assume \mathfrak{K}_{k_1} in (2.34) is large enough, so that s may be assumed large enough).

Then (2.38), (2.39), (2.40), and an application of the mean-value theorem imply

$$\mathbb{P}(X > \rho_1 | T|) \leq \frac{1}{s}(\mathfrak{K}_{k_1} + I) \quad \text{where} \quad I := \int_0^{2(k_1 - 1) \ln s} z \varphi(\tilde{a}_1) p(v) dv, \quad (2.42)$$

where φ is the density of Z and p is the density of V . It remains only to demonstrate that $I \leq \mathfrak{K}_{k_1}$. Since $\varphi(\tilde{a}_1) \leq 1/\sqrt{2\pi}$, we may restrict the region of integration for I to be $v \leq \frac{z}{2}$ (since $\int_{v > z/2} zp(v) dv < \infty$) and also replace the integrand $z \varphi(\tilde{a}_1) p(v)$ of I with $(z - v) \varphi(\tilde{a}_1) p(v)$ (since $\int vp(v) dv < \infty$). Finally, in the region $v \leq \frac{z}{2}$ we have $(z - v)_+ = z - v \geq \frac{z}{2} \geq \frac{z^2}{s}$ by (2.41). Then, since $\int p(v) dv = 1$ and $(z - v) \varphi(\tilde{a}_1) = (z - v) \varphi(\nu \sqrt{z - v}) < \infty$ (by (2.40)), we see that $I \leq \mathfrak{K}_{k_1}$. \square

Proof of Lemma 2.4.2. From (2.29) and (2.32), note that Y and T are independent. Then let $Z \sim \mathcal{N}_{\mathbb{R}}(0, 1)$ be independent of $\Pi_{V_1} Z_V$, so that $Y = \sqrt{\gamma^2 - |\Pi_{V_1} \varsigma|^2} Z = \gamma \sqrt{1 - \rho_1^2} Z$ may be asserted without any loss of generality. Let p denote the density of W , and let

$$s := \frac{n}{\gamma} \sqrt{\frac{1 - \rho_1}{1 + \rho_1}}.$$

Then

$$\mathbb{P}(Y \geq (1 - \rho_1) | T|) = \mathbb{P}\left(Z > s \left| \frac{W}{z} - 1 \right| \right) = I_+ + I_-,$$

where

$$I_+ := \mathbb{P}\left(z < W < z\left(1 + \frac{Z}{s}\right)\right) \leq \frac{\mathbb{E}[\{Z > 0\}Z]}{s} \sup_{z < w} zp(w) \leq \frac{\mathfrak{R}_{k_1}}{s}$$

and

$$I_- := \mathbb{P}\left(z\left(1 - \frac{Z}{s}\right) < W < z\right) \leq \mathbb{P}\left(Z > \frac{s}{2}\right) + \frac{\mathbb{E}[\{0 < Z < \frac{s}{2}\}Z]}{s} \sup_{\frac{z}{2} < w} zp(w) \leq \frac{\mathfrak{R}_{k_1}}{s}.$$

Thus, (2.35) is proved. \square

Chapter 3

Relative efficiency between Kendall's, Spearman's, and Pearson's correlation statistics

3.1 Introduction

The Kendall T and Spearman S correlation statistics, denoted here by T and S , are the most common nonparametric measures of association between two random variables X and Y . We are interested in relative efficiency between T and S . Daniels [25] and Durbin and Stuart [30] observed that for certain samples the values of T and S may differ much from each other. Of course, for large sample sizes n such samples may or may not be rare, depending on the underlying distribution of the pair (X, Y) . There are a number of considerations that suggest that, at least under the null hypothesis, H_0 , of independence between X and Y , the test statistics T and S will perform rather similarly. Indeed, already the work by Kendall [68] (on T) and Hotelling and Pabst [56] (on S) showed that $\mathbf{E} T = \mathbf{E} S = 0$, $\text{Var } T \sim \frac{4}{9n}$, and $\text{Var } S \sim \frac{1}{n}$ under H_0 ; further, the conjecture by Kendall et al. [77] that the correlation $\rho_{T,S}$ between T and S converges to 1 was later proved in Daniels' [24] treatment of a general class of correlation coefficients. One then immediately has $\sqrt{n}(3T - 2S)$ converging to 0 in probability, so that T and S are almost proportional when H_0 is true. So, if the alternative distribution is close enough to the null one, one may expect that the Pitman asymptotic relative efficiency between T and S , $\text{ARE}_{T,S}$, will equal 1 for the independence test in a general class of models. Indeed, this was the conjecture we posited that led us to the present paper.

In fact, Farlie [33] showed that this is so for a general class of models in which the joint distribution function (d.f.) F_θ of X and Y depends linearly on the association parameter θ . Namely, if G and H are the fixed marginal d.f.'s of X and Y , and A and B are some functions satisfying some rather mild regularity conditions, then $\text{ARE}_{T,S} = 1$ in the test for independence when

$$F_\theta(x, y) = G(x)H(y)(1 + \theta A(G(x))B(H(y))). \quad (3.1)$$

Of course, the linearity in θ is a strong assumption, which does not hold in a number of models of interest. One such nonlinear model was studied by Plackett [122, 78], whose

The material contained in this chapter is in preparation for submission to the pre-print server <http://arxiv.org>.

construction consists in letting $F_\theta(x, y)$ be the root (lying between the Fréchet-Hoeffding bounds on bivariate d.f.'s) to the reparameterized equation

$$F_\theta(x, y) - G(x)H(y) = (e^\theta - 1)(G(x) - F_\theta(x, y))(H(y) - F_\theta(x, y)) \quad (3.2)$$

for $\theta \in (-\infty, \infty)$, where G and H are fixed marginals; then $\theta = 0$ corresponds to independence between X and Y . Mardia [79] showed that for Plackett's model as well, $\text{ARE}_{T,S} = 1$ in the independence test. Frank's model [35, 90, 39] is also nonlinear in θ :

$$(e^{-\theta F_\theta(x, y)} - 1)(e^{-\theta} - 1) = (e^{-\theta G(x)} - 1)(e^{-\theta H(y)} - 1) \quad (3.3)$$

for $\theta \in (-\infty, \infty)$ (where $F_0(x, y) = G(x)H(y)$ is defined by continuity). Johnson and Tenenbein [64, 63] construct models (using a method called weighted linear combinations) which allow one to specify a single marginal distribution (so that $G = H$) and an association parameter. In both of these types of models, it is straightforward to verify from the given expressions that $\text{ARE}_{T,S} = 1$ in the independence test.

Kimeldorf and Sampson [69] give a list of desirable properties for a bivariate model with fixed marginals; namely, in addition to certain regularity properties, they require that the model family of distributions represent all the range of possible degrees of association, from complete positive dependence to independence to complete negative dependence. One particularly simple construction was first proposed by Fréchet [36] – to let F_θ be a mixture of the distributions at the extremes of this range; however, such a model misses the independence case. Farlie's model, on the other hand, may miss the extremes. The reader is referred to [80] for more discussion on constructing bivariate models.

Our first, and perhaps the most interesting, result presented in this paper provides a sufficient condition for $\text{ARE}_{T,S} = 1$ in the independence test. Heuristically, this condition means that the family (F_θ) possesses a certain “smoothness” in the average association for distributions near that of independence; this condition is defined and elaborated upon in Section 3.3. It is also shown that this condition is necessary for $\text{ARE}_{T,S} = 1$ when the model has a certain ordered property which is commonly found in association models.

We then proceed to describe some novel examples of models where the aforementioned smoothness condition is not met, illustrating just how influential the departure from independence can be in an association model; to the best of our knowledge, it has not yet been demonstrated in the literature that such families of distributions exist. The models are quite simple in construction, consisting of regions of independence (where X and Y are uniformly distributed over certain squares in the plane) and regions of complete dependence (where X and Y are uniformly distributed along portions of the line segments $x = \pm y$). In two of these models, $\text{ARE}_{T,S} = 1$ in the test for independence, while $\text{ARE}_{T,S}$ is not 1 (and in one case is infinite) in the other two models.

We also make comparisons of the Pearson correlation statistic, R , with T and S in hypothesis testing problems. While the test of independence is likely the most common use for any three of these statistics, we note that they may also be used to test for some degree of positive or negative association between X and Y ; cf. Fieller et al. [32] and Kraemer [75]. The utility of R as a measure of association is greatest in parametric settings (particularly, when normality is assumed); we consider several models based on some particular transformation of the bivariate normal distribution, particularly in regards

to the effect on the efficiency of R as compared to T or S . It should be noted that in all of our models presented here the ARE between T , S , and R can be expressed explicitly.

Our last investigation concerns estimating the relative efficiency (RE) in a few select models previously mentioned. The usefulness of the ARE as a tool for comparing two test statistics lies in the ease with which it can be calculated. Its use, however, could be called into question as it is a limit, obtained by forcing the sample size of a hypothesis test to increase to ∞ . These advantages and disadvantages are reversed in the RE: the RE is a more natural comparison tool (being based on finite sample sizes), though its calculation is generally intractable. We present the results of several simulations of the RE in Section 3.5, and show that there is a remarkable agreement between the RE and ARE (at least, in the models considered there) as to which of R , S , or T is preferable in a testing problem.

3.2 An expression for the ARE between the correlation statistics T , S , and R

In this section we discuss the relative efficiency (RE) and the related concept of the asymptotic relative efficiency (ARE), between two sequences of competing hypothesis tests. A simple expression, suitable for our purposes, for the Pitman ARE is also provided; particularly, it will be shown that, under some rather mild conditions, the ARE between tests using the Pearson, Kendall, or Spearman correlation statistics is easily expressible in terms of certain expectations. Other notation and conventions, to be used throughout the remainder of the paper, are introduced in this section as well.

Let $\Theta \subseteq \mathbb{R}$ be an interval, possibly infinite, indexing a family of probability distributions $\{\mathbf{P}_\theta\}_{\theta \in \Theta}$. Take any (temporarily fixed) $\theta_0, \theta_1 \in \Theta$ and $\alpha, \beta \in (0, 1)$, and consider the problem of testing the null hypothesis $\theta = \theta_0$ against the alternative $\theta = \theta_1$ with significance level α and power at least $1 - \beta$. Suppose that we have two (sequences of) competing real-valued test statistics $T_1 = (T_{1,n})$ and $T_2 = (T_{2,n})$; that is, for each $n \in \mathbb{N}$ and $\theta \in \Theta$, $T_{j,n}$ is $\mathbf{P}_{\theta,n}$ -measurable for $j \in \{1, 2\}$, where $\mathbf{P}_{\theta,n}$ denotes the n -fold product measure of \mathbf{P}_θ . A natural way of comparing the two statistics T_1 and T_2 is to calculate

$$\text{RE}_{T_1, T_2} := \text{RE}_{T_1, T_2}(\theta_0, \theta_1, \alpha, \beta) := \frac{n_{T_2}(\theta_0, \theta_1, \alpha, \beta)}{n_{T_1}(\theta_0, \theta_1, \alpha, \beta)}, \quad (3.4)$$

where $n_{T_j} := n_{T_j}(\theta_0, \theta_1, \alpha, \beta)$ is the minimal sample size needed for the test statistic T_j to achieve a power of at least $1 - \beta$ while maintaining a significance level of α ; we assume that this “minimal sample size” is well-defined and finite for the test statistics under consideration. Then T_1 is seen to be a better choice for test statistic than T_2 whenever $\text{RE}_{T_1, T_2} > 1$, in that it requires fewer observations to realize the same power.

The greatest impediment to using the RE as a tool for comparing two test statistics is that the power function for a given test is generally infeasible to calculate. One common solution to this problem is to allow n_{T_1} and n_{T_2} to simultaneously increase to ∞ ; the limiting value of RE_{T_1, T_2} is then called the ARE between T_1 and T_2 . This limit is usually obtained by fixing two of the parameters θ_1 , α , and β , and then allowing the third to approach some appropriate limit; particularly, the Pitman [92], Bahadur [5], and Hodges-Lehmann [49] ARE’s are the limiting values of the RE when we let $\theta_1 \rightarrow \theta_0$, $\alpha \rightarrow 0$, or $\beta \rightarrow 0$, respectively. Certain conditions, concerning the test statistics and the manner in which the parameter

approaches its limit, must be met in order to assure the existence of the ARE.

We are concerned with the Pitman ARE in this paper. In order to ensure the existence of the (Pitman) ARE, it is enough to suppose we have the following type of uniform central limit theorem: for $j = 1, 2$, there exist continuous real-valued functions μ_{T_j} and σ_{T_j} on Θ such that

$$\sup_{\theta \in \mathcal{V}} \sup_{z \in \mathbb{R}} \left| \mathbb{P}_{\theta, n} \left(\frac{T_{j, n} - \mu_{T_j}(\theta)}{\sigma_{T_j}(\theta)/\sqrt{n}} \leq z \right) - \Phi(z) \right| \xrightarrow{n \rightarrow \infty} 0, \quad (3.5)$$

where Φ is the standard normal distribution function (d.f.) and \mathcal{V} is some neighborhood of θ_0 on which $\mu_{T_j}(\theta) - \mu_{T_j}(\theta_0) \neq 0$ (when $\theta \neq \theta_0$) and $\sigma_{T_j}(\theta) > 0$. The functions μ_{T_j} and σ_{T_j} are called the asymptotic mean and standard deviation, respectively, of the statistic T_j . Under the assumption that (3.5) holds, heuristics similar to those employed in [92] imply

$$\text{ARE}_{T_1, T_2}(\theta_0) = \frac{\sigma_{T_2}^2(\theta_0)}{\sigma_{T_1}^2(\theta_0)} \lim_{\theta \rightarrow \theta_0} \left(\frac{\mu_{T_1}(\theta) - \mu_{T_1}(\theta_0)}{\mu_{T_2}(\theta) - \mu_{T_2}(\theta_0)} \right)^2 \quad (3.6)$$

whenever the above limit exists; note that we allow for the possibility that $\text{ARE}_{T_1, T_2}(\theta_0) = \infty$. When (3.5) holds and the limit in (3.6) exists, we say that $\text{ARE}_{T_1, T_2}(\theta_0)$ is the ARE between T_1 and T_2 of the test of the null hypothesis $\theta = \theta_0$ against the two-sided alternative $\theta \neq \theta_0$; if we wish to test the one-sided alternative $\theta > \theta_0$ (or $\theta < \theta_0$), \mathcal{V} in (3.5) may be taken to be a right (or left) neighborhood of θ_0 , and the limit in (3.5) should be the directional limit as $\theta \downarrow \theta_0$ (or $\theta \uparrow \theta_0$). Note that we differ from other authors in our treatment of the ARE, in that we do not explicitly assume anything about the differentiability of the asymptotic means μ_{T_j} ; the reader is referred to the papers [92, 52, 72, 128] for various generalizations of the condition (3.5) and expression (3.6).

We are interested in the ARE between various correlation coefficients, so we now assume that the parameter space Θ indexes some type of association between two r.v.'s. Particularly, for each $\theta \in \Theta$ we shall henceforth assume that there exist two r.v.'s X and Y that are \mathbb{P}_θ -measurable, and then let

$$F_\theta(x, y) := \mathbb{P}_\theta(X \leq x, Y \leq y), \quad G_\theta(x) := \mathbb{P}_\theta(X \leq x), \quad \text{and} \quad H_\theta(y) := \mathbb{P}_\theta(Y \leq y)$$

denote the d.f.'s of (X, Y) , X , and Y , respectively. As we want θ to be some measure of association between X and Y , we shall assume that X and Y have fixed marginal distributions; that is, there exist d.f.'s G and H such that

$$G_\theta = G \text{ and } H_\theta = H \text{ for all } \theta \in \Theta.$$

Further assume that F_θ is continuous on \mathbb{R}^2 for each $\theta \in \Theta$, and note that this implies G and H are also continuous on \mathbb{R} .

The first correlation statistic we consider is the Pearson statistic

$$R := R_n := \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

where $(X_1, Y_1), \dots, (X_n, Y_n)$ are independent and identically distributed to (X, Y) , $\bar{X} =$

$\frac{1}{n} \sum_{i=1}^n X_i$, and $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$. Let

$$\mu_R := \mu_R(\theta) := \mathbb{E}_\theta \left(\frac{X - \mu_X}{\sigma_X} \frac{Y - \mu_Y}{\sigma_Y} \right), \quad (3.7)$$

and

$$\sigma_R^2 := \sigma_R^2(\theta) := \text{Var}_\theta \left(\frac{X - \mu_X}{\sigma_X} \frac{Y - \mu_Y}{\sigma_Y} - \frac{\mu_R}{2} \left(\left(\frac{X - \mu_X}{\sigma_X} \right)^2 + \left(\frac{Y - \mu_Y}{\sigma_Y} \right)^2 \right) \right)^2, \quad (3.8)$$

where $\mu_X := \mathbb{E} X$, $\mu_Y := \mathbb{E} Y$, $\sigma_X^2 := \text{Var} X$, and $\sigma_Y^2 := \text{Var} Y$; here and elsewhere, we tacitly assume that $\sigma = (\sigma^2)^{1/2}$ for any nonnegative number σ^2 . Our recent result in [116, Theorem 4.17] implies that the CLT-condition (3.5) holds (with R in place of T_j) whenever: (i) μ_R and σ_R^2 are continuous on \mathcal{V} ; (ii) $\mu_R(\theta) - \mu_R(\theta_0)$ is nonzero on $\mathcal{V} \setminus \{\theta_0\}$; (iii) $\inf_{\theta \in \mathcal{V}} \sigma_R^2(\theta) > 0$; and (iv) $\mathbb{E}(X^6 + Y^6) < \infty$. Note that μ_R and σ_R^2 appear in [23, (27.8.1)] as the asymptotic, in a certain sense, mean and variance of R .

Remark 3.2.1. By an application of [116, Theorem 3.2], we may replace the restriction that $\mathbb{E}(X^6 + Y^6) < \infty$ with the weaker moment condition $\mathbb{E}(|X|^p + |Y|^p) < \infty$ for some $p \in (4, 6)$.

Remark 3.2.2. By [116, Remark 4.18], $\sigma_R^2 = 0$ if and only if the pair $(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y})$ lies almost surely (a.s.) on the two lines through the origin with slopes $\frac{1}{\mu_R}(1 \pm \sqrt{1 - \mu_R^2})$ (with these two lines interpreted to mean the two coordinate axes if $\mu_R = 0$). If $|\mu_R| = 1$, then $\sigma_R^2 = 0$ is easily verified; otherwise, when $|\mu_R| < 1$, we must ensure that the distributions indexed by $\theta \in \mathcal{V}$ do not include the pathological case $\sigma_R^2 = 0$.

Next let

$$T := T_n := \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{sgn}\{(X_i - X_j)(Y_i - Y_j)\}$$

denote Kendall's difference sign correlation, where $\text{sgn}\{x\} = x/|x|$ whenever $x \neq 0$ and $\text{sgn}\{0\} = 0$. The above expression clearly shows that T is an example of Hoeffding's U -statistics [50]; there, it is shown that $\sqrt{n}(T - \mu_T)/\sigma_T$ converges to normality, where

$$\mu_T := \mu_T(\theta) := 4 \mathbb{E}_\theta F_\theta(X, Y) - 1, \quad (3.9)$$

and

$$\sigma_T^2 := \sigma_T^2(\theta) := 16 \text{Var}_\theta \{2F_\theta(X, Y) - G(X) - H(Y)\}. \quad (3.10)$$

The last correlation statistic we consider is Spearman's rank correlation S ; S is simply the result of calculating Pearson's R after replacing pair (X_i, Y_i) with $(r(X_i), r(Y_i))$, where $r(X_i) = \sum_{j=1}^n \mathbf{1}\{X_j \leq X_i\}$, $r(Y_i) = \sum_{j=1}^n \mathbf{1}\{Y_j \leq Y_i\}$, and $\mathbf{1}\{\cdot\}$ denotes the indicator function. By our assumption that G and H are continuous, we note that $\{r(X_1), \dots, r(X_n)\} = \{r(Y_1), \dots, r(Y_n)\} = \{1, \dots, n\}$ a.s., so that

$$\begin{aligned} S := S_n &:= \frac{\sum_{i=1}^n (r(X_i) - \frac{n+1}{2})(r(Y_i) - \frac{n+1}{2})}{\sqrt{\sum_{i=1}^n (r(X_i) - \frac{n+1}{2})^2} \sqrt{\sum_{i=1}^n (r(Y_i) - \frac{n+1}{2})^2}} \\ &= \frac{1}{n^3} \sum_{i,j,k=1}^n 12 \left(\frac{n^2}{n^2 - 1} \mathbf{1}\{X_j \leq X_i, Y_k \leq Y_i\} - \frac{1}{4} \frac{n+1}{n-1} \right). \end{aligned}$$

We note that S as expressed above is clearly an example of a V -statistic, which can further

be expressed as a U -statistic (cf. [51, Section 5c]), so that $\sqrt{n}(S - \mu_S)/\sigma_S$ tends to normality, where

$$\mu_S := \mu_S(\theta) := 12 \mathbf{E}_\theta G(X)H(Y) - 3, \quad (3.11)$$

and

$$\begin{aligned} \sigma_S^2 := \sigma_S^2(\theta) := & 144 \mathbf{Var}_\theta \{ (1 - G(X))(1 - H(Y)) \\ & + \int F_\theta(X, v) dH(v) + \int F_\theta(u, Y) dG(u) \}; \end{aligned} \quad (3.12)$$

unless specified otherwise, integrals shall be assumed to be evaluated over the entire domain of the integrand (which is \mathbb{R} in the two integrals of (3.12)).

Convergence to normality of the statistics T and S is not enough to ensure the CLT condition (3.5). However, a Berry-Esseen bound for the general U -statistic, as in [19, Theorem 3.1], implies (3.5) holds for $U \in \{T, S\}$ in place of T_j whenever (i) μ_U and σ_U^2 are continuous on \mathcal{V} ; (ii) $\mu_U(\theta) - \mu_U(\theta_0)$ is nonzero on $\mathcal{V} \setminus \{\theta_0\}$; and (iii) $\inf_{\theta \in \mathcal{V}} \sigma_U^2 > 0$. Note that no moment restrictions are needed for the nonparametric statistics T and S ; μ_T and μ_S always exist (and are bounded by 1), as do σ_T and σ_S .

Remark 3.2.3. Let ψ_X and ψ_Y be strictly increasing functions on the real line. Then the distributions of T_n and S_n will not change if the X_i 's and Y_j 's are replaced by $\psi_X(X_i)$ and $\psi_Y(Y_j)$, respectively. Particularly, if $\psi_X = G$ and $\psi_Y = H$, we may assume without loss of generality (w.l.o.g.) that X and Y are uniformly distributed on the unit interval $[0, 1]$. For $(x, y) \in \mathbb{R}^2$, let C_θ be implicitly defined by $C_\theta(G(x), H(y)) = F_\theta(x, y)$; C_θ is then uniquely determined by F_θ (under our assumption that F_θ is continuous) and is called the copula associated with F_θ [138]. Thus, at least when considering T or S , we may assume that the model consists of a class of copulas. The reader is referred to [91] for a treatment of copulas and their role in bivariate modeling.

Remark 3.2.4. Assume w.l.o.g. that X and Y are uniformly distributed on the unit interval. From (3.10), it is easy to see that $\sigma_T^2 = 0$ if and only if $F_\theta(X, Y) - \frac{1}{2}(X + Y)$ is a.s. a constant, and similarly from (3.12) it follows that $\sigma_S^2 = 0$ if and only if $(1 - X)(1 - Y) + \int F_\theta(X, v) dv + \int F_\theta(u, Y) du$ is a.s. a constant. By Fréchet [36], $W \leq C \leq M$ for any copula C , where $W(x, y) = 0 \vee (x + y - 1)$ and $M(x, y) = x \wedge y$ for $(x, y) \in [0, 1]^2$. It is easy to verify that $|\mu_T| = |\mu_S| = 1$ and $\sigma_T^2 = \sigma_S^2 = 0$ whenever $F_\theta \in \{W, M\}$. Pinelis [106] further shows that if the support of (X, Y) has positive Lebesgue measure (a rather common property in most practical models), then $\sigma_T^2 > 0$ and $\sigma_S^2 > 0$.

3.3 On $\text{ARE}_{T,S} = 1$ for tests of independence

Arguably the most common use of the statistics T and S is in testing for the independence of two populations against an alternative of an increasing (or decreasing) monotonic relation between the two populations. When X and Y are independent, one has $\mu_T = \mu_S = 0$, $\sigma_T^2 = \frac{4}{9}$, $\sigma_S^2 = 1$, and $\text{Cov}(T_n, S_n) \sim \frac{2}{3n}$ (implying that the asymptotic correlation between T and S is 1), as easily verified from an application of the theory of U -statistics [50]. However, more is required in order to ensure that $\text{ARE}_{T,S} = 1$ in a test for independence, since the manner in which a model departs from independence (or any null hypothesis, for that matter) has significant influence on the ARE. In this section we investigate a certain “smoothness” condition which is sufficient for $\text{ARE}_{T,S} = 1$ when the null hypothesis is that

of independence between X and Y .

From hereon we shall assume, unless the contrary is explicitly stated, that 0 is a point interior to the interval Θ , and moreover that $F_0(x, y) = G(x)H(y)$ for all $x, y \in \mathbb{R}^2$; that is, P_0 corresponds to independence between X and Y . For any $\theta \in \Theta$, we define the *association function* $a_\theta: \mathbb{R}^2 \rightarrow [-\frac{1}{4}, \frac{1}{4}]$ by

$$a_\theta(x, y) := F_\theta(x, y) - F_0(x, y) = P_\theta(X \leq x, Y \leq y) - P(X \leq x)P(Y \leq y). \quad (3.13)$$

Let us say that a model is *orderly* whenever $\theta > 0$ implies $a_\theta \geq 0$ and $\theta < 0$ implies $a_\theta \leq 0$; a model will be called *quite orderly* if $a_\theta(x, y)$ is an increasing function of θ for all pairs $(x, y) \in \mathbb{R}^2$.

In the following theorem we provide a sufficient condition for $\text{ARE}_{T,S} = 1$ in the test of independence; this condition is also necessary in an orderly model.

Theorem 3.3.1. *Let $(F_\theta)_{\theta \in \Theta}$ be a family of continuous d.f.'s such that $F_0(x, y) = G(x)H(y)$ for all $(x, y) \in \mathbb{R}^2$. Assume that μ_T , μ_S , σ_T^2 , and σ_S^2 are continuous on some neighborhood \mathcal{V} of 0, and also that μ_T and μ_S are both nonzero on $\mathcal{V} \setminus \{0\}$. Then $\text{ARE}_{T,S}(0) = 1$ whenever*

$$\lim_{\theta \rightarrow 0} \frac{E_\theta a_\theta(X, Y)}{E_0 a_\theta(X, Y)} = 1. \quad (3.14)$$

Further, (3.14) is a necessary condition for $\text{ARE}_{T,S}(0) = 1$ whenever the family $(F_\theta)_{\theta \in \Theta}$ is orderly.

Proof. Upon recalling that $\sigma_T^2(0) = \frac{4}{9}$ and $\sigma_S^2(0) = 1$, the continuity of σ_T^2 and σ_S^2 implies $\inf_{\theta \in \mathcal{V}} \sigma_T^2 > 0$ and $\inf_{\theta \in \mathcal{V}} \sigma_S^2 > 0$ (upon choosing a smaller \mathcal{V} if necessary); then the definition of the ARE in (3.6) and the remarks of Section 3.2 show that

$$\text{ARE}_{T,S}(0) = \frac{1}{4/9} \lim_{\theta \rightarrow 0} \left(\frac{\mu_T(\theta)}{\mu_S(\theta)} \right)^2 = \lim_{\theta \rightarrow 0} \left(\frac{3\mu_T(\theta)}{2\mu_S(\theta)} \right)^2. \quad (3.15)$$

Let (X_0, Y_0) and (X_θ, Y_θ) denote two independent random vectors with d.f.'s F_0 and F_θ , respectively, and note that

$$\begin{aligned} \int F_\theta dF_0 &= P(X_\theta < X_0, Y_\theta < Y_0) \\ &= 1 - P(X_0 < X_\theta) - P(Y_0 < Y_\theta) + P(X_0 < X_\theta, Y_0 < Y_\theta) \\ &= \int F_0 dF_\theta, \end{aligned}$$

where we use the fact that $P(X_0 < X_\theta) = P(Y_0 < Y_\theta) = \frac{1}{2}$. Then, using the definitions (3.9) and (3.11) of μ_T and μ_S , for nonzero $\theta \in \mathcal{V}$ we have

$$\begin{aligned} \frac{3\mu_T(\theta)}{2\mu_S(\theta)} &= \frac{12(E_\theta F_\theta(X, Y) - \frac{1}{4})}{24(E_\theta F_0(X, Y) - \frac{1}{4})} = \frac{\int F_\theta dF_\theta - \int F_0 dF_0}{2 \int F_0 d(F_\theta - F_0)} \\ &= \frac{\int (F_\theta - F_0) dF_\theta + \int (F_\theta - F_0) dF_0}{2 \int (F_\theta - F_0) dF_0} = \frac{1}{2} \left(1 + \frac{E_\theta a_\theta(X, Y)}{E_0 a_\theta(X, Y)} \right). \end{aligned} \quad (3.16)$$

Then $\text{ARE}_{T,S}(0) = 1$ follows from (3.14), (3.15), and (3.16).

Conversely, $\text{ARE}_{T,S}(0) = 1$, (3.15), and (3.16) imply that the limit in (3.14) is

either 1 or -3 . The assumption that the family $(F_\theta)_{\theta \in \Theta}$ is orderly implies ratio $\mathbb{E}_\theta a_\theta(X, Y) / \mathbb{E}_0 a_\theta(X, Y)$ is nonnegative for all nonzero $\theta \in \mathcal{V}$, so that (3.14) follows. \square

We refer to (3.14) as a “smoothness” condition, and note that it is equivalent to

$$\lim_{\theta \rightarrow 0} \frac{\mathbb{E}_\theta a_\theta(X, Y) - \mathbb{E}_0 a_\theta(X, Y)}{\mathbb{E}_0 a_\theta(X, Y)} = \lim_{\theta \rightarrow 0} \frac{\int (F_\theta - F_0) d(F_\theta - F_0)}{\int (F_\theta - F_0) dF_0} = 0. \quad (3.17)$$

Thus, if the family (F_θ) of distributions is smooth enough, $F_\theta - F_0$ is on the order of θ ; this suggests that the numerator of the ratio in (3.17) is on the order of θ^2 , while the denominator is normally on the order of θ – so that the ratio is on the order of θ , and (3.17) holds. In particular, (3.14) will always hold when the dependence of F_θ on θ is linear. More precisely, (3.14) will hold if $F_\theta(x, y)$ is of the form $G(x)H(y) + \theta \Delta(x, y)$, where Δ is some function such that $\int F_0 d\Delta \neq 0$. From these remarks we immediately have

Corollary 3.3.2. *For any family of distributions (F_θ) in the Farlie model (3.1), $\text{ARE}_{T,S}(0) = 1$.*

The assumption that (F_θ) is orderly in order for (3.14) to be a necessary condition for $\text{ARE}_{T,S}(0) = 1$ cannot be discarded. Indeed, let X and Y be uniformly distributed on the unit interval $(0, 1)$, and suppose that their joint d.f. takes the form $F_\theta = F_0 + \theta \Delta_\theta$, where Δ_θ is some polynomial (to be specified shortly) in x , y , and θ . Since $\Delta_\theta(x, 1) = \Delta_\theta(1, y) = \Delta_\theta(x, 0) = \Delta_\theta(0, y) = 0$ for all $(x, y, \theta) \in [0, 1]^2 \times \Theta$, let us further say that $\Delta_\theta(x, y) = xy(1-x)(1-y)(p(\theta) + x + y)$ for some polynomial p . We see from (3.15) and (3.16) that $\text{ARE}_{T,S}(0) = 1$ if the limit in (3.14) is -3 (instead of 1); accordingly, set

$$\frac{\mathbb{E}_\theta a_\theta(X, Y)}{\mathbb{E}_0 a_\theta(X, Y)} = \frac{\mathbb{E}_\theta \Delta(X, Y)}{\mathbb{E}_0 \Delta(X, Y)} = 1 - \frac{\theta}{50(1 + p(\theta))} = -3,$$

so that $p(\theta) = \frac{\theta}{200} - 1$ and

$$F_\theta(x, y) = xy \left(1 - \theta(1-x-y)(1-x)(1-y) + \frac{\theta^2}{200}(1-x)(1-y) \right).$$

It is routine to verify that F_θ is indeed a d.f. when $\theta \in \Theta = [-1, 1]$, with

$$\mu_T(\theta) = -\frac{\theta^2}{900}, \quad \text{and} \quad \mu_S(\theta) = \frac{\theta^2}{600} = -\frac{3}{2} \mu_T(\theta).$$

Thus, we have $\text{ARE}_{T,S}(0) = 1$ while the condition (3.14) fails to hold; of course, this model is not orderly, as evidenced by $\mu_T < 0$ while $\mu_S > 0$ for all nonzero $\theta \in \Theta$.

In models which are “nearly linear”, the denominator of (3.17) will typically be on the order of θ , so that Theorem 3.3.1 implies $\text{ARE}_{T,S}(0) = 1$ whenever the numerator in (3.17) is $o(\theta)$. The following propositions provide sufficient conditions that ensure the numerator in (3.14) is $o(\theta)$ (or $o(\mathbb{E}_0 a_\theta(X, Y))$ in Proposition 3.3.3), which should then provide sufficient conditions for $\text{ARE}_{T,S}(0) = 1$.

Proposition 3.3.3. *Let*

$$d(\theta) = \|F_\theta - F_0\| = \sup_{A \in \mathfrak{R}^2} |\mathbb{P}_\theta(A) - \mathbb{P}_0(A)| \quad \text{and} \quad \rho(\theta) = \sup_{(x, y) \in \mathbb{R}^2} |F_\theta(x, y) - F_0(x, y)|.$$

Then $\text{ARE}_{T,S}(0) = 1$ whenever $d(\theta)\rho(\theta) = o(\mathbb{E}_0 a_\theta(X, Y))$ as $\theta \rightarrow 0$.

Proof. One has

$$\left| \frac{\int (F_\theta - F_0) d(F_\theta - F_0)}{\int (F_\theta - F_0) dF_0} \right| \leq \frac{\rho(\theta) \int |d(F_\theta - F_0)|}{|\mathbb{E}_0 a_\theta(X, Y)|} = \frac{2d(\theta)\rho(\theta)}{|\mathbb{E}_0 a_\theta(X, Y)|} \xrightarrow{\theta \rightarrow 0} 0,$$

by the hypotheses of the proposition. Then $\text{ARE}_{T,S}(0) = 1$ follows by Theorem 3.3.1. \square

Corollary 3.3.4. *If $d(\theta)^2 = \|F_\theta - F_0\|^2 = o(\mathbb{E}_0 a_\theta(X, Y))$, then $\text{ARE}_{T,S}(0) = 1$.*

Proof. Apply Proposition 3.3.3 after noting that $\rho(\theta) \leq d(\theta)$. \square

Note that $d(\theta)^2 = o(\mathbb{E}_0 a_\theta(X, Y))$ implies convergence in variation of F_θ to F_0 as $\theta \rightarrow 0$. The assumption that $F_\theta \rightarrow F_0$ in variation is a natural one and will be satisfied in most hypothesis testing problems; informally, the convergence in variation means that the competing hypotheses are not too easy to distinguish. We will use Proposition 3.3.3 in the next section to demonstrate that $\text{ARE}_{T,S}(0) = 1$ in two of the MICD models.

Remark 3.3.5. By Scheffé [129], in order for $F_\theta \rightarrow F_0$ in variation, it is enough that $f_\theta(x, y) \rightarrow f_0(x, y)$ for μ -almost all $(x, y) \in \mathbb{R}^2$, where μ is any nonnegative Borel measure on \mathbb{R}^2 and F_θ is absolutely continuous with respect to μ with density f_θ for all $\theta \in \mathcal{V}$. Moreover, this convergence is uniform and implies the uniform weak convergence $F_\theta \Rightarrow F_0$.

Proposition 3.3.6. *If, for some $M < \infty$ and some finite nonnegative Borel measure μ such that for all $\theta \in \mathcal{V}$ there exists a density f_θ of F_θ with respect to μ , one has that*

$$\sup_{\theta \in \mathcal{V} \setminus \{0\}} \left| \frac{1}{\theta} (f_\theta(x, y) - f_0(x, y)) \right| \leq M \quad \mu\text{-almost everywhere},$$

then $\int (F_\theta - F_0) d(F_\theta - F_0) = o(\theta)$ as $\theta \rightarrow 0$.

Proof. Given the hypotheses, $f_\theta \rightarrow f_0$ μ -almost everywhere, so that $F_\theta \rightarrow F_0$ in variation (cf. Remark 3.3.5); particularly, since F_0 is continuous over \mathbb{R}^2 , we have $F_\theta \rightarrow F_0$ pointwise (uniformly, moreover). Then,

$$\left| \frac{1}{\theta} \int (F_\theta - F_0) d(F_\theta - F_0) \right| = \left| \int (F_\theta - F_0) \frac{f_\theta - f_0}{\theta} d\mu \right| \leq M \int |F_\theta - F_0| d\mu \xrightarrow{\theta \rightarrow 0} 0,$$

with the last step following by dominated convergence. \square

Remark 3.3.7. If $F_\theta \xrightarrow{\theta \rightarrow 0} F_0$ weakly, and F_0 is continuous, then $F_\theta \xrightarrow{\theta \rightarrow 0} F_0$ uniformly on \mathbb{R}^2 .

Proposition 3.3.8. *If $F_\theta(x, y)$ is differentiable in θ at $\theta = 0$ uniformly over $(x, y) \in \mathbb{R}^2$ and $\frac{\partial}{\partial \theta} F_\theta|_{\theta=0}$ is continuous and bounded, then $\int (F_\theta - F_0) d(F_\theta - F_0) = o(\theta)$ as $\theta \rightarrow 0$.*

Proof. By uniformly differentiable over $(x, y) \in \mathbb{R}^2$, we mean exactly that

$$s(\theta) := \sup_{(x, y) \in \mathbb{R}^2} \left| \frac{F_\theta(x, y) - F_0(x, y)}{\theta} - \frac{\partial F_\theta(x, y)}{\partial \theta} \Big|_{\theta=0} \right| \xrightarrow{\theta \rightarrow 0} 0.$$

Then

$$\left| \frac{1}{\theta} \int (F_\theta - F_0) d(F_\theta - F_0) \right| \leq \left| \int \left(\frac{\partial F_\theta}{\partial \theta} \Big|_{\theta=0} \right) d(F_\theta - F_0) \right| + s(\theta) \int |d(F_\theta - F_0)| \xrightarrow{\theta \rightarrow 0} 0,$$

since $s(\theta) \rightarrow 0$ implies $F_\theta(x, y) \rightarrow F_0(x, y)$ for all $(x, y) \in \mathbb{R}^2$, whence $F_\theta \Rightarrow F_0$ weakly. \square

Corollary 3.3.9. *If $\frac{\partial}{\partial \theta} F_\theta$ and $\frac{\partial^2}{\partial \theta^2} F_\theta$ are uniformly bounded over $\theta \in \mathcal{V}$ and $(x, y) \in \mathbb{R}^2$, then $\int (F_\theta - F_0) d(F_\theta - F_0) = o(\theta)$ as $\theta \rightarrow 0$.*

Proposition 3.3.8 is easily applied to Plackett's (3.2) and Frank's (3.3) models in order to show that $\text{ARE}_{T,S}(0) = 1$:

Proposition 3.3.10. *For F_θ in the Plackett (3.2) or Frank (3.3) models, $\text{ARE}_{T,S}(0) = 1$.*

Proof. For simplicity of notation, assume in what follows that $F_\theta := F_\theta(x, y)$, $G := G(x)$, $H := H(y)$, and $F'_\theta := \frac{\partial}{\partial \theta} F_\theta(x, y)$.

Suppose F_θ lies in Plackett's model (3.2). Then

$$\frac{\text{E}_0 a_\theta(X, Y)}{\theta} = \frac{e^\theta - 1}{\theta} \text{E}_0(G - F_\theta)(H - F_\theta) \xrightarrow{\theta \rightarrow 0} \text{E}_0(G - F_0)(H - F_0) = \frac{1}{36}$$

by dominated convergence. Further, $F'_\theta|_{\theta=0} = (G - F_0)(H - F_0)$, whence

$$\begin{aligned} \left| \frac{F_\theta - F_0}{\theta} - F'_\theta|_{\theta=0} \right| &= \left| \frac{e^\theta - 1}{\theta} (G - F_\theta)(H - F_\theta) - (G - F_0)(H - F_0) \right| \\ &= \left| \frac{e^\theta - 1 - \theta}{\theta} (G - F_0)(H - F_0) + \frac{e^\theta - 1}{\theta} (F_0 - F_\theta)(G + H - F_0 - F_\theta) \right| \\ &\leq \left| \frac{e^\theta - 1 - \theta}{\theta} \right| + 2 \frac{e^\theta - 1}{\theta} |F_\theta - F_0|. \end{aligned}$$

Since $F_\theta \rightarrow F_0$ for all $(x, y) \in \mathbb{R}^2$, it does so uniformly (cf. Remark 3.3.7); that is, F_θ is differentiable uniformly over $(x, y) \in \mathbb{R}^2$ at $\theta = 0$ (and $F'_\theta|_{\theta=0}$ is clearly continuous and bounded), so that Proposition 3.3.8 implies $\int (F_\theta - F_0) d(F_\theta - F_0) = o(\theta)$ as $\theta \rightarrow 0$. Then (3.17) holds, and $\text{ARE}_{T,S}(0) = 1$ follows from Theorem 3.3.1.

In Frank's class of models (3.3), a Taylor expansion of $F_\theta = -\frac{1}{\theta} \ln(1 + (e^{-\theta G} - 1)(e^{-\theta H} - 1)/(e^{-\theta} - 1))$ about $\theta = 0$ yields

$$F_\theta = GH + \frac{1}{2} GH(1 - G)(1 - H)\theta + \frac{1}{12} GH(1 - 3G + 2G^2)(1 - 3H + 2H^2)\theta^2 + O(\theta^3).$$

Then $\frac{\partial}{\partial \theta} F_\theta$ and $\frac{\partial^2}{\partial \theta^2} F_\theta$ are uniformly bounded near the origin, so that Corollary 3.3.9 implies $\int (F_\theta - F_0) d(F_\theta - F_0) = o(\theta)$. The above Taylor expansion of F_θ further shows us that $\text{E}_0 a_\theta(X, Y)/\theta \rightarrow \frac{1}{2} \text{E}_0 G(X)H(Y)(1 - G(X))(1 - H(Y)) = 1/72$, so that (3.17) holds and $\text{ARE}_{T,S}(0) = 1$ follows from Theorem 3.3.1. \square

3.4 The ARE under some specific models

In this section we consider three general classes of bivariate models, investigating various properties of the ARE in each. We first introduce four novel *Mixtures of Independence* and *Complete Dependence* (MICD) models. The MICD models have a simple construction

in that their supports consist of unions of square planar regions and linear segments; of particular note is the existence of two models where $\text{ARE}_{T,S}(0) \neq 1$. We next consider two Durbin-Stuart models, named for their construction based on the Durbin-Stuart inequality [30] relating μ_S and μ_T ; these models are pathological in the sense that each of their distributions is singular with respect to Lebesgue measure in the plane. Lastly, we consider several classes of models where the dependence structure between X and Y is obtained from the bivariate normal distribution. It is known that R is more efficient than either T or S in the bivariate normal model (cf. [82]), though if the normality assumption fails to hold (say, due to heavy tails) then the distribution of R can be affected quite drastically (see e.g. [74] and the references there). We examine a few transformations of the bivariate normal distribution which demonstrate that R can be much less efficient than T (or S) when the marginal distributions of X and Y vary.

3.4.1 Mixtures of independence and complete dependence (MICD)

Each of the MICD models is a mixture of the uniform distribution in a plane region and the uniform distribution on a part of the line $y = x$ (or $y = -x$); X and Y are independent if and only if $\theta = 0$, and $Y = \pm X$ a.s. corresponds to $\theta = \pm 1$ (i.e., the two Fréchet-Hoeffding bounds are at the boundaries of Θ). All of the distributions in these models have support contained in the square $\Omega = [-\frac{1}{2}, \frac{1}{2}]^2$ and are constructed to have uniform marginal distributions; the d.f. of $(X + \frac{1}{2}, Y + \frac{1}{2})$ is then a copula, with this shift in location being performed solely for aesthetic and interpretive purposes. Each model m is described by a “region of independence” $A_{\theta,m}$ and a “region of complete dependence” $B_{\theta,m} := \{(x, \text{sgn}(\theta)x) : x \in [-\frac{1}{2}, \frac{1}{2}]\} \setminus A_{\theta,m}$. Further, the $A_{\theta,m}$ (and hence the $B_{\theta,m}$) are defined by a relatively simple heuristic: X and Y should be independent given that $|X|$ and/or $|Y|$ are either “small” or “large,” whence the model labels AS, AL, OS, and OL (with the symbols A, O, S, and L representing “and”, “or”, “small”, and “large”, respectively). Table 3.1 defines the $A_{\theta,m}$ and Figure 3.1 illustrates the support of these four models for choices of $\theta = 0.4$; for instance, in the OS model we see that X and Y are conditionally independent given the event $\{|X| \leq \frac{1-|\theta|}{2}\} \cup \{|Y| \leq \frac{1-|\theta|}{2}\}$.

Table 3.1: Description of the four MICD models

Model m	X and Y are independent when ...	$A_{\theta,m}$	π_m
AS	$ X $ <u>and</u> $ Y $ are <u>small</u>	$[-\frac{1- \theta }{2}, \frac{1- \theta }{2}]^2$	$ \theta $
AL	$ X $ <u>and</u> $ Y $ are <u>large</u>	$([-\frac{1}{2}, -\frac{ \theta }{2}] \cup [\frac{ \theta }{2}, \frac{1}{2}])^2$	$ \theta $
OS	$ X $ <u>or</u> $ Y $ is <u>small</u>	$\Omega \setminus ([-\frac{1}{2}, -\frac{1- \theta }{2}] \cup [\frac{1- \theta }{2}, \frac{1}{2}])^2$	θ^2
OL	$ X $ <u>or</u> $ Y $ is <u>large</u>	$\Omega \setminus [-\frac{ \theta }{2}, \frac{ \theta }{2}]^2$	θ^2

Let $F_{A_{\theta,m}}$ denote the d.f. of a random vector uniformly distributed over the region $A_{\theta,m}$, and let $F_{B_{\theta,m}}$ denote the d.f. of a random vector uniformly distributed along the line segment(s) $B_{\theta,m}$. The d.f. of (X, Y) is then

$$F_{\theta,m} := (1 - \pi_m)F_{A_{\theta,m}} + \pi_m F_{B_{\theta,m}}, \quad (3.18)$$

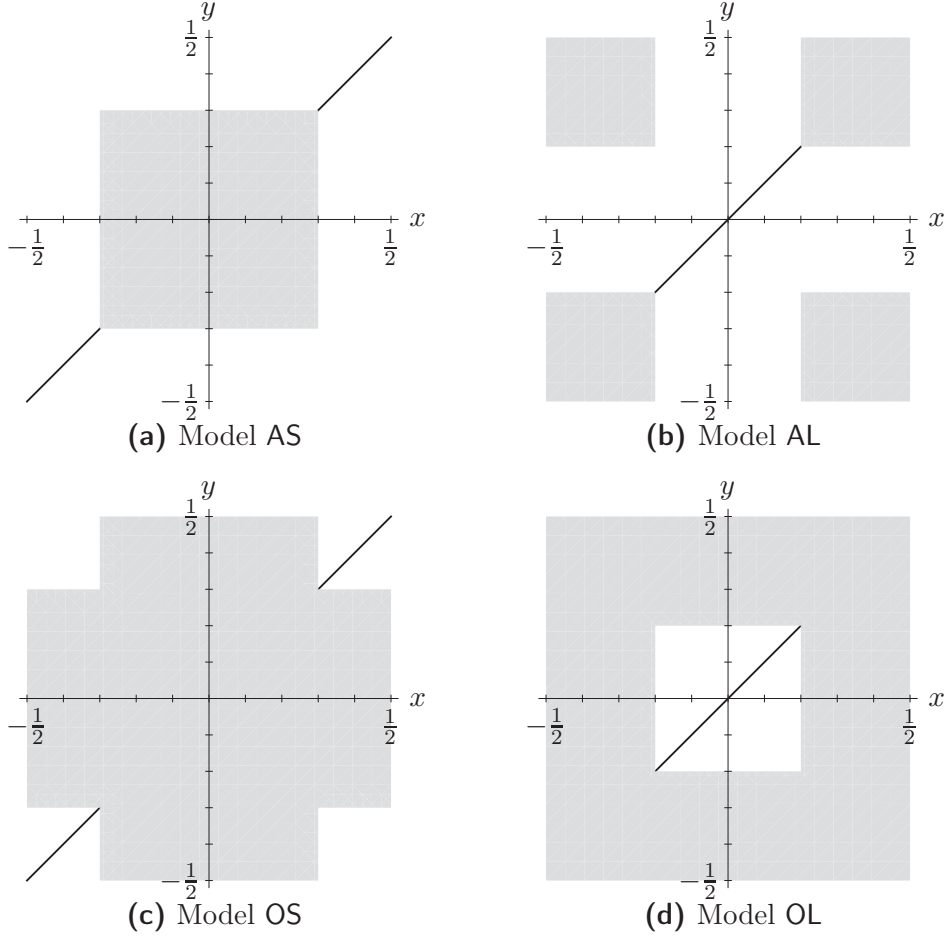


Figure 3.1: Supports of the MICD models for $\theta = 0.4$; note that reflection of each figure across the y -axis gives the support for $\theta = -0.4$

where $\pi_m = |\theta|$ for models $m \in \{\text{AS}, \text{AL}\}$ and $\pi_m = \theta^2$ for models $m \in \{\text{OS}, \text{OL}\}$ are chosen so as to ensure the marginal distributions of X and Y are uniform.

Given this characterization of the MICD models, the asymptotic means and variances of the three statistics R , S , and T are rather straightforward to calculate using the formulas (3.7)–(3.12); expressions for these asymptotic moments are tabulated in Table 3.2 of the appendix, and plots of the ARE's are given in Figure 3.2. Note that we use here (and subsequently hereafter) a superscript on the ARE to distinguish between the various models. By the construction of the models, the asymptotic means are odd functions of θ and the asymptotic variances are even functions, so that the ARE's are also even. As such, the moments in Table 3.2 are given only for $\theta \geq 0$, with the understanding that the reader may generalize the formulas there for arbitrary $\theta \in [-1, 1]$.

Table 3.2: Asymptotic moments of R , S , and T in the MICD models for $\theta \geq 0$; note that $\mu_R = \mu_S$

Model AS

$$\begin{aligned}\mu_T &= \theta(2 - \theta) & \sigma_T^2 &= \frac{4}{9}(1 - \theta)^3(1 + 9\theta) \\ \mu_S &= \theta(3 - 3\theta + \theta^2) & \sigma_S^2 &= (1 - \theta)^5(1 + 9\theta) \\ & & \sigma_R^2 &= \frac{1}{5}(1 - \theta)^5(5 + 45\theta - 72\theta^2 + 54\theta^3 - 30\theta^4 + 12\theta^5 - 2\theta^6)\end{aligned}$$

Model AL

$$\begin{aligned}\mu_T &= \theta^2 & \sigma_T^2 &= \frac{2}{9}(2 + 3\theta + 6\theta^2 + 7\theta^3 - 18\theta^4) \\ \mu_S &= \theta^3 & \sigma_S^2 &= 1 + \theta + \theta^2 - \theta^3 - \theta^4 + 8\theta^5 - 9\theta^6 \\ & & \sigma_R^2 &= \frac{1}{10}(10 + 10\theta + 10\theta^2 - 10\theta^3 - 10\theta^4 + 8\theta^5 + 14\theta^6 + 5\theta^7 - 31\theta^8 \\ & & & \quad - 5\theta^9 - 5\theta^{10} + 4\theta^{11})\end{aligned}$$

Model OS

$$\begin{aligned}\mu_T &= \theta^2(2 - 2\theta + \theta^2) & \sigma_T^2 &= \frac{4}{9}(1 - 12\theta^4 + 42\theta^5 - 58\theta^6 + 36\theta^7 - 9\theta^8) \\ \mu_S &= \theta^2(3 - 3\theta + \theta^2) & \sigma_S^2 &= \frac{1}{10}(10 - 240\theta^4 + 630\theta^5 - 667\theta^6 + 333\theta^7 - 66\theta^8) \\ & & \sigma_R^2 &= \frac{1}{10}(10 - 384\theta^4 + 1338\theta^5 - 2122\theta^6 + 1896\theta^7 - 859\theta^8 - 117\theta^9 \\ & & & \quad + 495\theta^{10} - 387\theta^{11} + 165\theta^{12} - 39\theta^{13} + 4\theta^{14})\end{aligned}$$

Model OL

$$\begin{aligned}\mu_T &= \theta^4 & \sigma_T^2 &= \frac{4}{9}(1 + 8\theta^6 - 9\theta^8) \\ \mu_S &= \theta^4 & \sigma_S^2 &= \frac{1}{5}(5 + 4\theta^6 + 24\theta^7 - 33\theta^8) \\ & & \sigma_R^2 &= \frac{1}{5}(5 + 4\theta^6 + 7\theta^8 - 18\theta^{10} + 2\theta^{14})\end{aligned}$$

3.4.1.1 On $\text{ARE}_{T,S}(0) = 1$ in the MICD models

The plots in Figure 3.2 show that $\text{ARE}_{T,S}(0) = 1$ in the two “small” models, whereas $\text{ARE}_{T,S}^{(\text{OL})}(0) = \frac{9}{4}$ and $\text{ARE}_{T,S}^{(\text{AL})}(0) = \infty$. We examine this phenomenon in regards to Theorem 3.3.1 and Proposition 3.3.3.

For any of the four MICD models, consider this:

$$\mathbb{E}_0 a_\theta(X, Y) = \mathbb{P}(V_\theta - V_0 > (0, 0)) - \frac{1}{4},$$

where $V_\theta := (X_\theta, Y_\theta)$ and $V_0 := (X_0, Y_0)$ are independent, with d.f.’s F_θ and F_0 , respectively, and $V_\theta - V_0 > (0, 0)$ means that $X_\theta > X_0$ and $Y_\theta > Y_0$. Note that the distribution of V_θ is invariant with respect to the group of rotations, say \mathcal{G} , through integer multiples of $\frac{\pi}{2}$, so that the order of \mathcal{G} is 4. It follows that $\mathbb{P}(V_\theta - V_0 > (0, 0) | V_\theta \in A_\theta) = \frac{1}{4}$, where $A_\theta := A_{\theta,m}$

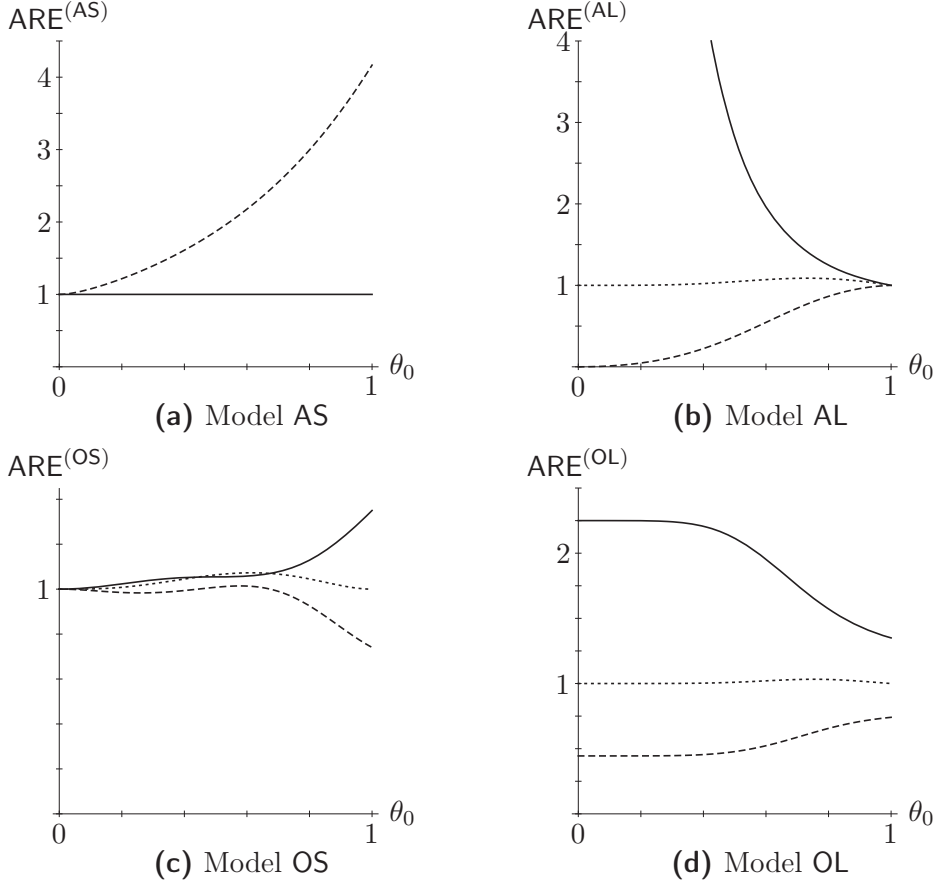


Figure 3.2: $\text{ARE}_{T,S}$ (solid), $\text{ARE}_{R,T}$ (dashed), and $\text{ARE}_{R,S}$ (dotted) for the MICD models

is the “independence” region. Indeed,

$$\begin{aligned}
 16 \mathbb{P} (V_\theta - V_0 > (0, 0), V_\theta \in A_\theta) &= \sum_{R_\theta \in \mathcal{G}} \sum_{R_0 \in \mathcal{G}} \mathbb{P} (R_\theta V_\theta - R_0 V_0 > (0, 0), R_\theta V_\theta \in A_\theta) \\
 &= \sum_{R_\theta \in \mathcal{G}} \sum_{R_0 \in \mathcal{G}} \mathbb{P} (R_\theta V_\theta - R_0 V_0 > (0, 0), V_\theta \in A_\theta) \\
 &= \mathbb{E} \mathbb{I} \{V_\theta \in A_\theta\} \sum_{R_0 \in \mathcal{G}} \sum_{R_\theta \in \mathcal{G}} \mathbb{I} \{R_\theta (V_\theta - R_0 V_0) > (0, 0)\} \\
 &= \mathbb{E} \mathbb{I} \{V_\theta \in A_\theta\} \sum_{R_0 \in \mathcal{G}} 1 = 4 \mathbb{P} (V_\theta \in A_\theta).
 \end{aligned}$$

On the other hand, it is clear that for $\theta \downarrow 0$ the conditional probability $\mathbb{P} (V_\theta - V_0 > (0, 0) | V_\theta \notin A_\theta)$ tends to $\frac{1}{2}$ in the “small” models (AS and OS), and it tends to $\frac{1}{4}$ in the “large” models (AL and OL). Working a bit harder, one can see that in fact $\mathbb{P} (V_\theta - V_0 >$

$(0, 0) | V_\theta \notin A_\theta) = \frac{1}{4} + O(\theta^2)$ in the “large” models; this follows because

$$\begin{aligned} & \mathbf{P}((x, x) - V_0 > (0, 0)) + \mathbf{P}(-(x, x) - V_0 > (0, 0)) \\ &= (\tfrac{1}{2} + x)^2 + (\tfrac{1}{2} - x)^2 = \tfrac{1}{2} + 2x^2 = \tfrac{1}{2} + O(\theta^2) \end{aligned}$$

over all $x \in [0, \theta/2]$ (constituting half of the region $B_{\theta,m}$). So, for $\theta \downarrow 0$,

$$\mathbf{E}_0 a_\theta(X, Y) \sim (\tfrac{1}{2} - \tfrac{1}{4}) \mathbf{P}(V_\theta \notin A_{\theta,m}) = \tfrac{1}{4} \pi_m = \begin{cases} \theta/4 & \text{if } m = \text{AS}, \\ \theta^2/4 & \text{if } m = \text{OS}, \end{cases}$$

and

$$\mathbf{E}_0 a_\theta(X, Y) \sim (\tfrac{1}{4} + O(\theta^2) - \tfrac{1}{4}) \mathbf{P}(V_\theta \notin A_{\theta,m}) = O(\pi_m \theta^2) = \begin{cases} O(\theta^3) & \text{if } m = \text{AL}, \\ O(\theta^4) & \text{if } m = \text{OL}. \end{cases}$$

This explains the order of magnitude of $\mu_S = 12 \mathbf{E}_0 a_\theta(X, Y)$, quite in concordance with the exact expressions given in Table 3.2.

It is also instructive to compute the variation distance $d(\theta) = \|F_\theta - F_0\|$. Let $P_\theta(A) := \mathbf{P}_\theta((X, Y) \in A)$ and $\lambda_\theta := P_\theta - P_0$. Then $d(\theta) = \|\lambda_\theta\| = \lambda_\theta(C_\theta)$, where C_θ is a Borel subset of \mathbb{R}^2 such that $\lambda_\theta(E) \geq 0$ for all $E \subseteq C_\theta$ and $\lambda_\theta(E) \leq 0$ for all $E \subseteq \mathbb{R}^2 \setminus C_\theta$. In all four MICD models, it is easy to see that C_θ is the support of \mathbf{P}_θ , i.e. $C_\theta = A_{\theta,m} \cup B_{\theta,m}$; thus, $d(\theta)$ equals θ^2 for each of the two “or” models and $\theta(2 - \theta)$ for each of the “and” models.

Moreover, it is fairly straightforward to compute the Kolmogorov distance $\rho(\theta) = \sup |F_\theta - F_0|$. In the two “large” models, one finds that $\rho(\theta) = |F_\theta(0, 0) - F_0(0, 0)|$; that is, $\rho(\theta)$ is $\frac{\theta}{2}$ in the AL model and is $\frac{\theta^2}{4}$ in the OL model. In the two “small” models, $|F_\theta - F_0|$ attains its maximum at the “corners” of $A_{\theta,m}$; namely, letting $a = \frac{1-\theta}{2}$, $\rho(\theta) = |F_\theta(a, a) - F_0(a, a)| = |F_\theta(-a, -a) - F_0(-a, -a)|$ for $\theta > 0$, and $\rho(\theta) = |F_\theta(-a, a) - F_0(-a, a)| = |F_\theta(a, -a) - F_0(a, -a)|$ for $\theta < 0$. Thus, $\rho(\theta) = \frac{\theta}{4}(2 - \theta)$ in the AS model and $\rho(\theta) = \frac{\theta^2}{4}$ in the OS model. Note that $\rho(\theta)$ is of the same order as $d(\theta)$ in all four of the MICD models.

Thus, $d(\theta)\rho(\theta)$ is on the order of $O(\theta^2)$ in the two “and” models, and is on the order of $O(\theta^4)$ in the two “or” models. Taking into account the previously described asymptotics of $\mathbf{E}_0 a_\theta(X, Y)$, our heuristics show that the ratio in the limit of (3.17) is on the order of $O(\theta)$, $O(1/\theta)$, $O(\theta)$, and $O(1)$ for the models AS, AL, OS, and OL, respectively, which explains why in the four models $\text{ARE}_{T,S}^{(\text{AS})}(0) = \text{ARE}_{T,S}^{(\text{OS})}(0) = 1$, $\text{ARE}_{T,S}^{(\text{AL})}(0) = \infty$, and $\text{ARE}_{T,S}^{(\text{OL})}(0) \neq 1$.

3.4.1.2 ARE’s between R , S , and T in the MICD models

Some general remarks on the plots in Figure 3.2 are in order. First, $\text{ARE}_{R,S}(0) = 1$ in all four of the MICD models. This is due to the fact that $\mu_R = \mu_S$ (recall the d.f.’s in these models are copulas with a change of location, so that the product-moment correlation coincides with the grade correlation) and $\sigma_R^2(0) = \sigma_S^2(0) = 1$. Indeed, $\text{ARE}_{R,S}$ is determined solely by σ_R^2 and σ_S^2 in these models, with $\text{ARE}_{R,S} = \sigma_S^2/\sigma_R^2$.

Note that $\text{ARE}_{R,S}^{(\text{AS})}$ is strictly increasing on $(0, 1)$ to $\text{ARE}_{R,S}(1-) = 4.166\dots$. This is in contrast to the other three models, where $\text{ARE}_{R,S}$ is increasing then decreasing on $(0, 1)$,

with relatively little variation; particularly, $\text{ARE}_{R,S}(1-) = 1$ in all MICD models except AS, $\max \text{ARE}_{R,S}^{(\text{AL})} = 1.087\dots$, $\max \text{ARE}_{R,S}^{(\text{OS})} = 1.072\dots$, and $\max \text{ARE}_{R,S}^{(\text{OL})} = 1.032\dots$.

The plots of $\text{ARE}_{T,S}$ possess even more interesting features. We see that $\text{ARE}_{T,S}^{(\text{AS})} = 1$, not only for the test of independence ($\theta_0 = 0$), but for any null hypothesis with $\theta_0 \in (-1, 1)$. As already noted, $\text{ARE}_{T,S}(0) = 1$ in OS; we also see that $\text{ARE}_{T,S}$ is strictly increasing on $(0, 1)$. In the other two models (AL and OL), $\text{ARE}_{T,S}(0) > 1$ and $\text{ARE}_{T,S}$ is strictly decreasing on $(0, 1)$. In OL, we have $\mu_T = \mu_S = \text{sgn}(\theta)\theta^4$, so that $\text{ARE}_{T,S} = \sigma_S^2/\sigma_T^2$; in AL, $\text{ARE}_{T,S}(0) = \infty$ (and $\text{ARE}_{T,S}(1-) = 1$).

3.4.1.3 The Neyman-Pearson test for the MICD AS model

We provide here a short derivation of the Neyman-Pearson test of the null hypothesis $H_0: \theta = \theta_0$ against the alternative $H_1: \theta = \theta_1$, $0 \leq \theta_0 < \theta_1 < 1$, for the MICD model AS; similar methods may of course be used to derive the Neyman-Pearson test for the other three MICD models. The test statistic is the likelihood ratio $\Lambda := \prod_{i=1}^n f_{\theta_1}(X_i, Y_i)/f_{\theta_0}(X_i, Y_i)$, where

$$f_{\theta}(x, y) = \frac{1}{1-\theta} \mathbf{I}\{(x, y) \in A_{\theta}\} + \mathbf{I}\{(x, y) \in B_{\theta}\}$$

for $\theta \in [0, 1]$ and $(x, y) \in \Omega = [-\frac{1}{2}, \frac{1}{2}]^2$,

$$A_{\theta} = \left\{-\frac{1-\theta}{2}, \frac{1-\theta}{2}\right\}^2, \quad \text{and} \quad B_{\theta} = \{(x, x): |x| \in \left\{\frac{1-\theta}{2}, \frac{1}{2}\right\}\}.$$

Under the null hypothesis, $\Lambda \in \{0, 1, \lambda, \lambda^2, \dots, \lambda^n\}$ a.s., where $\lambda := \frac{1-\theta_0}{1-\theta_1} > 1$, with $\Lambda = 0$ occurring on the event $\cup_{i=1}^n \{(X_i, Y_i) \in A_{\theta_0} \setminus A_{\theta_1}\}$ (an “impossible” event under the alternative hypothesis) and $\Lambda = \lambda^k$ corresponding to the event of exactly k observations lying in $A_{\theta_0} \cap A_{\theta_1} = A_{\theta_1}$ and the remaining $n - k$ observations lying in $B_{\theta_0} \cap B_{\theta_1} = B_{\theta_0}$. One has $\mathbf{P}_{\theta_0}(\Lambda = 0) = 1 - (\theta_0 + \frac{(1-\theta_1)^2}{1-\theta_0})^n$ and $\mathbf{P}_{\theta_0}(\Lambda = \lambda^k) = \binom{n}{k} (\frac{(1-\theta_1)^2}{1-\theta_0})^k \theta_0^{n-k}$ for $k = 0, \dots, n$. The null hypothesis is rejected whenever $\Lambda > \gamma$, where $\gamma \geq 0$ is some prescribed “critical value”; that is, H_0 is rejected when there are too many observations inside the smaller square A_{θ_1} (and hence too many observations along the line segments B_{θ_0}). By the Neyman-Pearson lemma, this test is a most powerful size α test, where $\alpha = \mathbf{P}_{\theta_0}(\Lambda > \gamma)$.

3.4.2 Durbin-Stuart models

When Kendall’s rank correlation statistic was first introduced, relations between it and Spearman’s S were studied extensively. One important relation, the Durbin-Stuart inequality [30], states that

$$\frac{3}{2} \frac{n}{n+1} T_n - \frac{1}{2} \frac{n-2}{n+1} \leq S_n \leq \frac{1}{2} \frac{n-1}{n+1} + T_n - \frac{1}{2} \frac{n-1}{n+1} T_n^2 \quad (3.19)$$

a.s. on the event $\{T_n \geq 0\}$. The lower bound in (3.19) is attained precisely when, upon sorting the ranks of X , the ranks of Y are a cyclical permutation of the natural ordering $(1, 2, \dots, n)$. The upper bound in (3.19) can be attained for some values of n and T_n , particularly when the ranks of Y are partitioned into “compact sets” (cf. [30] for a definition

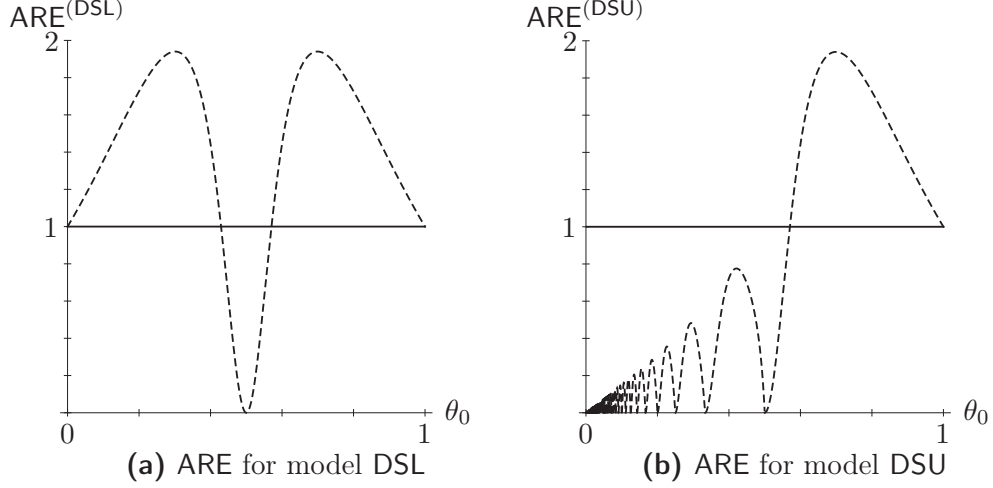


Figure 3.3: $\text{ARE}_{T,S}$ (solid) and $\text{ARE}_{R,T}$ (dashed) for Durbin-Stuart models

of this term). Allowing $n \rightarrow \infty$ in (3.19) yields the inequalities

$$\frac{3}{2}\mu_T - \frac{1}{2} \leq \mu_S \leq \frac{1}{2} + \mu_T - \frac{1}{2}\mu_T^2 \quad (3.20)$$

when $\mu_T \geq 0$. Using the criterion for which the two bounds in (3.19) are attained, we construct two Durbin-Stuart models, labeled DSL and DSU, in which the bounds in (3.20) are attained; for these two models, we depart from the convention that $\theta = 0$ corresponds to independence between X and Y .

Model DSL is characterized by letting $Y = X \ominus \theta$ a.s., where X is uniformly distributed on $[0, 1]$, $\theta \in \Theta := [0, 1]$, and \ominus denotes the subtraction modulo 1 operator. It is straightforward to verify that

$$\begin{aligned} \mu_T(\theta) &= (1 - 2\theta)^2, & \sigma_T^2(\theta) &= 16\theta(1 - \theta)(1 - 2\theta)^2, \\ \mu_R(\theta) = \mu_S(\theta) &= \frac{3}{2}\mu_T - \frac{1}{2}, & \text{and } \sigma_S^2(\theta) &= \frac{9}{4}\sigma_T^2; \end{aligned}$$

an expression for $\sigma_R^2(\theta)$ is rather lengthy and so is not included here. Plots of $\text{ARE}_{T,S}^{(\text{DSL})}$ and $\text{ARE}_{R,T}^{(\text{DSL})}$ are included in Figure 3.3a. As expected, the lower bound in (3.20) is attained for all distributions in this model; further, $\text{ARE}_{T,S}(\theta_0) = 1$ for all $\theta_0 \in [0, 1] \setminus \{\frac{1}{2}\}$ ($\sigma_T^2(\frac{1}{2}) = \sigma_S^2(\frac{1}{2}) = 0$, so $\text{ARE}_{T,S}(\frac{1}{2})$ is undefined).

Model DSU is characterized by letting X be uniformly distributed along $[0, 1]$ and setting the support of (X, Y) as the union of the $m + 1$ line segments joining the points $(0, \theta)$ and $(\theta, 0)$, $(\theta, 2\theta)$ and $(2\theta, \theta)$, \dots , $((m - 1)\theta, m\theta)$ and $(m\theta, (m - 1)\theta)$, and $(m\theta, 1)$ and $(1, m\theta)$,

where $m := m(\theta) := \lfloor \frac{1}{\theta} \rfloor$. One then has

$$\begin{aligned}\mu_T(\theta) &= -1 + 4m\theta - 2m\theta^2(1+m), \\ \sigma_T^2(\theta) &= 16m\theta(1-m\theta)(1-m\theta-\theta)^2, \\ \mu_R(\theta) = \mu_S(\theta) &= -1 + 6m\theta(1-m\theta) + 2m\theta^3(m^2-1), \\ \text{and } \sigma_S^2(\theta) &= \frac{9}{4}(1+\theta-m\theta)^2\sigma_T^2(\theta);\end{aligned}$$

again, an expression for $\sigma_R^2(\theta)$ is omitted due to its length. Note here that the upper bound in (3.20) is attained when $m = \frac{1}{\theta}$ (i.e. when $\frac{1}{\theta} \in \mathbb{N}$), though in such a case $\sigma_T^2 = \sigma_S^2 = 0$ and hence the ARE is undefined; however, $\text{ARE}_{T,S}(\theta_0) = 1$ for all other values of θ_0 . Plots of $\text{ARE}_{T,S}^{(\text{DSU})}$ and $\text{ARE}_{R,T}^{(\text{DSU})}$ are found in Figure 3.3.

3.4.3 Transformations of the bivariate normal model

Consider next the bivariate normal model, given the label \mathbf{N} , where (X, Y) has a normal distribution; the model is indexed by the correlation coefficient of X and Y . The asymptotic moments of R , S , and T are well-known in this model:

$$\mu_R(\theta) = \theta \quad \text{and} \quad \sigma_R^2(\theta) = (1 - \theta^2)^2$$

follows easily from (3.7) and (3.8) (or see [40] for further asymptotic expansions of the moments of R), the expressions

$$\mu_T(\theta) = \frac{2}{\pi} \sin^{-1} \theta, \quad \sigma_T^2(\theta) = \frac{4}{9} - \frac{16}{\pi^2} (\sin^{-1} \frac{\theta}{2})^2, \quad \text{and} \quad \mu_S(\theta) = \frac{6}{\pi} \sin^{-1} \frac{\theta}{2}$$

are derived in e.g. [67], and the formula

$$\sigma_S^2(\theta) = 1 - \frac{324}{\pi^2} (\sin^{-1} \frac{\theta}{2})^2 + \frac{72}{\pi^2} (I_1(\theta) + 2I_2(\theta) + 2I_3(\theta) + 4I_4(\theta)),$$

where

$$\begin{aligned}I_1(\theta) &:= \int_0^\theta \frac{\sin^{-1} \frac{u^3}{4(2-u^2)}}{\sqrt{4-u^2}} du, & I_2(\theta) &:= \int_0^\theta \frac{\sin^{-1} \frac{u}{2(3-u^2)}}{\sqrt{4-u^2}} du, \\ I_3(\theta) &:= \int_0^\theta \frac{\sin^{-1} \frac{u(4-u^2)}{2\sqrt{2}\sqrt{8-6u^2+u^4}}}{\sqrt{4-u^2}} du, & \text{and } I_4(\theta) &:= \int_0^\theta \frac{\sin^{-1} \frac{u(4-u^2)}{2\sqrt{12-7u^2+u^4}}}{\sqrt{4-u^2}} du\end{aligned}$$

is found in [82] (or see [27] for a much lengthier expression for $\text{Var}_\theta S_n$). It is also well-known that $\text{ARE}_{T,S}^{(\mathbf{N})}(0) = 1$ and $\text{ARE}_{R,T}^{(\mathbf{N})}(0) = \text{ARE}_{R,S}^{(\mathbf{N})}(0) = \frac{\pi^2}{9} = 1.0966\dots$ in this model. As can be seen from the plots of the pairwise ARE's illustrated in Figure 3.4, R is more efficient than T or S for any of the one-sided or two-sided tests in this model (note the ARE's are even functions of θ); this should not be surprising considering that R is the maximum-likelihood estimator of θ (at least when σ_X and σ_Y are also unknown parameters in the model).

In [82], we show that the three ARE's plotted in Figure 3.4 are strictly increasing on $(0, 1)$, so that $\text{ARE}_{R,T}^{(\mathbf{N})}(\theta_0) \in [1.0966\dots, 1.2091\dots)$ and $\text{ARE}_{R,S}^{(\mathbf{N})}(\theta_0) \in [1.0966\dots, 1.4395\dots)$ for $\theta_0 \in (-1, 1)$. Particularly, the efficiency of R over its two nonparametric competitors is not

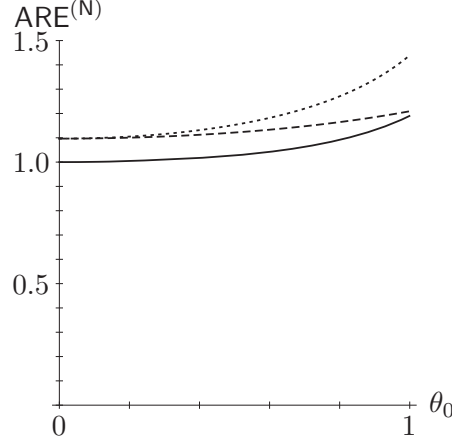


Figure 3.4: $\text{ARE}_{T,S}^{(N)}$ (solid), $\text{ARE}_{R,T}^{(N)}$ (dashed), and $\text{ARE}_{R,S}^{(N)}$ (dotted)

that great, which begs the question if R retains its efficiency in models where the normality assumption is perturbed in some fashion.

The main goal of this subsection is to examine some models which inherit their dependence structure from the bivariate normal distribution but in which the marginal distributions are no longer normal. We conjectured that when the tails of X and Y are heavier than those of a standard normal distribution, then R is no longer an efficient estimator of θ (and hence not an efficient test statistic in tests concerning θ), and this indeed appears to be the case in the models considered.

All of the subsequent models have the following general construction: let Z_1 and Z_2 be two standard normal r.v.'s with $\text{Cov}(Z_1, Z_2) = \theta \in \Theta = [-1, 1]$ (we may include the degenerate distributions corresponding to $\theta = \pm 1$ in the model, though the ARE will not be defined at these points), and let ψ be a strictly increasing function defined on the entire line \mathbb{R} . The model then consists of the distributions of $(X, Y) = (\psi(Z_1), \psi(Z_2))$, indexed by Θ . Note that the distributions of T and S are invariant to such a transformation (cf. 3.2.3), so that we need only find expressions for μ_R and σ_R^2 in the following models; since $\text{ARE}_{T,S}(\theta_0) \in [1, 1.1904\dots]$, and $\text{ARE}_{R,S} = \text{ARE}_{R,T} \cdot \text{ARE}_{T,S}$, we consider only plots of $\text{ARE}_{R,T}$ in what follows. It will further be helpful to note that, if ψ is an odd function, one has $\mu_X = \mu_Y = 0$ and hence

$$\sigma_R^2 = \frac{1}{\sigma_X^4} \mathbb{E}_\theta \left(\left(1 + \frac{\mu_R^2}{2}\right) X^2 Y^2 + \frac{\mu_R^2}{2} X^4 - 2\mu_R X^3 Y \right),$$

which follows from (3.8) after noting that X and Y have identical (marginal) distributions.

3.4.3.1 Bivariate lognormal model

Let $\psi(x) = e^{\alpha x}$ for $x \in \mathbb{R}$ and $\alpha > 0$; we call this model the bivariate lognormal model, and label it by LN, since X and Y both have a lognormal distribution when $\alpha = 1$. For $(s, t) \in \mathbb{R}^2$, let

$$M(s, t) := \mathbb{E}_\theta e^{sZ_1 + tZ_2} = e^{\frac{1}{2}(s^2 + 2\theta st + t^2)}, \quad (3.21)$$

denote the moment-generating function (m.g.f.) of (Z_1, Z_2) . Then $E_\theta X^j Y^k = E e^{\alpha(jZ_1 + kZ_2)} = M(\alpha j, \alpha k)$ for any $j, k \in \mathbb{N}$, so that

$$\mu_R(\theta) = \frac{E_\theta XY - \mu_X \mu_Y}{\sigma_X \sigma_Y} = \frac{M(\alpha, \alpha) - M(\alpha, 0)^2}{M(2\alpha, 0) - M(\alpha, 0)^2} = \frac{e^{\alpha^2 \theta} - 1}{e^{\alpha^2} - 1},$$

we omit an expression for $\sigma_R^2(\theta)$ due to its length, though it is clear that σ_R^2 may be readily expressed in terms of M . Figure 3.5 illustrates plots of $\text{ARE}_{R,T}^{(\text{LN})}(\theta_0)$ for several choices of α . Note the lack of symmetry (due of course to the lack of symmetry in the choice of ψ) in these plots. For small values of α , we see that R retains its efficiency over T for some values of θ_0 ; of particular interest, in the test for independence, one has $\text{ARE}_{R,T}^{(\text{LN})}(0) = \frac{\pi^2 \alpha^4}{9(e^{\alpha^2} - 1)^2}$, which is easily seen to be decreasing with respect to α . That is, $\text{ARE}_{R,T}^{(\text{LN})}(0) \geq 1$ for $\alpha \leq 0.3025 \dots$

3.4.3.2 Hyperbolic sine model

Let next $\psi(x) = \sinh(\alpha x) = \frac{1}{2}(e^{\alpha x} - e^{-\alpha x})$ for $x \in \mathbb{R}$ and $\alpha > 0$; call this model the hyperbolic sine model, and label it HS. Expressions for the moment $E_\theta X^j Y^k$ are again easily expressed in terms of the m.g.f. M in (3.21):

$$\begin{aligned} E_\theta X^j Y^k &= E_\theta \left(\frac{e^{\alpha Z_1} - e^{-\alpha Z_1}}{2} \right)^j \left(\frac{e^{\alpha Z_2} - e^{-\alpha Z_2}}{2} \right)^k \\ &= \frac{1}{2^{j+k}} \sum_{a=0}^j \sum_{b=0}^k \binom{j}{a} \binom{k}{b} (-1)^{j+k-a-b} M(\alpha(2a-j), \alpha(2b-k)). \end{aligned}$$

One then has

$$\mu_R(\theta) = \frac{e^{\alpha^2(1-\theta)}(e^{2\alpha^2\theta} - 1)}{e^{2\alpha^2} - 1},$$

and again an expression for $\sigma_R^2(\theta)$ is omitted due to its length. Figure 3.6 illustrates plots of $\text{ARE}_{R,T}^{(\text{HS})}(\theta_0)$ for several choices of α . We see that R retains its efficiency over T for relatively moderate values of α (as compared to the bivariate lognormal model); particularly, $\text{ARE}_{R,T}^{(\text{HS})}(0) = \frac{4\pi^2 \alpha^4 e^{2\alpha^2}}{9(e^{2\alpha^2} - 1)^2}$, which is routinely verified to be decreasing with respect to α , so that $\text{ARE}_{R,T}^{(\text{HS})}(0) \geq 1$ for $\alpha \leq 0.7269 \dots$

Note that, when α is small, $X \approx 1 + \alpha Z_1$ in the bivariate lognormal model and $X \approx \alpha Z_1$ in the hyperbolic sine model; as R is invariant to linear transformations, we would expect the distribution of R to be well-approximated by that in the bivariate normal model for small α . It is an easy exercise in calculus to show that $\mu_R(\theta_0)$ and $\sigma_R^2(\theta_0)$ in either of these models converge pointwise as $\alpha \rightarrow 0$ (we shall not worry about the entire distribution of R), so we simply state the following proposition without proof:

Proposition 3.4.1. *For $U \in \{T, S\}$ and any $\theta \in (-1, 1)$, $\text{ARE}_{R,U}^{(\text{LN})}(\theta) \rightarrow \text{ARE}_{R,U}^{(\text{N})}(\theta)$ and $\text{ARE}_{R,U}^{(\text{HS})}(\theta) \rightarrow \text{ARE}_{R,U}^{(\text{N})}(\theta)$ as $\alpha \downarrow 0$.*

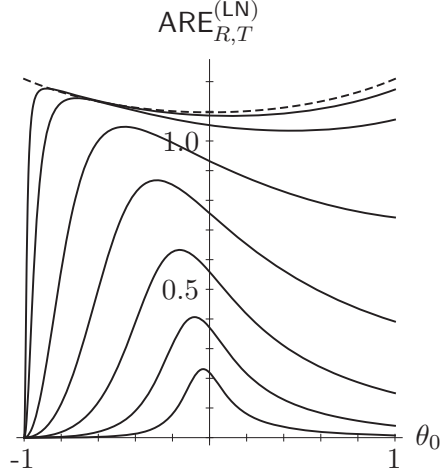


Figure 3.5: $\text{ARE}_{R,T}^{(\text{LN})}(\theta_0)$ (solid) in the bivariate lognormal model for $\alpha = 1.2, 1, 0.8, 0.6, 0.4, 0.2, 0.1$ in increasing order, respectively; $\text{ARE}_{R,T}^{(\text{N})}(\theta_0)$ (dashed) provided for visual comparison

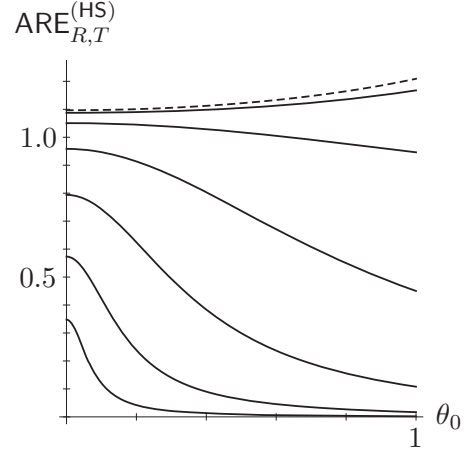


Figure 3.6: $\text{ARE}_{R,T}^{(\text{HS})}(\theta_0)$ (solid) in the hyperbolic sine model for $\alpha = 1.4, 1.2, 1, 0.8, 0.6, 0.4$ in increasing order, respectively; $\text{ARE}_{R,T}^{(\text{N})}(\theta_0)$ (dashed) provided for visual comparison

3.4.3.3 “Power tails” model

Next let $\psi(x) = (e^{\alpha x^2} - 1) \text{sgn}(x)$, where $0 < \alpha < \frac{1}{8}$. We dub this class of models with the label “power tails” (and label it as PT) since, using Mills’ ratio, one has

$$\mathbb{P}(X > x) = \mathbb{P}(Z > \sqrt{\ln(x+1)/\alpha}) \sim \sqrt{\frac{\alpha}{2\pi \ln(x+1)}} (1+x)^{-1/(2\alpha)} \sim \sqrt{\frac{\alpha}{2\pi}} x^{-1/(2\alpha)}$$

as $x \rightarrow \infty$.

Note that

$$\mathbb{E}_\theta X^j Y^k = \sum_{a=0}^j \sum_{b=0}^k (-1)^{j+k-a-b} \binom{j}{a} \binom{k}{b} \mathbb{E}_\theta e^{\alpha(aZ_1^2 + bZ_2^2)} \text{sgn}(Z_1)^j \text{sgn}(Z_2)^k.$$

Rewriting the above expectation as an integral, it is easy to see that the integrand (aside from the factor $\text{sgn}(Z_1)^j \text{sgn}(Z_2)^k$) is the kernel of another bivariate normal density with covariance matrix Σ_3 , where

$$\Sigma_1 := \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}, \quad \Sigma_2 := \begin{bmatrix} 2\alpha a & 0 \\ 0 & 2\alpha b \end{bmatrix}, \quad \text{and} \quad \Sigma_3 := (\Sigma_1^{-1} - \Sigma_2)^{-1} =: \begin{bmatrix} \sigma_1^2 & \tilde{\theta} \sigma_1 \sigma_2 \\ \tilde{\theta} \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix};$$

σ_1 , σ_2 , and $\tilde{\theta}$ are implicitly defined by the above equations. Using the transformation to polar coordinates $Z_1 \mapsto r\sigma_1 \cos \varphi$, $Z_2 \mapsto r\sigma_2(\tilde{\theta} \cos \varphi + \sqrt{1 - \tilde{\theta}^2} \sin \varphi) = r\sigma_2 \cos(\varphi - \xi)$, where

$\xi := \cos^{-1} \tilde{\theta}$, we then have

$$\begin{aligned} \mathbb{E}_{\theta} X^j Y^k &= \sum_{a=0}^j \sum_{b=0}^k (-1)^{j+k-a-b} \binom{j}{a} \binom{k}{b} \frac{|\Sigma_3|^{1/2}}{2\pi|\Sigma_1|^{1/2}} \\ &\quad \times \int_0^{\infty} r e^{-r^2/2} dr \int_0^{2\pi} \operatorname{sgn}(\cos \varphi)^j \operatorname{sgn}(\cos(\varphi - \xi))^k d\varphi \\ &= \sum_{a=0}^j \sum_{b=0}^k (-1)^{j+k-a-b} \binom{j}{a} \binom{k}{b} \frac{|\Sigma_3|^{1/2}}{2\pi|\Sigma_1|^{1/2}} \times \begin{cases} 2\pi, & j \equiv k \equiv 0 \pmod{2} \\ 4 \sin^{-1} \tilde{\theta}, & j \equiv k \equiv 1 \pmod{2} \\ 0, & j \not\equiv k \pmod{2}, \end{cases} \end{aligned}$$

where in the second case above the identity $\frac{\pi}{2} - \cos^{-1}(\tilde{\theta}) = \sin^{-1}(\tilde{\theta})$ is used,

$$\tilde{\theta} = \frac{\theta}{\sqrt{(1 - 2\alpha a(1 - \theta^2))(1 - 2\alpha b(1 - \theta^2))}},$$

and

$$\frac{|\Sigma_3|^{1/2}}{|\Sigma_1|^{1/2}} = \frac{1}{\sqrt{1 - 2\alpha(a + b) + 4\alpha^2 ab(1 - \theta^2)}}.$$

Note that, by Remark 3.2.1, there exists a $p > 4$ such that the moment restriction on R is satisfied for a given α . Plots of $\operatorname{ARE}_{R,T}^{(\text{PT})}(\theta)$ are shown in Figure 3.7 for selected values of α . We again see that T is asymptotically more efficient than R in this heavy-tailed distribution.

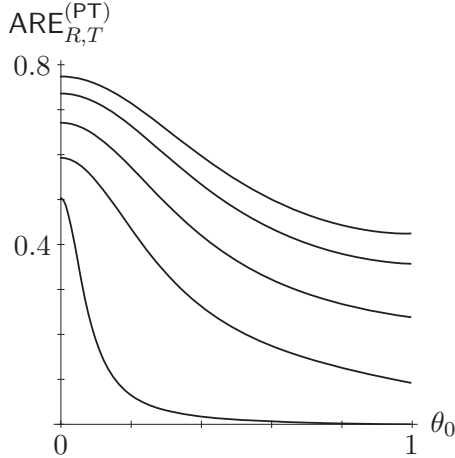


Figure 3.7: $\operatorname{ARE}_{R,T}^{(\text{PT})}(\theta_0)$ with $\alpha = 0.01, 1/32, 1/16, 3/32, 0.124$ (in decreasing order, respectively)

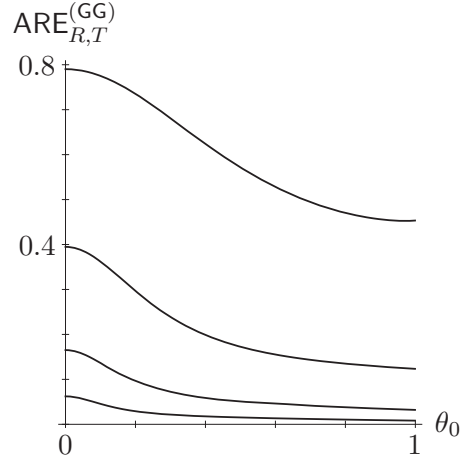


Figure 3.8: $\operatorname{ARE}_{R,T}^{(\text{GG})}(\theta_0)$ with $p = 2, 3, 4, 5$ (in decreasing order, respectively)

3.4.3.4 Generalized gamma model

Consider lastly the model defined by $\psi(x) = |x|^p \operatorname{sgn}(x)$ for $p \in \mathbb{N}$; we dub this class of models the (signed) generalized gamma models and label it by **GG**, as X shares the distribution of

$W - W^*$, where W and W^* are independent generalized gamma r.v.'s with density

$$f_W(x) = \frac{\beta}{\Gamma(k)\lambda} \left(\frac{x}{\lambda}\right)^{k\beta-1} e^{-(x/\lambda)^\beta}$$

for $x > 0$, scale parameter $\lambda = 2^{p/2}$, and shape parameters $k = \frac{1}{2}$ and $\beta = \frac{2}{p}$. In order to calculate μ_R and σ_R^2 we need a general formula for the mixed moments

$$\mathbb{E}_\theta X^j Y^k = \mathbb{E}_\theta |Z_1|^{pj} |Z_2|^{pk} \operatorname{sgn}(Z_1)^j \operatorname{sgn}(Z_2)^k.$$

In general, transforming (Z_1, Z_2) to a pair of independent normal r.v.'s and then to polar coordinates via $Z_1 \mapsto r \cos(\varphi)$ and $Z_2 \mapsto r(\theta \cos(\varphi) + \sqrt{1 - \theta^2} \sin(\varphi)) = r \cos(\varphi - \xi)$, where $\xi := \cos^{-1} \theta$, we have

$$\begin{aligned} \mathbb{E}_\theta X^j Y^k &= \left(\int_0^\infty \frac{1}{2\pi} r^{p(j+k)+1} e^{-r^2/2} dr \right) \\ &\times \left(\int_0^{2\pi} |\cos(\varphi)|^{pj} |\cos(\varphi - \xi)|^{pk} \operatorname{sgn}(\cos \varphi)^j \operatorname{sgn}(\cos(\varphi - \xi))^k d\varphi \right). \end{aligned} \quad (3.22)$$

The first integral above is easily seen to be equal to $\frac{(\frac{p}{2}(j+k))!}{2\pi} 2^{\frac{p}{2}(j+k)}$. A general formula for the second integral above (which will depend on the parities of j , k , and p) is straightforward to obtain using the identity $\cos x + i \sin x = e^{xi}$ for all real x coupled with a binomial expansion. Explicit formulas for μ_R and σ_R^2 for specific values of p are not provided due to their length, though Figure 3.8 illustrates the $\operatorname{ARE}_{R,T}^{(\text{GG})}$ curves for this class of models with $p = 2, 3, 4, 5$. We see that T is much more efficient in the limit than is R , with $\operatorname{ARE}_{R,T}^{(\text{GG})}$ fast approaching 0 as p increases.

As suggested by the plots in Figures 3.7 and 3.8, it appears that $\operatorname{ARE}_{R,T}^{(\text{PT})}$ converges pointwise, as $\alpha \downarrow 0$, to $\operatorname{ARE}^{(\text{GG})}$ with $p = 2$. This is intuitively clear, since for small α one has $X = (e^{\alpha Z_1^2} - 1) \operatorname{sgn}(Z_1) \approx \alpha Z_1^2 \operatorname{sgn}(Z_1)$. This heuristic leads us to a correct conclusion, as indicated by the following proposition:

Proposition 3.4.2. *Let $\operatorname{ARE}_\alpha^{(\text{PT})}$ and $\operatorname{ARE}_2^{(\text{GG})}$ denote the ARE (between any of R , S , or T) in the power-tails model with parameter $\alpha \in (0, \frac{1}{8})$ and the generalized gamma model with parameter $p = 2$, respectively. Then, for any $\theta \in (-1, 1)$, $\operatorname{ARE}_\alpha^{(\text{PT})}(\theta) \rightarrow \operatorname{ARE}_2^{(\text{GG})}(\theta)$ as $\alpha \rightarrow 0$.*

Proof. Again, this is a straightforward (though tedious) exercise in calculus. Details are omitted, though the general method is as follows: for $\alpha \in (0, \frac{1}{8})$, let $\mathbb{E}_\theta^{(\alpha)}$ and $\mathbb{E}_\theta^{(0)}$ denote the expectation operator in the power-tails model with parameter α and in the generalized gamma model with parameter $p = 2$, respectively. Further let $\sigma^2(\alpha) := \mathbb{E}_\theta^{(\alpha)} X^2 = 1 - \frac{2}{\sqrt{1-2\alpha}} + \frac{1}{\sqrt{1-4\alpha}}$ and $\sigma^2(0) := \mathbb{E}_\theta^{(0)} X^2 = 3$. Then verify that $\mathbb{E}_\theta^{(\alpha)} X^j Y^k / \sigma^{j+k}(\alpha)$ converges to $\mathbb{E}_\theta^{(0)} X^j Y^k / \sigma^{j+k}(0)$ for all $\theta \in (-1, 1)$ as $\alpha \rightarrow 0$; this shows that the asymptotic moments of R (in the power-tails model) converge pointwise to those in the generalized gamma model, which is sufficient to demonstrate the statement of the proposition. \square

3.5 Simulations of the relative efficiency

The ARE is convenient as a measure of the relative performance between two test statistics primarily due to its ease in calculation, though the practicing statistician might balk at its relevance to actual experiments. A more natural, and more difficult to calculate, measure of relative performance is the RE, as defined in (3.4). We present here the results of estimating the RE between R , S , and T for a few selected models of Section 3.4 (namely, the bivariate normal and the four MICD models), with the goal of evaluating the utility of using the ARE in lieu of the RE when deciding what test statistic to use for a particular problem.

The simulations considered here are based on an upper-tail test of $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$ for $\theta_0 \in \{0, 0.1, \dots, 0.9\}$, where $\alpha = \beta = 0.05$ are kept fixed. We allow for these tests to be randomized so as to achieve an exact size of α ; particularly, for $W \in \{R, S, T\}$, let

$$p_{W,n}(\theta) := P_\theta(W_n > \gamma) + \pi P_\theta(W_n = \gamma) \quad (3.23)$$

denote the power function, where $\gamma \in \mathbb{R}$ and $\pi \in [0, 1)$ are numbers uniquely chosen to satisfy $p_{W,n}(\theta_0) = \alpha$.

We wish to produce plots of the RE which are visually comparable to those of the ARE (as in Figures 3.4 and 3.2), with the goal of observing them converge to the theoretical ARE plot as θ_1 nears θ_0 , or alternatively as n_W increases. There are two obvious routes to achieving this goal: fix $\delta := \theta_1 - \theta_0$ and estimate each of n_R , n_S , and n_T as functions of δ ; alternatively, fix n_R , say, and estimate θ_1 , n_T , and n_S as functions of n_R .

We adopt here a modification of the latter approach. Temporarily accept the fantasy that R is normally distributed with mean $\mu_R(\theta)$ and variance $\sigma_R^2(\theta)/n$ for all $n \in \mathbb{N}$ and $\theta \in (-1, 1)$; then, for each pair $(\theta_0, \tilde{n}_R) \in \{0, 0.1, \dots, 0.9\} \times \{10, 40, 70\}$, the power function is known exactly. The equation $p_{R,\tilde{n}_R}(\theta_1) = 0.95$ will either have a unique solution or no solutions (if, say, θ_0 is too near 1 or \tilde{n}_R is too small); particularly, we solve the equation

$$z_{0.05} = \frac{\sigma_R(\theta_0)}{\sigma_R(\theta_1)} \left(z_{0.05} + \frac{\mu_R(\theta_0) - \mu_R(\theta_1)}{\sigma_R(\theta_0)/\sqrt{\tilde{n}_R}} \right) \quad (3.24)$$

for $\theta_1 \in (\theta_0, 1)$ whenever a solution exists, where $z_q := \Phi^{-1}(1 - q)$ for $q \in (0, 1)$. Then, for each combination of θ_0 and \tilde{n}_R for which a solution to (3.24) exists, we estimate each of n_R , n_S , and n_T (as functions of the now fixed parameters θ_0 , θ_1 , α , and β) through simulation; it is at this point we step out of the fantasy that R is normally distributed, as n_R will generally not be equal to \tilde{n}_R .

Assume then that θ_0 and θ_1 are fixed, as well as $W \in \{R, S, T\}$. We obtain an estimate \hat{n}_W of n_W through the following procedure. For any value of n under consideration, estimate the power function $p_{W,n}$ in (3.23) by estimating π and γ from the empirical distribution of $M_1 = 10^4$ instances of W_n (obtained by pseudorandomly generating pairs (X, Y) with d.f. F_{θ_0}). The accuracy of the estimate $\hat{p}_{W,n}$ is assessed by generating two more samples of $M_2 = 4000$ instances of W_n from the null distribution. If the number of tests which reject the null hypothesis lies in the interval $M_2\alpha \pm 2\sqrt{M_2\alpha(1-\alpha)}$ for each of the two new samples then we are satisfied with the accuracy of $\hat{p}_{W,n}$; otherwise we generate M_1 more observations of W_n , update the estimate $\hat{p}_{W,n}$ (from the empirical distribution of $2M_1$ instances of W_n), recheck the accuracy of $\hat{p}_{W,n}$, and continue this iterative procedure until the accuracy is deemed acceptable. We then generate a sample of $M_3 = 2 \times 10^4$ observations of W_n from

the alternative distribution, obtaining an estimate $\hat{p}_{W,n}(\theta_1)$ of the power. In order to obtain an estimate \hat{n}_W of n_W , we iterate this procedure by obtaining a sequence of estimates $\hat{p}_{W,n}(\theta_1)$ until either: (1) $\hat{p}_{W,n-1}(\theta_1) < 0.95 \leq \hat{p}_{W,n}(\theta_1) \leq \hat{p}_{W,n+1}(\theta_1) \leq \hat{p}_{W,n+2}(\theta_1)$, or (2) $\hat{p}_{W,n-1}(\theta_1) < 0.95$ and $\hat{p}_{W,n+i}(\theta_1) \geq 0.95$ for $i = 0, \dots, 4$; this seemingly peculiar stopping rule was designed to account for the nonmonotonicity (in n) of the power functions of the discrete statistics T and S . When such an n is found, we let \hat{n}_W be obtained through linear interpolation:

$$\hat{n}_W := n - \frac{\hat{p}_{W,n}(\theta_1) - 0.95}{\hat{p}_{W,n}(\theta_1) - \hat{p}_{W,n-1}(\theta_1)}.$$

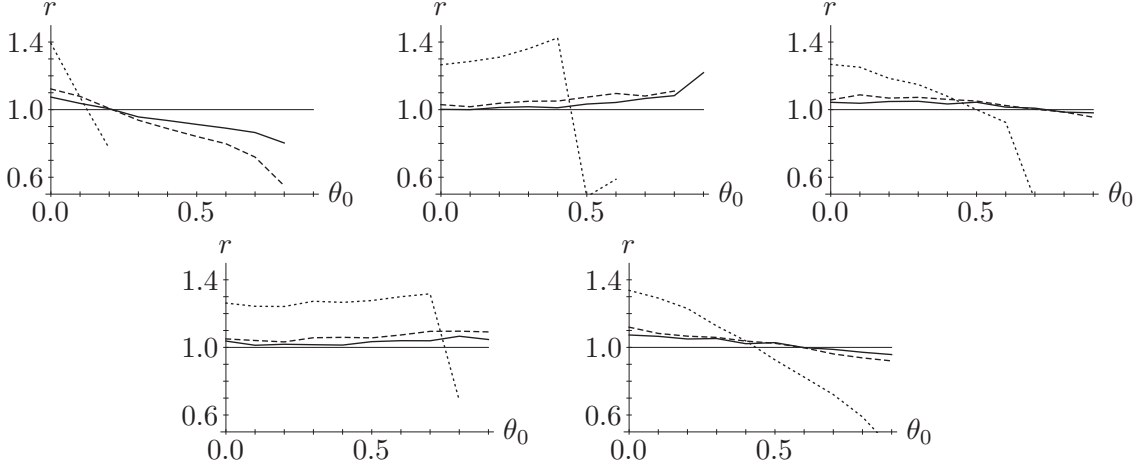
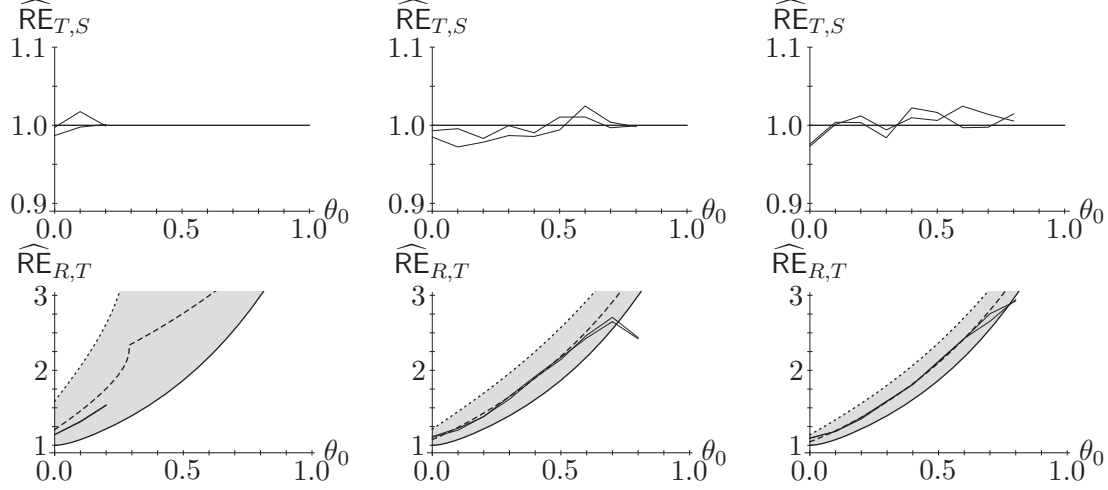


Figure 3.9: Plots of $r := \hat{n}_R / \tilde{n}_R$ against θ_0 for the MICD AS, AL, OS, OL models and the bivariate normal N model (left to right, respectively); $\tilde{n}_R = 10$ (dotted), $\tilde{n}_R = 40$ (dashed), $\tilde{n}_R = 70$ (thick solid)

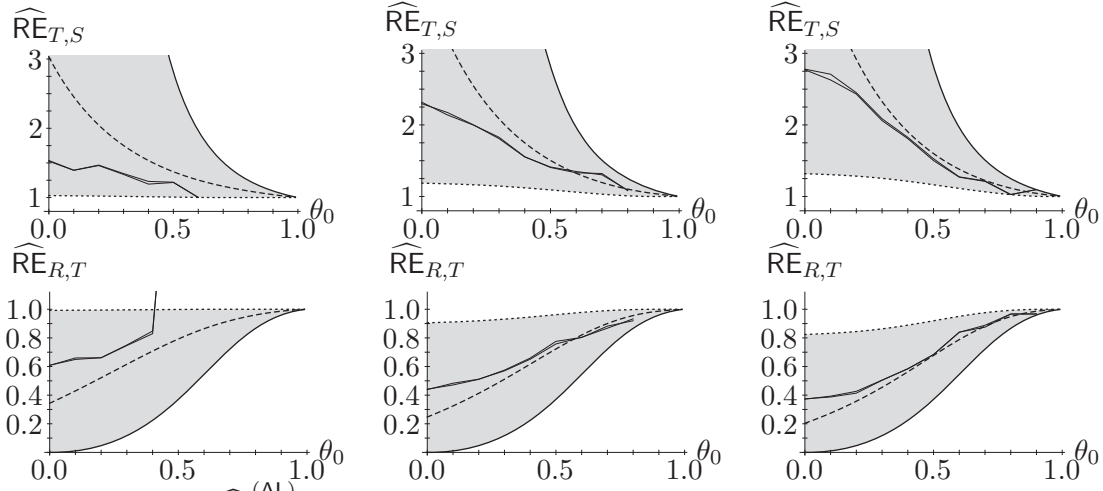
For each combination of $(W, \theta_0, \tilde{n}_R) \in \{\{R, S, T\}, \{0, 0.1, \dots, 0.9\}, \{10, 40, 70\}\}$ (for which a solution to (3.24) exists), two independent estimates \hat{n}_W were obtained. Figure 3.9 shows plots of the ratio of the mean of the two estimates \hat{n}_R to the target value \tilde{n}_R for the five models under consideration. We hope to see that \hat{n}_R is close to \tilde{n}_R (suggesting that our “fantasy assumption” that R is normally distributed isn’t quite so fantastical), and indeed that is observed; for $\tilde{n}_R = 10$ the ratio \hat{n}_R / \tilde{n}_R can deviate fairly far from 1, though for the larger values of \tilde{n}_R this ratio is generally rather close to 1.

From the two independent estimates \hat{n}_W we obtain two estimates $\widehat{\text{RE}}_{T,S}(\theta_0) = \hat{n}_S / \hat{n}_T$ and $\widehat{\text{RE}}_{R,T}(\theta_0) = \hat{n}_T / \hat{n}_R$ for the five models under consideration; these estimates of the RE are plotted in Figure 3.10. While $\widehat{\text{RE}}(\theta_0)$ is expected to approach $\text{ARE}(\theta_0)$ for very large \tilde{n}_R , there is no apparent reason that we should expect these two values to be close for small or moderate values of \tilde{n}_R . Indeed, considering that $\theta_1 - \theta_0$ will be relatively large for small \tilde{n}_R , it is plausible that $\text{RE}(\theta_0)$ is closer to $\text{ARE}(\theta^*)$ for some $\theta^* \in [\theta_0, \theta_1]$. Our simulations appear to suggest this, and plots of $\text{ARE}(\theta_0)$, $\text{ARE}((\theta_0 + \theta_1)/2)$, and $\text{ARE}(\theta_1)$ are provided alongside the plots of $\widehat{\text{RE}}(\theta_0)$ to provide a graphical justification for this heuristics. The space between the two “extreme” plots of $\text{ARE}(\theta_0)$ and $\text{ARE}(\theta_1)$ is lightly shaded to represent the plots of all possible such $\text{ARE}(\theta^*)$.

As evident from Figure 3.10, the estimates $\widehat{\text{RE}}$ get rather close to the limiting value



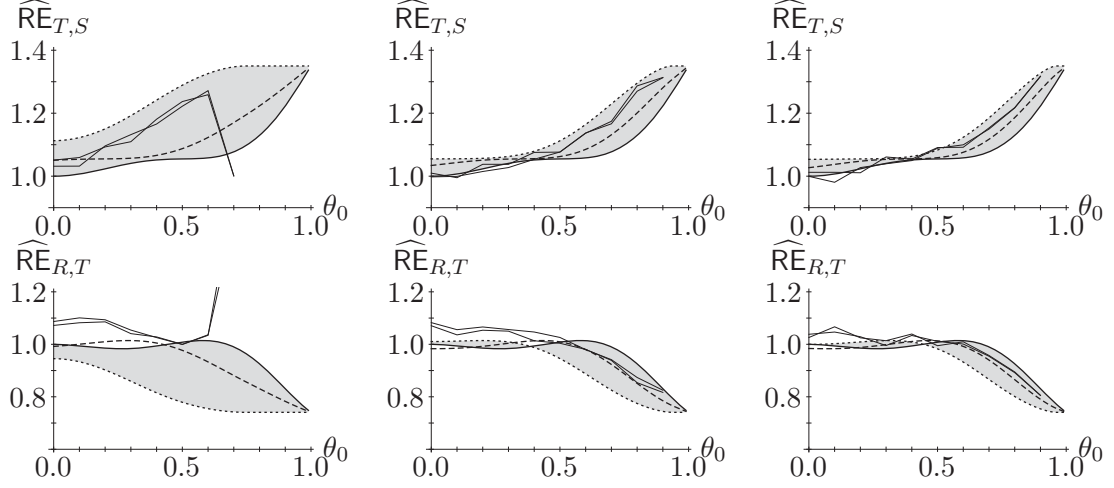
(a) $\widehat{\text{RE}}^{(\text{AS})}$ for $\tilde{n}_R = 10$ (left), $\tilde{n}_R = 40$ (middle), $\tilde{n}_R = 70$ (right)



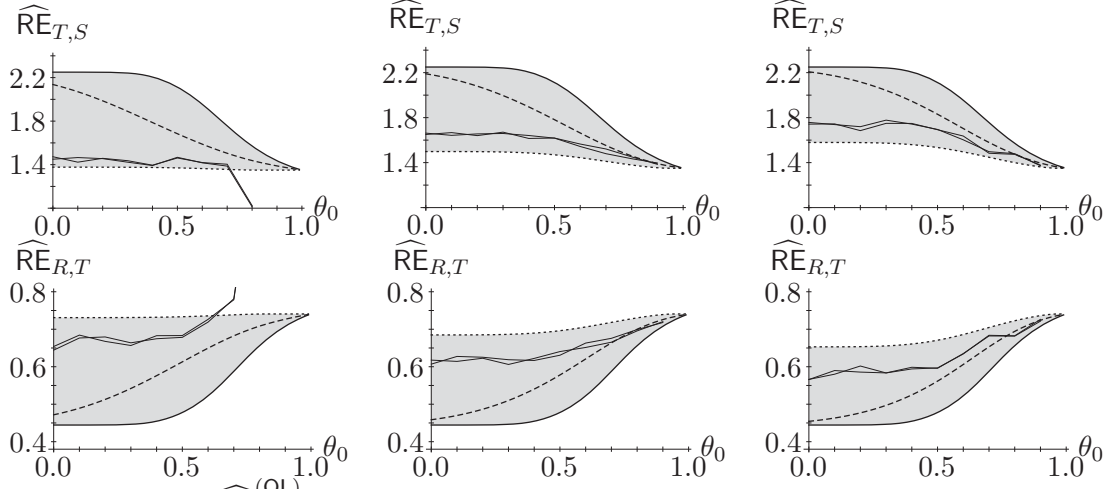
(b) $\widehat{\text{RE}}^{(\text{AL})}$ for $\tilde{n}_R = 10$ (left), $\tilde{n}_R = 40$ (middle), $\tilde{n}_R = 70$ (right)

Figure 3.10: Plots of $\widehat{\text{RE}}$ with $\text{ARE}(\theta_0)$ (solid), $\text{ARE}((\theta_0 + \theta_1)/2)$ (dashed), and $\text{ARE}(\theta_1)$ (dotted) provided for visual comparison

$\text{ARE}(\theta_0)$ as \tilde{n}_R increases, though of course the rate of this convergence depends on θ_0 and also the model in question. There is a remarkable convergence of the estimates $\widehat{\text{RE}}(\theta_0)$ to $\text{ARE}((\theta_0 + \theta_1)/2)$ (as opposed to converging to $\text{ARE}(\theta_0)$) as \tilde{n}_R increases; this is evident most clearly in the MICD models. This convergence is quickest for the MICD model AS; moreover, the accuracy of the approximation of $\widehat{\text{RE}}^{(\text{AS})}(\theta_0)$ to $\text{ARE}^{(\text{AS})}((\theta_0 + \theta_1)/2)$ appears to be rather uniform in θ_0 . For the remaining three MICD models, this convergence appears to be quicker for large values of θ_0 , as opposed to the convergence appearing to be quicker for θ_0 nearer 0 in MICD model AS and the bivariate normal model. We see that $\text{RE}^{(\text{N})}$ is much larger than 1; this agrees with our intuition, as R is the likelihood-ratio test statistic in the normal model. We see that even for small values of \tilde{n}_R , $\widehat{\text{RE}}(\theta_0) - 1$ generally shares



(c) $\widehat{\text{RE}}^{(\text{OS})}$ for $\tilde{n}_R = 10$ (left), $\tilde{n}_R = 40$ (middle), $\tilde{n}_R = 70$ (right)



(d) $\widehat{\text{RE}}^{(\text{OL})}$ for $\tilde{n}_R = 10$ (left), $\tilde{n}_R = 40$ (middle), $\tilde{n}_R = 70$ (right)

Figure 3.10: Plots of $\widehat{\text{RE}}$ with $\text{ARE}(\theta_0)$ (solid), $\text{ARE}((\theta_0 + \theta_1)/2)$ (dashed), and $\text{ARE}(\theta_1)$ (dotted) provided for visual comparison

the same sign as $\text{ARE}(\theta_0) - 1$ for any of the models under consideration; this lends some empirical evidence that the practice of using the ARE as a selection tool for test statistics is justifiable.

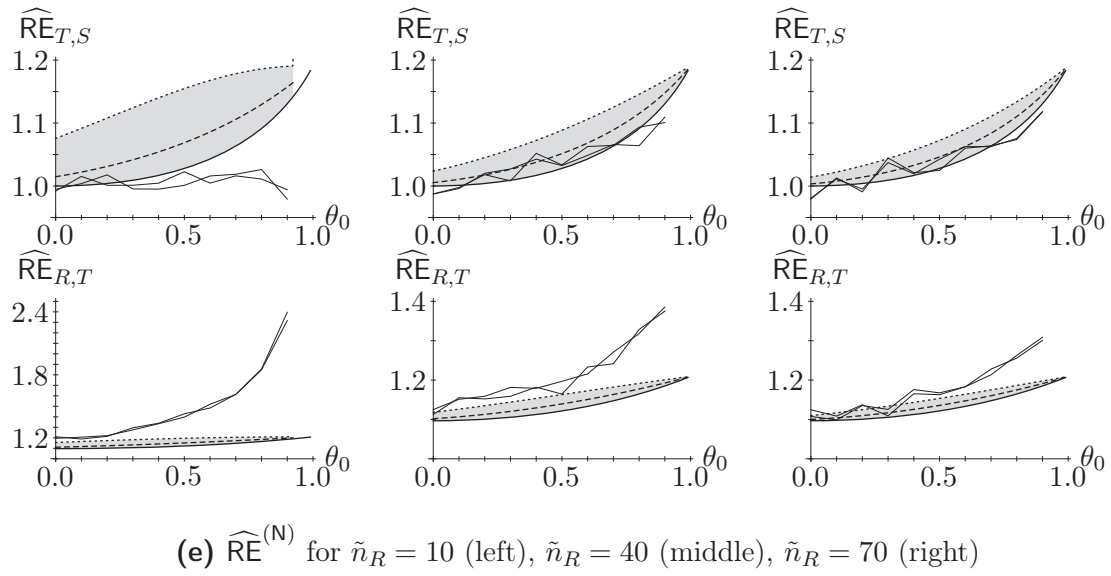


Figure 3.10: Plots of $\widehat{\text{RE}}$ with $\text{ARE}(\theta_0)$ (solid), $\text{ARE}((\theta_0 + \theta_1)/2)$ (dashed), and $\text{ARE}(\theta_1)$ (dotted) provided for visual comparison

Chapter 4

Monotonicity of the asymptotic relative efficiency between common correlation statistics in the bivariate normal model

4.1 Introduction

Pearson's R , Spearman's S and Kendall's T are the three most commonly used correlation statistics, the latter two especially in nonparametric studies. When the population distribution is bivariate normal, the question of independence between the two random variables (r.v.'s) reduces to deciding if the population correlation ρ is 0. In the case of testing the null hypothesis $\rho = 0$, it is known that the Pitman asymptotic relative efficiency (ARE) of R to S is $\frac{\pi^2}{9}$ [56] and that of T to S is 1 [83] (and hence the ARE of R to T is $\frac{\pi^2}{9}$ as well). While perhaps less common in practice, one could also use any three of these statistics to test hypotheses of the form $\rho = \rho_0$ (against alternatives $\rho > \rho_0$, $\rho < \rho_0$, or $\rho \neq \rho_0$) for arbitrary $\rho_0 \in (-1, 1)$. In [15], values of $\text{ARE}_{S,R}(\rho_0)$ (the ARE of S to R for the null hypothesis $\rho = \rho_0$) are tabulated for several values of $\rho_0 \in [0, 1]$; several values of $\text{ARE}_{T,R}(\rho_0)$ are given in [84] as well.

In this paper, we show that $\text{ARE}_{R,T}(|\rho_0|)$ is strictly increasing in $|\rho_0| \in [0, 1]$ from 1.096... to 1.209..., $\text{ARE}_{R,S}(|\rho_0|)$ increases from 1.096... to 1.439..., and $\text{ARE}_{T,S}(|\rho_0|)$ increases from 1 to 1.190.... Thus, all these ARE's stay rather close to 1 for all values of $\rho_0 \in (-1, 1)$. Additionally, we prove the existence of several upper and lower quadratic bounds for each of $\text{ARE}_{R,T}$, $\text{ARE}_{R,S}$ and $\text{ARE}_{T,S}$. All of these results are immediate corollaries to a stronger result, stated in this paper as Theorem 4.2.1.

For testing the null hypothesis $\theta = \theta_0$ (against any of the alternatives $\theta \neq \theta_0$, $\theta > \theta_0$, or $\theta < \theta_0$) in the framework of a given statistical model, under certain general conditions there exists an easily applicable formula for computing the ARE between two (sequences of) real-valued test statistics $T_1 = (T_{1,n})_{n \in \mathbb{N}}$ and $T_2 = (T_{2,n})_{n \in \mathbb{N}}$. The main condition (see e.g. [92, 52, 72]) is that the distribution function (d.f.) of either properly normalized test statistic uniformly converges to the standard normal d.f. Φ as the sample size n tends to ∞ . Particularly, if there exist continuous real-valued functions μ_{T_j} and σ_{T_j} on the parameter

The material contained in this chapter has been submitted to the pre-print server <http://arxiv.org>.

space Θ such that

$$\sup_{\theta \in \mathcal{V}} \sup_{z \in \mathbb{R}} \left| \mathbb{P}_\theta \left(\frac{T_{j,n} - \mu_{T_j}(\theta)}{\sigma_{T_j}(\theta)/\sqrt{n}} \leq z \right) - \Phi(z) \right| \xrightarrow{n \rightarrow \infty} 0, \quad (4.1)$$

where \mathcal{V} is some neighborhood of θ_0 chosen such that μ_{T_j} is continuously differentiable on \mathcal{V} and $\inf_{\theta \in \mathcal{V}} \sigma_{T_j}(\theta) > 0$ for $j = 1, 2$, then the ARE of T_1 to T_2 may be expressed by the formula

$$\text{ARE}(\theta_0) := \text{ARE}_{T_1, T_2}(\theta_0) = \frac{\sigma_{T_2}^2(\theta_0)}{\sigma_{T_1}^2(\theta_0)} \frac{\mu'_{T_1}(\theta_0)^2}{\mu'_{T_2}(\theta_0)^2}, \quad (4.2)$$

assuming that $\mu'_{T_j}(\theta_0) > 0$. The functions μ_{T_j} and σ_{T_j}/\sqrt{n} may be called the asymptotic mean and standard deviation, respectively, of the sequence T_j .

Berry-Esseen bounds provide a nice way to verify the condition (4.1). Such bounds for the Kendall and Spearman statistics, which are instances of so-called U - and V -statistics, are essentially well known; see e.g. [73]. In fact, we are using here a result by Chen and Shao [19] and a convenient representation of any V -statistic as a U -statistic [51]. As for a Berry-Esseen bound for the Pearson correlation statistic, we are using an apparently previously unknown result in [116].

According to the formula (4.2), the ARE between two test statistics can be expressed in terms of the asymptotic means and variances of the two statistics. In turn, the asymptotic variance of either T or S in the bivariate normal model can be expressed using Schläfli's formula [131] for the volume of the spherical tetrahedron in \mathbb{R}^4 . Such formulas have been of significant interest to a number of authors; see e.g. the recent papers [70] and [86]. We remark also that Plackett [121] obtained a result more general than Schläfli's. Actually, here we are using formulas by David and Mallows [27] which are based on [121].

To prove the main result, we use l'Hospital-type rules for the monotonicity pattern of a function $r = \frac{f}{g}$ on some interval (a, b) . Knowledge of the monotonicity of $\frac{f'}{g'}$ on (a, b) , along with the sign of gg' on (a, b) , allows one to obtain the monotonicity pattern of r ; see Pinelis [103] and the bibliography there for several variants of these rules and applications to various problems. For convenient reference these rules are restated in Section 4.3.2.

4.2 Monotonicity properties of the ARE in the bivariate normal model

Let $(V_n) =: ((X_n, Y_n))$ be a sequence of independent, identically distributed (i.i.d.) nondegenerate bivariate normal r.v.'s with

$$\mathbb{E} V_1 =: (\mu_X, \mu_Y) \quad \text{and} \quad \text{Cov}(V_1) =: \begin{bmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix};$$

note that $\rho \in (-1, 1)$ is the correlation coefficient between X_1 and Y_1 . Let

$$R := R_n := \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}, \quad (4.3)$$

where $(\bar{X}, \bar{Y}) := \frac{1}{n} \sum_{i=1}^n V_i$; R is commonly called Pearson's product-moment correlation coefficient, and it is the maximum-likelihood estimator of ρ . Spearman's rank correlation is

$$S := S_n := \frac{12}{n^3 - n} \sum_{i=1}^n r(X_i)r(Y_i) - \frac{3(n+1)}{n-1}, \quad (4.4)$$

where $r(X_i) := \sum_{j=1}^n \mathbf{I}\{X_j \leq X_i\}$ and $r(Y_i) := \sum_{k=1}^n \mathbf{I}\{Y_k \leq Y_i\}$ are the ranks (and $\mathbf{I}\{\cdot\}$ denotes the indicator function). Note that S is simply the product-moment correlation of the sample of ranks $(r(X_1), r(Y_1)), \dots, (r(X_n), r(Y_n))$. Let next

$$J_{ij} := \mathbf{I}\{X_j < X_i\} \mathbf{I}\{Y_j < Y_i\},$$

and let

$$T := T_n := \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} h_T(V_i, V_j) \quad (4.5)$$

denote Kendall's correlation statistic, where $h_T(V_i, V_j) := 2(J_{ij} + J_{ji}) - 1$, so that almost surely (a.s.) $h_T(V_i, V_j) = \pm 1$ depending on whether the pair (V_i, V_j) is concordant or discordant; also, $\mathbf{E} h_T(V_i, V_j) = 0$ if V_i and V_j are independent.

Consider the hypothesis test of $\rho = \rho_0$ against the alternative $\rho \neq \rho_0$ (or again, either of the two one-sided alternatives), where $\rho_0 \in (-1, 1)$. We shall show that each of R , S , and T satisfies the condition (4.1), so that (4.2) may be used to express the ARE between any two of these statistics. Further, it is easy to see, and also will be clear from what follows, that σ_R^2 , σ_S^2 , and σ_T^2 are all even functions of ρ , and also μ_R , μ_S , and μ_T are odd functions, so that the ARE of any pair of these statistics is even. See Figure 4.1 for a plot of these three functions, and note it suggests each of the pairwise

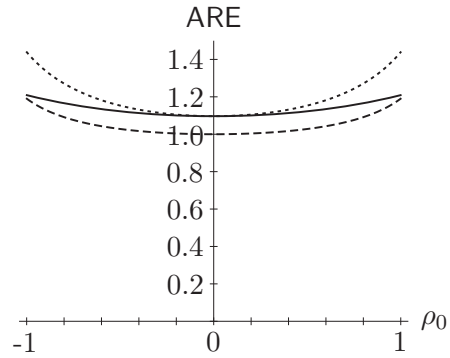


Figure 4.1: Plots of $\text{ARE}_{R,T}(\rho_0)$ (solid), $\text{ARE}_{T,S}(\rho_0)$ (dashed) and $\text{ARE}_{R,S}(\rho_0)$ (dotted)

ARE's is strictly increasing on $(0, 1)$. We also see that, while S and T are asymptotically equally efficient in the test for independence (i.e. $\rho_0 = 0$), the Kendall rank correlation has a better performance than the Spearman rank correlation for any other test of a positive (or negative) correlation; neither of these statistics performs better than the Pearson correlation, which is hardly surprising given that R is the maximum likelihood estimator of ρ in the bivariate normal model. Further, the shapes of these plots suggest the functions may be well-approximated by a quadratic polynomial. Indeed, the monotonicity of the ARE and a quadratic approximation shall be immediate results of the following:

Theorem 4.2.1. *For the test of the null hypothesis $\rho = \rho_0$ against any of the three alternative hypotheses: $\rho \neq \rho_0$, $\rho > \rho_0$, or $\rho < \rho_0$, let*

$$q_a(\rho_0) := q_{T_1, T_2; a}(\rho_0) := \frac{\text{ARE}_{T_1, T_2}(\rho_0) - \text{ARE}_{T_1, T_2}(a) - \text{ARE}'_{T_1, T_2}(a)(\rho_0 - a)}{(\rho_0 - a)^2}$$

for $\rho_0 \in [0, 1) \setminus \{a\}$ and $a \in [0, 1]$, where T_i is one of the statistics R , S , or T ; here and in what follows, $\text{ARE}_{T_1, T_2}(a)$ and $\text{ARE}'_{T_1, T_2}(a)$ are understood to mean $\text{ARE}_{T_1, T_2}(1-)$ and $\text{ARE}'_{T_1, T_2}(1-)$, respectively, when $a = 1$. Then

(RT0) $q_{R, T; 0}$ is increasing from 0.0966... to 0.1125...;

(RT1) $q_{R, T; 1}$ is increasing from 0.1510... to 0.2247...;

(TS0) $q_{T, S; 0}$ is increasing from 0.0984... to 0.1904...;

(TS1) $q_{T, S; 1}$ is increasing from 0.5516... to 1.8200...;

(RS0) $q_{R, S; 0}$ is increasing from 0.2045... to 0.3428...;

(RS1) $q_{R, S; 1}$ is increasing from 0.8682... to 2.6639....

The term “increasing” will mean for us “strictly increasing,” and similarly “decreasing” will mean “strictly decreasing.” Exact expressions for the endpoint values 0.0966..., 0.1125..., ... are found in the proof of Theorem 4.2.1. All proofs are deferred to Section 4.3.

Corollary 4.2.2. *For the test of the null hypothesis $\rho = \rho_0$ against any of the three alternative hypotheses: $\rho \neq \rho_0$, $\rho > \rho_0$, or $\rho < \rho_0$, one has*

(RT) $\text{ARE}_{R, T}(|\rho_0|)$ is increasing in $|\rho_0| \in (0, 1)$ from $\frac{\pi^2}{9} = 1.0966...$ to $\frac{2\pi\sqrt{3}}{9} = 1.2091...$;

(TS) $\text{ARE}_{T, S}(|\rho_0|)$ is increasing in $|\rho_0| \in (0, 1)$ from 1 to $\frac{9\sqrt{3}(11\sqrt{5}-15)}{40\pi} = 1.1904...$;

(RS) $\text{ARE}_{R, S}(|\rho_0|)$ is increasing in $|\rho_0| \in (0, 1)$ from $\frac{\pi^2}{9} = 1.0966...$ to $\frac{3(11\sqrt{5}-15)}{20} = 1.4395....$

This corollary justifies the conjecture that the pairwise ARE's are increasing on $(0, 1)$, which one would make from observing Figure 4.1. One immediate consequence of Corollary 4.2.2 is that the ARE's between R , S , and T remain in relatively small intervals. Indeed, $\text{ARE}_{R, T}$ increases by only about 10% as the null distribution moves from independence to the extreme case of almost sure dependence ($\text{ARE}_{R, T}(1-)/\text{ARE}_{R, T}(0) = 1.1026...$), with the largest increase in the ARE being attributed to the comparison of R to S ($\text{ARE}_{R, S}(1-)/\text{ARE}_{R, S}(0) = 1.3126...$).

Figure 4.1 also suggests the conjecture that these functions are convex on $(-1, 1)$; we have not yet proven the truth of such a conjecture, though it is conceivable that the methods used to prove Theorem 4.2.1 could be adapted to prove that each of $\text{ARE}'_{R, T}$, $\text{ARE}'_{R, S}$, and $\text{ARE}'_{T, S}$ are increasing on $(0, 1)$ (which, along with the aforementioned evenness of the ARE's, would imply convexity on $(-1, 1)$).

Corollary 4.2.3. *Let*

$$L_a(x) := L_{T_1, T_2; a}(x) := \text{ARE}_{T_1, T_2}(a) + \text{ARE}'_{T_1, T_2}(a)(|x| - a) + q_{T_1, T_2; a}(0+)(|x| - a)^2,$$

$$U_a(x) := U_{T_1, T_2; a}(x) := \text{ARE}_{T_1, T_2}(a) + \text{ARE}'_{T_1, T_2}(a)(|x| - a) + q_{T_1, T_2; a}(1-)(|x| - a)^2,$$

$$L(x) := L_{T_1, T_2}(x) := L_{T_1, T_2; 0}(x) \vee L_{T_1, T_2; 1}(x),$$

$$\text{and } U(x) := U_{T_1, T_2}(x) := U_{T_1, T_2; 0}(x) \wedge U_{T_1, T_2; 1}(x)$$

for $(T_1, T_2) \in \{(R, T), (T, S), (R, S)\}$, $x \in (-1, 1)$, and $a \in \{0, 1\}$. Then for all $\rho_0 \in (-1, 0) \cup (0, 1)$

$$L_a(\rho_0) \leq L(\rho_0) < \text{ARE}(\rho_0) < U(\rho_0) \leq U_a(\rho_0).$$

These piecewise quadratic bounds are illustrated in Figure 4.2. Note that L_0 and U_0 give good quadratic approximations to the ARE near the origin, while L_1 and U_1 are better approximations when ρ_0 is near ± 1 .

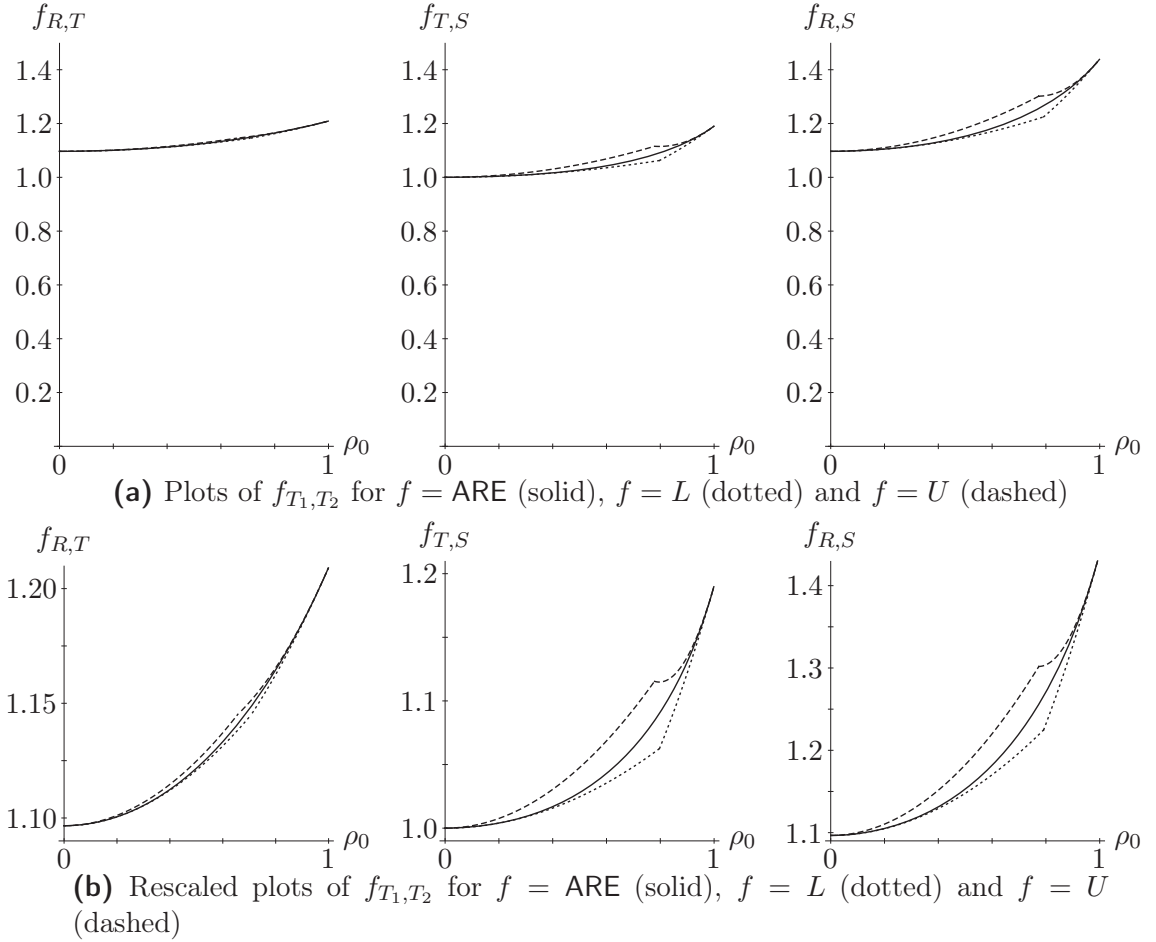


Figure 4.2: Illustration of piecewise quadratic bounds of Corollary 4.2.3

Remark 4.2.4. Numerical approximations to the various bounds in Corollary 4.2.3 are given

below.

$$\begin{aligned}
L_{R,T;0}(x) &\approx 1.0966 + 0.0966x^2; & L_{R,T;1}(x) &\approx 1.0966 - 0.0384|x| + 0.1510x^2; \\
U_{R,T;0}(x) &\approx 1.0966 + 0.1126x^2; & U_{R,T;1}(x) &\approx 1.1704 - 0.1860|x| + 0.2248x^2; \\
L_{T,S;0}(x) &\approx 1 + 0.0984x^2; & L_{T,S;1}(x) &\approx 1 - 0.3612|x| + 0.5516x^2; \\
U_{T,S;0}(x) &\approx 1 + 0.1905x^2; & U_{T,S;1}(x) &\approx 2.2684 - 2.8980|x| + 1.8200x^2; \\
L_{R,S;0}(x) &\approx 1.0966 + 0.2046x^2; & L_{R,S;1}(x) &\approx 1.0966 - 0.5254|x| + 0.8683x^2; \\
U_{R,S;0}(x) &\approx 1.0966 + 0.3429x^2; & U_{R,S;1}(x) &\approx 2.8924 - 4.1169|x| + 2.6640x^2.
\end{aligned}$$

Further, one has

$$\begin{aligned}
L_{R,T;0}(x) &= L_{R,T;1}(x) \text{ when } x \approx 0.7067; & U_{R,T;0}(x) &= U_{R,T;1}(x) \text{ when } x \approx 0.6573; \\
L_{T,S;0}(x) &= L_{T,S;1}(x) \text{ when } x \approx 0.7969; & U_{T,S;0}(x) &= U_{T,S;1}(x) \text{ when } x \approx 0.7784; \\
L_{R,S;0}(x) &= L_{R,S;1}(x) \text{ when } x \approx 0.7916; & U_{R,S;0}(x) &= U_{R,S;1}(x) \text{ when } x \approx 0.7737.
\end{aligned}$$

Remark 4.2.5. We note that piecewise quadratic bounds even tighter than the L_{T_1,T_2} and U_{T_1,T_2} could be obtained from Theorem 4.2.1. The bounds on the ARE given in Corollary 4.2.3 are derived by appropriately rewriting the inequalities $q_a(0+) < q_a(x) < q_a(1-)$ for $x \in (0,1)$ and $a \in \{0,1\}$. Of course, one may use any finite partition $0 = x_0 < x_1 < \dots < x_{n-1} < x_n = 1$ of the interval $(0,1)$ to obtain the corresponding piecewise quadratic bounds based on the inequalities $q_a(x_{i-1}+) < q_a(x) < q_a(x_i-)$ for $x \in (x_{i-1}, x_i)$, for each $i = 1, \dots, n$. We state this as another corollary, whose proof will be omitted due to its similarity to that of Corollary 4.2.3.

Corollary 4.2.6. *Let $0 = x_0 < x_1 < \dots < x_{n-1} < x_n = 1$. Then for $\text{ARE} = \text{ARE}_{T_1,T_2}$ and $q_a = q_{T_1,T_2;a}$ with $(T_1, T_2) \in \{(R, T), (T, S), (R, S)\}$ and $a \in \{0,1\}$, one has*

$$L_a \leq L \leq \text{ARE} \leq U \leq U_a$$

on $(0,1)$, where

$$L_a(x) := L_{T_1,T_2;a}(x) := \text{ARE}(a) + \text{ARE}'(a)(x-a) + q_a(x_{i-1}+)(x-a)^2$$

and

$$U_a(x) := U_{T_1,T_2;a}(x) := \text{ARE}(a) + \text{ARE}'(a)(x-a) + q_a(x_i-)(x-a)^2$$

for $x \in (x_{i-1}, x_i)$, and $L := L_0 \vee L_1$ and $U := U_0 \wedge U_1$.

Corollary 4.2.6 is illustrated by Figure 4.3; the bounds L and U are based on the partition $0 = x_0 < x_1 < x_2 = 1$, where $x_1 = (x_1)_{T_1,T_2}$ is chosen as the mean of the solutions to $L_0 = L_1$ and $U_0 = U_1$ (from Corollary 4.2.3), whose approximate values are given in Remark 4.2.4. That is, $(x_1)_{R,T} \approx \frac{1}{2}(0.7067 + 0.6573) \approx 0.6820$, $(x_1)_{T,S} \approx \frac{1}{2}(0.7969 + 0.7784) \approx 0.7876$ and $(x_1)_{R,S} \approx \frac{1}{2}(0.7916 + 0.7737) \approx 0.7826$.

Note also that Corollary 4.2.3 immediately implies even better quartic bounds on $\text{ARE}_{R,S}$:

Corollary 4.2.7. *Let*

$$\tilde{L}_{R,S;a} := L_{R,T;a} \cdot L_{T,S;a} \quad \text{and} \quad \tilde{U}_{R,S;a} := U_{R,T;a} \cdot U_{T,S;a}$$

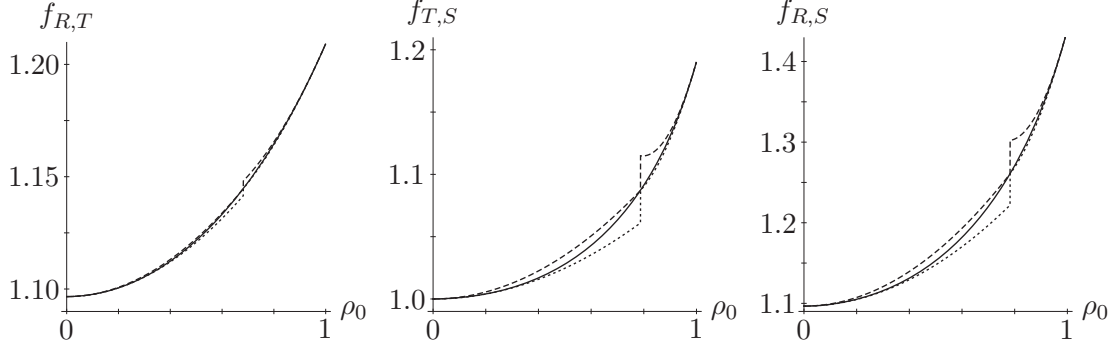


Figure 4.3: Illustration of Corollary 4.2.6, using partition $0 = x_0 < x_1 < x_2 = 1$, where $(x_1)_{R,T} \approx 0.682$, $(x_1)_{T,S} \approx 0.788$, $(x_1)_{R,S} \approx 0.783$; plots are of f_{T_1, T_2} for $f = \text{ARE}$ (solid), $f = L$ (dotted) and $f = U$ (dashed)

for $a \in \{0, 1\}$, and also let

$$\tilde{L}_{R,S} := \tilde{L}_{R,S;0} \vee \tilde{L}_{R,S;1} \quad \text{and} \quad \tilde{U}_{R,S} := \tilde{U}_{R,S;0} \wedge \tilde{U}_{R,S;1}.$$

Then

$$L_{R,S;a} < \tilde{L}_{R,S;a} < \text{ARE}_{R,S} < \tilde{U}_{R,S;a} < U_{R,S;a}$$

for $a \in \{0, 1\}$ and

$$L_{R,S} < \tilde{L}_{R,S} < \text{ARE}_{R,S} < \tilde{U}_{R,S} < U_{R,S}$$

on $(-1, 0) \cup (0, 1)$.

Consider also the test statistic commonly known as Fisher's Z variance-stabilizing transformation, where $Z := \tanh^{-1}(R)$. The utility of Z lies in the fact that its distribution converges to normality at a quicker rate than does that of R , at least when (X, Y) has the bivariate normal distribution; see e.g. the discussion in [34, 55]. We will later show that (3.5) holds when Z replaces T_j , and further that $\text{ARE}_{R,Z}(\rho_0) = 1$ for all $\rho_0 \in (-1, 1)$, which immediately yields the following:

Corollary 4.2.8. *The statements of Theorem 4.2.1 and Corollaries 4.2.2, 4.2.3, 4.2.6, and 4.2.7 all hold when R is replaced by $Z = \tanh^{-1}(R)$.*

While it is generally recommended to use R , or Z , as an estimator and test statistic when the sampled population can safely be assumed normal, departures from normality can drastically affect the distributions of either of these statistics. In [115, Section 4.3], we study the ARE when (X, Y) is the image of one of several “heavy-tail transformations” applied to a (light-tailed) bivariate normal random vector. As T and S are invariant to the transformations considered there, $\text{ARE}_{T,S}$ remains unaffected; however, $\text{ARE}_{R,T}$ (and hence $\text{ARE}_{Z,T}$) is seen to be less than 1 for most of the transformations considered there. Simulations of the relative efficiency (RE) between R , S , and T in the bivariate normal model are also discussed in [115]; the results there show that $\text{ARE} - 1$ and $\text{RE} - 1$ are estimated to

generally share the same sign, suggesting that the ARE is, at least in some instances, a suitable replacement for the RE as a comparison tool.

4.3 Proofs

We first provide Berry-Esseen bounds for the distributions of the test statistics R , S , and T and explicit expressions for the asymptotic mean and variance for each of these statistics. Once these facts are established, Theorem 4.2.1 will be proven with the aid of l'Hospital-type rules for determining the monotonicity pattern of a ratio.

4.3.1 Berry-Esseen bounds and expressions for the asymptotic means and variances of R , S , and T

Each of R , S , and T shall be shown to satisfy (4.1). For each of these statistics, it will be clear that \mathcal{V} in (4.1) may be taken to be any open interval containing ρ_0 whose closure does not contain the points -1 or 1 . Further note that each of these three statistics is invariant to linear transformations of the form $X_i \mapsto aX_i + b$ and $Y_i \mapsto cY_i + d$ with $a > 0$ and $c > 0$. So, let us assume without loss of generality (w.l.o.g.) that $\mu_X = \mu_Y = 0$ and $\sigma_X = \sigma_Y = 1$. For convenience we allow the values $\rho = \pm 1$; then the bivariate normal distribution is degenerate: $Y_i = \pm X_i$ a.s.

Our recent result [116, Theorem 4.17] shows that (4.1) holds with R in place of T_j when

$$\mu_R(\rho) := \rho \quad (4.6)$$

and

$$\sigma_R^2(\rho) := \mathbb{E}_\rho(X_1 Y_1 - \frac{\rho}{2}(X_1^2 + Y_1^2))^2 = (1 - \rho^2)^2; \quad (4.7)$$

note that an explicit expression for \mathfrak{C} there is found by an application of [116, Theorem 4.1], from which it is seen that \mathfrak{C} can be bounded independently of $\rho \in \mathcal{V}$. Note also that [116, Theorem 4.1] can be used to show that (4.1) holds with Z in place of T_j when we let

$$\mu_Z(\rho) = \tanh^{-1}(\rho) \quad \text{and} \quad \sigma_Z^2(\rho) = 1.$$

Then $\sigma_R^2/(\mu'_R)^2 = \sigma_Z^2/(\mu'_Z)^2$, so that $\text{ARE}_{R,Z}(\rho_0) = 1$ for all $\rho_0 \in (-1, 1)$ follows from (4.2), which proves Corollary 4.2.8.

By (4.5), T is a U -statistic with kernel h_T of degree $m = 2$. Further, S (defined in (4.4)) is a V -statistic with a kernel of degree $m = 3$; Hoeffding [51, Section 5c] describes how any V -statistic can be expressed as a U -statistic of the same degree, so that S is a U -statistic with a symmetric kernel $h_{S,n}$ of degree $m = 3$. Namely,

$$S = \frac{1}{\binom{n}{3}} \sum_{1 \leq i < j < k \leq n} h_{S,n}(V_i, V_j, V_k),$$

where

$$h_{S,n}(V_i, V_j, V_k) := \frac{n-2}{n+1} h_S(V_i, V_j, V_k) + \frac{1}{n+1} (h_T(V_i, V_j) + h_T(V_i, V_k) + h_T(V_j, V_k)), \quad (4.8)$$

$$h_S := 2(K_{ijk} + K_{ikj} + K_{jik} + K_{jki} + K_{kij} + K_{kji}) - 3 \quad (4.9)$$

and

$$K_{ijk} := \mathbb{I}\{X_j < X_i\} \mathbb{I}\{Y_k < Y_i\}.$$

It follows by Chen and Shao's result [19, (3.4) in Theorem 3.1] that for $m \in \{2, 3\}$ and $n \geq m$

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P} \left(\frac{U - \mathbb{E} U}{\sigma_1 / \sqrt{n}} \leq z \right) - \Phi(z) \right| \leq \frac{A}{\sigma_1^3 \sqrt{n}}, \quad (4.10)$$

for some absolute constant A , where $U = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} h(V_{i_1}, \dots, V_{i_m})$ is any U -statistic with a bounded symmetric kernel h ,

$$\sigma_1^2 := m^2 \text{Var } g(V_1) > 0,$$

and

$$g(V_1) := \mathbb{E}[h(V_1, \dots, V_m) | V_1].$$

Now consider T as expressed in (4.5), and recall that $|h_T| = 1$. One also has

$$\mu_T(\rho) := \mathbb{E}_\rho T = \mathbb{E}_\rho h_T(V_1, V_2) = 2 \mathbb{E}_\rho (J_{12} + J_{21}) - 1 = 4 \mathbb{E}_\rho J_{12} - 1 = \frac{2}{\pi} \sin^{-1} \rho. \quad (4.11)$$

In order to see this, note that $\mathbb{E}_\rho J_{12} = \mathbb{P}_\rho(X_1 - X_2 > 0, Y_1 - Y_2 > 0) = \mathbb{P}(Z_1 > 0, \rho Z_1 + \sqrt{1 - \rho^2} Z_2 > 0)$, where Z_1 and Z_2 are independent standard normal r.v.'s. By the circular symmetry of the distribution of (Z_1, Z_2) on the plane, we see $\mathbb{E}_\rho J_{12}$ is simply the proportion of the length of the arc of the unit circle between the points $(0, 1)$ and $(\sqrt{1 - \rho^2}, -\rho)$; that is,

$$\mathbb{E}_\rho J_{12} = \frac{1}{2\pi} \left(\frac{\pi}{2} - \sin^{-1}(-\rho) \right) = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho,$$

whence (4.11).

One can use a similar geometric reasoning to obtain an expression for the asymptotic variance of T . Let

$$g_T(V_1) := \mathbb{E}[h_T(V_1, V_2) | V_1] = 2 \mathbb{E}[J_{12} + J_{21} | V_1] - 1,$$

so that

$$\sigma_T^2(\rho) := 4 \text{Var}_\rho g_T(V_1) = 16 (\mathbb{E}_\rho J_{12} J_{13} + 2 \mathbb{E}_\rho J_{12} J_{31} + \mathbb{E}_\rho J_{21} J_{31} - 4 [\mathbb{E}_\rho J_{12}]^2).$$

Consider first $\mathbb{E}_\rho J_{12} J_{13} = \mathbb{P}(U_1 > 0, U_2 > 0, U_3 > 0, U_4 > 0)$, where the U_i 's are standard normal r.v.'s with

$$\Sigma := \text{Cov} \begin{bmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} & \rho & \frac{\rho}{2} \\ \frac{1}{2} & 1 & \frac{\rho}{2} & \rho \\ \rho & \frac{\rho}{2} & 1 & \frac{1}{2} \\ \frac{\rho}{2} & \rho & \frac{1}{2} & 1 \end{bmatrix}.$$

That is, $\mathbb{E}_\rho J_{12} J_{13}$ is the probability that the random point $\Sigma^{1/2} [Z_1, Z_2, Z_3, Z_4]^T$ lies in the first orthant of 4-dimensional space, where the Z_i 's are independent standard normal r.v.'s; further, this is simply the ratio of the volume $V(\rho)$ of the spherical tetrahedron $A_1 A_2 A_3 A_4$ to

the volume $2\pi^2$ of the unit sphere $S_3 := \{x \in \mathbb{R}^4 : \|x\| = 1\}$, where the vertices A_1, A_2, A_3, A_4 of the tetrahedron are the columns of $\Sigma^{-1/2}$ normalized to be unit vectors. One can use the classical result of Schläfli [131] to obtain the volume of this spherical tetrahedron. But, in fact, this work has been indirectly done by David and Mallows in their derivation of the variance of S ; the probabilities $\mathbb{E}_\rho J_{12}J_{13}$ and $\mathbb{E}_\rho J_{12}J_{31}$ correspond to correlation matrices (r) and (w), respectively, in Appendix 2 of [27]. Using the formulas there, and noting $\mathbb{E}_\rho J_{21}J_{31} = \mathbb{E}_\rho J_{12}J_{13}$ by the symmetry of the normal distribution, one sees

$$\sigma_T^2(\rho) = \frac{4}{9} - \frac{16}{\pi^2} \left(\sin^{-1} \frac{\rho}{2} \right)^2, \quad (4.12)$$

which is bounded away from 0 over any closed subinterval of $(-1, 1)$, so, by (4.10), one has (4.1) for any $\theta_0 = \rho_0 \in (-1, 1)$ when $T_j = T$.

We remark that Kendall's monograph [67, Chapter 10] contains derivations of (4.11) and (4.12). Further, Plackett [121] has obtained a more general method for calculating $\mathbb{P}(U_1 > a_1, U_2 > a_2, U_3 > a_3, U_4 > a_4)$ which reduces to the Schläfli method when the a_i are all 0.

Directing attention to S , first note that $h_{S,n}$ is bounded (in fact, one can check that $|h_{S,n}(V_1, V_2, V_3)| \in \{1, \frac{n-1}{n+1}\}$ a.s.). Using geometric reasoning similar to that used to compute $\mathbb{E}_\rho J_{12}$ (only now using the fact that $X_1 - X_2$ and $Y_1 - Y_3$ have a correlation of $\frac{\rho}{2}$), one finds

$$\mathbb{E}_\rho K_{123} = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \frac{\rho}{2},$$

so that

$$\mu_{S,n}(\rho) := \mathbb{E}_\rho S = \mathbb{E}_\rho h_{S,n}(V_1, V_2, V_3) = \frac{n-2}{n+1} \frac{6}{\pi} \sin^{-1} \frac{\rho}{2} + \frac{3\mu_T(\rho)}{n+1};$$

accordingly, let

$$\mu_S(\rho) := \lim_{n \rightarrow \infty} \mu_{S,n}(\rho) = \mathbb{E}_\rho h_S(V_1, V_2, V_3) = \frac{6}{\pi} \sin^{-1} \frac{\rho}{2} \quad (4.13)$$

and note that $\sqrt{n}(\mu_{S,n} - \mu_S) \xrightarrow[n \rightarrow \infty]{} 0$ uniformly for $\rho \in (-1, 1)$.

Let next

$$\begin{aligned} g_{S,n}(V_1) &:= \mathbb{E}[h_{S,n}(V_1, V_2, V_3)|V_1]; \\ g_S(V_1) &:= \mathbb{E}[h_S(V_1, V_2, V_3)|V_1] = 4\mathbb{E}[K_{123} + K_{213} + K_{231}|V_1] - 3; \\ \sigma_{S,n}^2(\rho) &:= 9\text{Var}_\rho g_{S,n}(V_1); \\ \sigma_S^2(\rho) &:= 9\text{Var}_\rho g_S(V_1) \\ &= 144(\mathbb{E}_\rho K_{123}K_{145} + 2\mathbb{E}_\rho K_{213}K_{415} + 4\mathbb{E}_\rho K_{123}K_{415} + 2\mathbb{E}_\rho K_{213}K_{451} \\ &\quad - 9[\mathbb{E}_\rho K_{123}]^2), \end{aligned}$$

where $\sigma_{S,n}(\rho) := \sqrt{\sigma_{S,n}^2(\rho)}$ and $\sigma_S(\rho) := \sqrt{\sigma_S^2(\rho)}$, noting that

$$\mathbb{E}_\rho K_{231}K_{451} = \mathbb{E}_\rho K_{213}K_{415}$$

and

$$\mathbb{E}_\rho K_{123}K_{451} = \mathbb{E}_\rho K_{132}K_{415} = \mathbb{E}_\rho K_{123}K_{415}$$

since the distributions of $((X_1, Y_1), \dots, (X_n, Y_n))$ and $((Y_1, X_1), \dots, (Y_n, X_n))$ are identical and permutation-invariant. It is clear that expressions for $\sigma_{S,n}^2$ and σ_S^2 may be derived in terms of the volumes of spherical tetrahedra via Schläfli's formula. For the sake of brevity, we omit these details and refer the reader to David and Mallows' derivation of $\text{Var } S$; note the probabilities $\mathbb{E}_\rho K_{123}K_{145}$, $\mathbb{E}_\rho K_{123}K_{415}$, $\mathbb{E}_\rho K_{213}K_{415}$, $\mathbb{E}_\rho K_{213}K_{451}$ correspond to the correlation matrices (c), (d), (e), and (f), respectively, found in Appendix 2 of [27]. Then one has

$$\sigma_S^2(\rho) = 1 - \frac{324}{\pi^2} (\sin^{-1} \frac{\rho}{2})^2 + \frac{72}{\pi^2} (I_1(\rho) + 2I_2(\rho) + 2I_3(\rho) + 4I_4(\rho)), \quad (4.14)$$

where

$$\begin{aligned} I_1(x) &:= \int_0^x \frac{\sin^{-1} \frac{u^3}{4(2-u^2)}}{\sqrt{4-u^2}} du, & I_2(x) &:= \int_0^x \frac{\sin^{-1} \frac{u}{2(3-u^2)}}{\sqrt{4-u^2}} du, \\ I_3(x) &:= \int_0^x \frac{\sin^{-1} \frac{u(4-u^2)}{2\sqrt{2}\sqrt{8-6u^2+u^4}}}{\sqrt{4-u^2}} du, & I_4(x) &:= \int_0^x \frac{\sin^{-1} \frac{u(4-u^2)}{2\sqrt{12-7u^2+u^4}}}{\sqrt{4-u^2}} du; \end{aligned}$$

an explicit expression of $\sigma_{S,n}^2$ is not of direct concern to us and so is omitted (though could also be obtained from [27]). Note the integrals I_1, \dots, I_4 are expressed differently than the corresponding ones found in [27], though a simple change of variables shows their equivalence.

Now, $\sigma_{S,n} \xrightarrow[n \rightarrow \infty]{} \sigma_S$ uniformly over all $\rho \in [-1, 1]$ (since, by (4.8), $h_{S,n} - h_S = O(1/n)$). It will be pointed out in the last paragraph of part (TS0) of the proof of Theorem 4.2.1 that $\sigma_S^2 > 0$ for $\rho \in (-1, 1)$. It is also clear from (4.14) that σ_S^2 is a continuous function of ρ , so that the minimum of σ_S over any closed subinterval of $(-1, 1)$ is strictly positive. Thus, $\inf_{\rho \in \mathcal{V}} \sigma_{S,n}(\rho) > 0$ for all large enough n , where \mathcal{V} is as introduced in the beginning of Section 4.3.1. Referring now to (4.10) (and replacing there U with S , $\mathbb{E}U$ with $\mu_{S,n}$ and σ_1 with $\sigma_{S,n}$), one finds that

$$\begin{aligned} & \sup_{z \in \mathbb{R}} \left| \mathbb{P}_\rho \left(\frac{S - \mu_S(\rho)}{\sigma_S(\rho)/\sqrt{n}} \leq z \right) - \Phi(z) \right| \\ & \leq \frac{A}{\sigma_{S,n}^3(\rho)\sqrt{n}} + \left| \Phi(z^*) - \Phi\left(\frac{\sigma_S(\rho)}{\sigma_{S,n}(\rho)} z\right) \right| + \left| \Phi\left(\frac{\sigma_S(\rho)}{\sigma_{S,n}(\rho)} z\right) - \Phi(z) \right|, \end{aligned}$$

where $z^* = \frac{\sigma_S(\rho)}{\sigma_{S,n}(\rho)} \left(z + \frac{\mu_S(\rho) - \mu_{S,n}(\rho)}{\sigma_S(\rho)/\sqrt{n}} \right)$; in turn, the last two terms in the above inequality vanish uniformly over $z \in \mathbb{R}$ and $\rho \in \mathcal{V}$ as n tends to ∞ (using well-known properties of the function Φ and the previously noted facts that $\sqrt{n}(\mu_S - \mu_{S,n}) \rightarrow 0$ and $\sigma_{S,n}/\sigma_S \rightarrow 1$ uniformly on \mathcal{V}), so that S satisfies (4.1).

The next result will be used in the proofs of the statements (TS0) – (RS1) in Theorem 4.2.1:

Lemma 4.3.1. *One has $\sigma_S^2(1-) = 0$.*

Proof. W.l.o.g., $Y_i = \rho X_i + \sqrt{1 - \rho^2} Z_i$ for all i , where the Z_i 's are i.i.d. $N(0, 1)$ r.v.'s

independent of the X_i 's. Further note that $\sigma_{S,n}^2(\rho)$ differs only by a positive constant factor from $\text{Var}_\rho \text{proj}_{\mathcal{L}} S$, where \mathcal{L} is the space of all linear statistics. Also, for $\rho = 1$, one has $S = 1$ a.s. and hence $\text{Var}_\rho \text{proj}_{\mathcal{L}} S \leq \text{Var}_\rho S = 0$, so $\sigma_{S,n}^2(1) = 0$ for all n . Now, letting $n \rightarrow \infty$, one has $\sigma_S^2(1) = 0$, since $h_{S,n} - h_S = O(1/n)$.

Next, $\frac{1}{9}\sigma_S^2(\rho) = \text{Var}_\rho g_S(V_1) = \mathbb{E}_\rho h_S(W_1, W_2, W_3)h_S(W_1, W_4, W_5) - \mathbb{E}_\rho^2 h_S(W_1, W_2, W_3)$, with $W_i := W_i(\rho) := (X_i, \rho X_i + \sqrt{1 - \rho^2} Z_i)$. Next, $h_S(W_1, W_2, W_3)$ and $h_S(W_1, W_4, W_5)$ are continuous in ρ on the complement of the union of all events of the form $\{X_i = X_j\}$ for $i \neq j$. The latter union has zero probability. So, by dominated convergence, $\sigma_S^2(\rho) \rightarrow \sigma_S^2(1) = 0$ as $\rho \uparrow 1$. \square

While the result of this last lemma should not be surprising, it should be noted that trying to assert $\sigma_S^2(1-) = 0$ using only the expression (4.14) is a more difficult task.

4.3.2 Proofs of monotonicity

As in [101, 102, 103], let $-\infty \leq a < b \leq \infty$, and suppose that f and g are differentiable functions on (a, b) . Let $r := \frac{f}{g}$ and $\rho := \frac{f'}{g'}$; from hereon, the symbol ρ should not be considered the correlation of a bivariate normal population, which latter will be denoted by x . Assume that either $g < 0$ or $g > 0$ on (a, b) , and also that $g' < 0$ or $g' > 0$ on (a, b) . For an arbitrary function h defined on (a, b) , adopt the notation “ $h \nearrow$ ” to mean h is (strictly) increasing on (a, b) and similarly let “ $h \searrow$ ” mean h is decreasing on (a, b) ; the juxtaposition of these arrows shall have the obvious meaning, e.g. “ $h \nearrow \searrow$ ” means that there exists some $c \in (a, b)$ such that $h \nearrow$ on (a, c) and $h \searrow$ on (c, b) . Further, let the notation “ h is $+-$ ” mean that there exists $c \in (a, b)$ such that $h > 0$ on (a, c) and $h < 0$ on (c, b) ; similar meaning will be given to other such strings composed of alternating “ $+$ ” and “ $-$ ” symbols.

Special-case rules (Proposition 4.1 of [103]).

Suppose that either $f(a+) = g(a+) = 0$ or $f(b-) = g(b-) = 0$.

- (i) If $\rho \nearrow$ on (a, b) , then $r' > 0$ on (a, b) and hence $r \nearrow$ on (a, b) ;
- (ii) If $\rho \searrow$ on (a, b) , then $r' < 0$ on (a, b) and hence $r \searrow$ on (a, b) .

General rules (Corollary 3.1 of [103]).

- (i) If $\rho \nearrow$ and $gg' > 0$ on (a, b) , then $r \searrow$, $r \nearrow$ or $r \searrow \nearrow$ on (a, b) ;
- (ii) If $\rho \nearrow$ and $gg' < 0$ on (a, b) , then $r \searrow$, $r \nearrow$ or $r \nearrow \searrow$ on (a, b) ;
- (iii) If $\rho \searrow$ and $gg' > 0$ on (a, b) , then $r \searrow$, $r \nearrow$ or $r \nearrow \searrow$ on (a, b) ;
- (iv) If $\rho \searrow$ and $gg' < 0$ on (a, b) , then $r \searrow$, $r \nearrow$ or $r \searrow \nearrow$ on (a, b) .

Refined general rules (Corollary 3.2 of [103]).

Let $\tilde{\rho} := g^2 \frac{r'}{|g'|} = \text{sgn}(g')(\rho g - f)$.

- (i) If $\rho \nearrow$ and $gg' > 0$ on (a, b) , then $\tilde{\rho} \nearrow$;
- (ii) If $\rho \nearrow$ and $gg' < 0$ on (a, b) , then $\tilde{\rho} \searrow$;

(iii) If $\rho \searrow$ and $gg' > 0$ on (a, b) , then $\tilde{\rho} \searrow$;

(iv) If $\rho \searrow$ and $gg' < 0$ on (a, b) , then $\tilde{\rho} \nearrow$.

In addition, $\text{sgn}(\tilde{\rho}) = \text{sgn}(r')$, so that the monotonicity pattern of r may be determined by the monotonicity of $\tilde{\rho}$ and knowledge of the signs of $\tilde{\rho}(a+)$ and/or $\tilde{\rho}(b-)$.

E.g. suppose it can be established that $\rho \nearrow$ and $gg' > 0$ on (a, b) ; if one also knows that $r(a+) = -\infty$ then the general rules imply $r \nearrow$. Alternatively, $\rho \nearrow$ and $gg' > 0$ imply $\tilde{\rho} \nearrow$; if it can be established that $\tilde{\rho}(a+) \geq 0$, then $\tilde{\rho} > 0$ on (a, b) and hence $r \nearrow$ on (a, b) . We shall make frequent use of these rules throughout the proof of Theorem 4.2.1. The special-case rules are proved in [102, Proposition 1.1], and a proof of the general rules is found in [101, Proposition 1.9]. A proof of the refined general rules, along with several other variants of these monotonicity rules, is found in [103, Lemma 2.1]. Note that Anderson et al. [2, Lemma 2.2] proved a variant of the special-case rules, wherein the function $\frac{f(x)-f(a)}{g(x)-g(a)}$ \nearrow (or \searrow) whenever $\rho \nearrow$ (or \searrow).

That (4.2) may be used to express any of the three pairwise ARE's has been justified by the work of the previous section. The proofs of the six statements (RT0) – (RS1) in Theorem 4.2.1 will follow the same general method. Fix an arbitrary $a \in [0, 1]$, and let

$$b := b(a) := \text{ARE}(a) \quad \text{and} \quad c := c(a) := \text{ARE}'(a). \quad (4.15)$$

Then

$$q_a(x) = \frac{\text{ARE}(x) - b - c(x-a)}{(x-a)^2} = \frac{f(x) - bg(x) - c(x-a)g(x)}{(x-a)^2g(x)}$$

when f and g are functions chosen so that $\text{ARE} = \frac{f}{g}$. Accordingly, let

$$f_0(x) := f(x) - bg(x) - c(x-a)g(x), \quad g_0(x) := (x-a)^2g(x), \quad r_0 := \frac{f_0}{g_0} = q_a,$$

and also

$$f_i := a_i f'_{i-1}, \quad g_i := a_i g'_{i-1}, \quad r_i := \frac{f_i}{g_i}, \quad \rho_{i-1} := \frac{f'_{i-1}}{g'_{i-1}} = r_i, \quad \text{and} \quad \tilde{\rho}_i = \text{sgn}(g_{i+1})(r_{i+1}g_i - f_i) \quad (4.16)$$

where the a_i are positive on $(0, 1)$. There is some freedom in choosing the functions a_i , though the goal is to ensure that, for some natural number $n \geq 1$, the ratio r_n is an algebraic function. In our case it will turn out that r_n is actually an algebraic function independent of the value of a . As r_n is algebraic, the problem of determining its monotonicity pattern on an interval is completely algorithmic (cf. [140, 22]); here, we use the Mathematica **Reduce** command to deduce the monotonicity of $r_n = \rho_{n-1}$. The specific choices of f , g and the a_i are given in Lemmas 4.3.2 – 4.3.4 below. One may refer to this first phase of the proof as the “reduction” phase.

Once the monotonicity of $r_n = \rho_{n-1}$ is established, the second and final stage of the proof is to “work backwards” by using the various l’Hospital-type rules stated above to deduce the monotonicity patterns of $r_{n-1} = \rho_{n-2}$, $r_{n-2} = \rho_{n-3}$, \dots , $r_1 = \rho_0$, $r_0 = q_a$. Throughout the proof, all functions shall be assumed to be defined on $(0, 1)$ unless otherwise stated.

As most of the functions being treated are rather unwieldy, all calculations are performed with the Mathematica (v. 7.0 or later) software. Detailed output from the notebooks we used in the following proofs are found in the appendices of the version of this paper [82] found on the pre-print server <http://arxiv.org>. Each of the appendices RT, TS, and RS follows the same general format: the first section (labeled RTr, TSr, or RSr – where “r” stands for “reduction (phase)”) is dedicated to proving one of the corresponding Lemmas 4.3.2–4.3.4 below (i.e., the “reduction” stage of the proofs), the second section (RT0, TS0, or RS0) provides numerical support for proving the monotonicity of q_0 , and the third section (RT1, TS1, or RS1) provides support for proving the monotonicity of q_1 .

We prove q_a is increasing only for $a \in \{0, 1\}$; the following three lemmas could perhaps be used as starting points for the “working backwards” phase for other choices of $a \in (0, 1)$ to get even more quadratic bounds on the ARE’s (cf. Corollary 4.2.3). It is of course desirable to demonstrate that $q_a \nearrow$ for arbitrary $a \in [0, 1]$ (should this be true), though a proof of such a statement has yet to be found; for any given $a \in (0, 1)$, this second phase of the proof is restricted only by computational capacities, since, as mentioned above, the expression for r_n is eventually algebraic. We remark also that this method could conceivably be adapted (by using an appropriate variant of the definition of q_a) to finding quadratic bounds on $\text{ARE}_{T,R} = 1/\text{ARE}_{R,T}$, $\text{ARE}_{S,T} = 1/\text{ARE}_{T,S}$, and $\text{ARE}_{S,R} = 1/\text{ARE}_{R,S}$, or possibly finding approximating polynomials of degree greater than 2.

Lemma 4.3.2. *Let $a \in [0, 1]$ be arbitrary, and let*

$$f(x) := \pi^2 - 36(\sin^{-1} \frac{x}{2})^2 \quad \text{and} \quad g(x) := 9(1 - x^2)$$

for $x \in (0, 1)$. Further let

$$\begin{aligned} a_1(x) &:= \sqrt{4 - x^2}, & a_2(x) &:= \frac{\sqrt{4 - x^2}}{2 - x^2}, \\ a_3(x) &:= \frac{(2 - x^2)^2}{50 - 29x^2 + 9x^4}, & \text{and } a_4(x) &:= \frac{(50 - 29x^2 + 9x^4)^2}{2 - x^2}. \end{aligned}$$

Then on the interval $(0, 1)$, one has $\text{ARE}_{R,T} = \frac{f}{g}$, $a_i > 0$ for $i = 1, \dots, 4$, $r_4 \nearrow$, $f_4 < 0$ and $g_4 < 0$, where and f_i , g_i , r_i are as defined in (4.16).

Lemma 4.3.3. *Let $a \in [0, 1]$ be arbitrary, and let*

$$f(x) := \sigma_S^2(x) \quad \text{and} \quad g(x) := \frac{4(1 - x^2)(\pi^2 - 36(\sin^{-1} \frac{x}{2})^2)}{\pi^2(4 - x^2)}, \quad (4.17)$$

where σ_S^2 is given in (4.14). Further let

$$\begin{aligned}
h_1(x) &:= 4 - x^2, \\
h_2(x) &:= 2 + x^2, \\
h_3(x) &:= 38 - 17x^2 - 3x^4, \\
h_4(x) &:= x(892 - 440x^2 + 61x^4 - 9x^6), \\
h_5(x) &:= 328256 - 60276x^2 - 28380x^4 + 12853x^6 - 678x^8 + 81x^{10}, \\
h_6(x) &:= 17418976 - 12356932x^2 + 3290736x^4 - 575137x^6 + 35011x^8 - 447x^{10} + 81x^{12}, \\
h_7(x) &:= 18745083424 - 14666397812x^2 + 4272900412x^4 - 473552785x^6 + 47852540x^8 \\
&\quad - 89482x^{10} + 1296x^{12} - 729x^{14}, \\
h_8(x) &:= 67393220864 - 66665518536x^2 + 25281966744x^4 - 4783210446x^6 + 320370996x^8 \\
&\quad - 26281941x^{10} - 170777x^{12} + 231x^{14} + 81x^{16}, \\
h_9(x) &:= 32482389470208 - 34864017237408x^2 + 16286313144464x^4 - 4430399397672x^6 \\
&\quad + 832485830428x^8 - 100457826796x^{10} + 7855470828x^{12} - 362114966x^{14} \\
&\quad + 14054393x^{16} + 127203x^{18} + 31x^{20} - 9x^{22},
\end{aligned}$$

and also

$$\begin{aligned}
a_1 &:= h_1^{1/2}, & a_2 &:= \frac{h_1^{5/2}}{h_2}, & a_3 &:= \frac{h_2^2}{h_1 h_3}, & a_4 &:= \frac{h_1^{5/2} h_3^2}{h_2 h_4}, & a_5 &:= \frac{h_4^2}{h_1 h_3 h_5}, \\
a_6 &:= \frac{h_1^{1/2} h_5^2}{h_4 h_6}, & a_7 &:= \frac{h_6^2}{h_5 h_7}, & a_8 &:= \frac{h_7^2}{h_6 h_8}, & a_9 &:= \frac{h_1^{5/2} h_8^2}{h_7 h_9}, & a_{10} &:= \frac{\pi^2 h_9^2}{27648 h_1^{3/2} h_8}.
\end{aligned}$$

Then on the interval $(0, 1)$, one has $\text{ARE}_{T,S} = \frac{f}{g}$, $a_i > 0$ for $i = 1, \dots, 10$, $r_{10} \nearrow$, $f_{10} > 0$ and $g_{10} > 0$, where f_i , g_i , and r_i are as defined in (4.16).

Lemma 4.3.4. Let $a \in [0, 1]$ be arbitrary, and let

$$f(x) := \sigma_S^2(x) \quad \text{and} \quad g(x) := \frac{36(1 - x^2)^2}{\pi^2(4 - x^2)},$$

where σ_S^2 is given in (4.14). Further let

$$\begin{aligned}
h_1(x) &:= 4 - x^2, \\
h_3(x) &:= x(41 - 20x^2 + 3x^4), \\
h_4(x) &:= 7052 + 30147x^2 - 35490x^4 + 13432x^6 - 2370x^8 + 189x^{10}, \\
h_5(x) &:= x,
\end{aligned}$$

and also

$$a_1 := h_1^{1/2}, \quad a_2 := h_1^{5/2}, \quad a_3 := \frac{1}{h_3}, \quad a_4 := \frac{h_3^2}{h_4}, \quad a_5 := \frac{\pi^2 h_4^2}{1728 h_3 h_5}.$$

Then on the interval $(0, 1)$, one has $\text{ARE}_{R,S} = \frac{f}{g}$, $a_i > 0$ for $i = 1, \dots, 5$, $r_5 \nearrow$, $f_5 > 0$, and $g_5 > 0$, where f_i , g_i , and r_i are as defined in (4.16).

The proofs of Lemmas 4.3.2–4.3.4 are omitted here; again, the appendices of the online version of this paper [82] contain the details for these proofs.

Before proving Theorem 4.2.1, recall the implications of (4.16). If on some open subinterval of $(0, 1)$ one has $f_i > 0$ (or $f_i < 0$), then on this subinterval $f_{i-1} \nearrow$ (or $f_{i-1} \searrow$), and similarly for the g_i 's. If g_i has k roots in $(0, 1)$, these shall be denoted by $x_{i,j}$, $j = 1, \dots, k$, with the assumption that $x_{i,1} < \dots < x_{i,k}$; if g_i has only a single root in $(0, 1)$, it will simply be denoted by x_i . Similarly, the roots of f_i whenever they exist will be denoted by $y_{i,1}, y_{i,2}, \dots$ (or simply y_i if f_i has a single root), and if ever r'_i is shown to have a root in $(0, 1)$ (there will only be at most one root in what follows), this root will be denoted by z_i . Numerical approximations of any of these roots are not of direct concern to us, but rather their positions relative to other roots. Such information is easily obtained from evaluation of the respective functions at specific points; for instance, if at some step we deduce that f_1 and g_1 are both $+-$, with $f_1(0.5) > 0 > g_1(0.5)$, then it is inferred that $x_1 < 0.5 < y_1$ (and further, that $r_1(x_1-) = \frac{f_1}{g_1}(x_1-) = \infty$ and $r_1(x_1+) = \frac{f_1}{g_1}(x_1+) = -\infty$).

Proof of Theorem 4.2.1, (RT0). Adopt the notation of Lemma 4.3.2, with $a = 0$, so that, in accordance with (4.15),

$$b = \text{ARE}_{R,T}(0) = \frac{\pi^2}{9} \quad \text{and} \quad c = \text{ARE}'_{R,T}(0) = 0.$$

Noting that $f_3(0+) = g_3(0+) = 0$, one has $f_3 < 0$, $g_3 < 0$ (since, by Lemma 4.3.2, $f_4 < 0$ and $g_4 < 0$), and also, by the special-case rules, $\rho_2 = r_3 \nearrow$ (since, by Lemma 4.3.2, $\rho_3 = r_4 \nearrow$).

Next, $g_2 \searrow$ (as $g_3 < 0$) and $g_2(0+) > 0 > g_2(1-)$ imply $g_2 > 0$ on $(0, x_2)$ and $g_2 < 0$ on $(x_2, 1)$; similarly, $f_2 \searrow$ and $f_2(0+) > 0 > f_2(1-)$ imply $f_2 > 0$ on $(0, y_2)$ and $f_2 < 0$ on $(y_2, 1)$. Verifying that $g_2(0.41) < 0 < f_2(0.41)$, one has $x_2 < 0.41 < y_2$, further implying $r_2(x_2-) = \infty$ and $r_2(x_2+) = -\infty$. Noting the sign of $g_2 g'_2$ (which is the sign of $g_2 g_3$) on each of $(0, x_2)$ and $(x_2, 1)$, the general rules imply $\rho_1 = r_2 \nearrow$ on each of these two intervals.

Next, $g_1 \nearrow \searrow$ on $(0, 1)$ (as g_2 is $+-$) and $g_1(0+) = 0 > g_1(1-)$ imply the existence of a single root x_1 , with $x_2 < x_1$; similarly, $f_1 \nearrow \searrow$ and $f_1(0+) = 0 > f_1(1-)$ imply the existence of a single root y_1 , with $y_2 < y_1$. The special-case rules imply $r_1 \nearrow$ on $(0, x_2)$ (as $f_1(0+) = g_1(0+) = 0$). Further, $g_1(0.71) < 0 < f_1(0.71)$ implies $x_1 < y_1$, which in turn shows $r_1(x_1-) = \infty$ and $r_1(x_1+) = -\infty$; noting the sign of $g_1 g'_1$ on each of the intervals (x_2, x_1) and $(x_1, 1)$, the general rules imply $r_1 \nearrow$ on these two intervals. The continuity of r_1 at x_2 implies $\rho_0 = r_1 \nearrow$ on $(0, x_1)$ and $(x_1, 1)$.

Finally, $f_0(0+) = g_0(0+) = f_0(1-) = g_0(1-) = 0$ imply both $g_0 > 0$ on $(0, 1)$ (since g_1 is $+-$ and hence $g_0 \nearrow \searrow$ on $(0, 1)$) and $r_0 \nearrow$ on each of the intervals $(0, x_1)$ and $(x_1, 1)$ (by the special-case rules); the continuity of r_0 at x_1 implies $q_{R,T;0} = r_0 \nearrow$ on $(0, 1)$. Further, the l'Hospital rule for limits implies $r_0(0+) = r_2(0+)$ and $r_0(1-) = r_1(1-)$. \square

Proof of Theorem 4.2.1, (RT1). Adopt the notation of Lemma 4.3.2, with $a = 1$, so that

$$b = \text{ARE}_{R,T}(1-) = \frac{f'(1-)}{g'(1-)} = \frac{2\pi}{3\sqrt{3}}$$

and

$$c = \text{ARE}'_{R,T}(1-) = \frac{f''(1-)g'(1-) - f'(1-)g''(1-)}{2g'(1-)^2} = \frac{2(9 - \sqrt{3}\pi)}{27},$$

which follows by repeated application of the l'Hospital rule for limits after noting $f(1-) = g(1-) = 0$.

Next, $g_3(0+) > 0 > g_3(1-)$ and $f_3(0+) > 0 > f_3(1-)$ along with $g_3 \searrow$ and $f_3 \searrow$ (since $f_4 < 0$ and $g_4 < 0$ by Lemma 4.3.2) shows that g_3 and f_3 each have a single root x_3 and y_3 , respectively. Also, $g_3(0.6) < 0 < f_3(0.6)$ shows $x_3 < y_3$ and hence $r_3(x_3-) = \infty$ and $r_3(x_3+) = -\infty$. Noting the sign of $g_3g'_3$ on each of the intervals $(0, x_3)$ and $(x_3, 1)$, the general rules imply $\rho_2 = r_3 \nearrow$ on these two intervals.

Next, $g_2 \nearrow \searrow$ (as g_3 is $+-$) and $g_2(0+) = g_2(1-) = 0$ imply $g_2 > 0$, whereas $f_2 \nearrow \searrow$ and $f_2(0+) < 0 = f_2(1-)$ imply f_2 has a single root y_2 . The special-case rules imply $r_2 \nearrow$ on $(x_3, 1)$; as $\rho_2 \nearrow$ and $g_2g'_2 > 0$ on $(0, x_3)$ and $\tilde{\rho}_2(0+) > 0$, the refined general rules imply $\tilde{\rho}_2 > 0$ and hence $r_2 \nearrow$ on $(0, x_3)$. Noting that r_2 is continuous at x_3 , one has $\rho_1 = r_2 \nearrow$ on $(0, 1)$.

Next, $g_1 \nearrow$ and $f_1(1-) = g_1(1-) = 0$ imply both $g_1 < 0$ and $\rho_0 = r_1 \nearrow$ on $(0, 1)$; similarly, $g_0 \searrow$ and $f_0(1-) = g_0(1-) = 0$ imply $g_0 > 0$ and $q_{R,T;1} = r_0 \nearrow$ on $(0, 1)$. Lastly, $r_0(0+) = \frac{f_0(0+)}{g_0(0+)}$ and also $r_0(1-) = r_3(1-)$, which follows by the l'Hospital rule for limits. \square

Proof of Theorem 4.2.1, (TS0). Adopt the notation of Lemma 4.3.3, with $a = 0$, so that

$$b = \text{ARE}_{T,S}(0) = 1 \quad \text{and} \quad c = \text{ARE}'_{T,S}(0) = 0.$$

Now, $g_9 \nearrow$, $f_9 \nearrow$, and $f_9(0+) = g_9(0+) = 0$ imply $f_9 > 0$, $g_9 > 0$, and $\rho_8 = r_9 \nearrow$ (using the results of Lemma 4.3.3 and the special-case rules) on $(0, 1)$. Also, $g_8(1-) < 0$, $f_8(0+) > 0$, and $\tilde{\rho}_8(0+) < 0$ imply $g_8 < 0$, $f_8 > 0$, and $\rho_7 = r_8 \searrow$ (by the refined general rules) on $(0, 1)$. Further, $f_7(0+) = g_7(0+) = 0$ imply $f_7 > 0$, $g_7 < 0$, and $\rho_6 = r_7 \searrow$ (again by the special-case rules) on $(0, 1)$.

Next, $g_6 \searrow$ and $g_6(0+) > 0 > g_6(1-)$ imply the existence of a single root x_6 ; $f_6 \nearrow$ and $f_6(0+) > 0$ imply $f_6 > 0$ on $(0, 1)$. The refined general rules imply $r_6 \nearrow$ on $(0, x_6)$ (as $\tilde{\rho}_6(0+) > 0$), and also that $\tilde{\rho}_6 \searrow$ on $(x_6, 1)$. As $x_6 < 0.75$ (since $g_6(0.75) < 0$), note that $\tilde{\rho}_6(x_6+) > \tilde{\rho}_6(0.75) > 0 > \tilde{\rho}_6(1-)$ implies $r_6 \nearrow \searrow$ on $(x_6, 1)$. That is, r'_6 has a single root z_6 , and hence we have $\rho_5 = r_6 \nearrow$ on each of $(0, x_6)$ and (x_6, z_6) and \searrow on $(z_6, 1)$.

Next, $g_5(0+) > 0$ and $g_5(1-) > 0$ (along with $g_5 \nearrow \searrow$) imply $g_5 > 0$ on $(0, 1)$; also, $f_5(0+) > 0$ implies $f_5 > 0$ on $(0, 1)$. As $x_6 > 0.5$ (since $g_6(0.5) > 0$) and $\tilde{\rho}_5 \nearrow$ on $(0, x_6)$ (by the refined general rules), one has $\tilde{\rho}_5(0+) < 0 < \tilde{\rho}_5(0.5) < \tilde{\rho}_5(x_6+)$; that is, $r_5 \searrow \nearrow$ on $(0, x_6)$, or r'_5 has a single root z_5 (with $z_5 < x_6$). Recall that f_5 , f'_5 and g_5 are all positive on $(0, 1)$, and also $g'_5 < 0$ on $(x_6, 1)$. Then $r'_5 = \frac{f'_5g_5 - f_5g'_5}{g_5^2} > 0$ and hence $r_5 \nearrow$ on $(x_6, 1)$. (Let us remark at this point that the l'Hospital-type rules could, in principle, be used to establish the monotonicity of r_5 on each of (x_6, z_6) and $(z_6, 1)$; however, this would necessitate proving that $\tilde{\rho}_5(z_6) > 0$, a task which requires more work than simply requesting the Mathematica program to evaluate the function at the *approximation* of the root z_6 .) As r_5 is continuous on $(0, 1)$, we have $\rho_4 = r_5 \searrow$ on $(0, z_5)$ and \nearrow on $(z_5, 1)$.

Next, $g_4(0+) = -\infty < 0 < g_4(1-)$ and $f_4(0+) = -\infty < 0 < f_4(1-)$ imply the existence of roots x_4 and y_4 (as $g_5 > 0$ and $f_5 > 0$). As $g_4(0.3) < 0 < r'_5(0.3)$, we see that $x_4 > 0.3 > z_5$; the refined general rules imply $\tilde{\rho}_4 \nearrow$ on $(0, z_5)$, and so, $\tilde{\rho}_4(0+) = 0$ implies $r_4 \nearrow$ on $(0, z_5)$. Also, $g_4(0.4) > 0 > f_4(0.4)$ implies $x_4 < 0.4 < y_4$, so that $r_4(x_4-) = \infty$ and $r_4(x_4+) = -\infty$. The general rules then imply $r_4 \nearrow$ on each of (z_5, x_4) and $(x_4, 1)$. Further, the continuity of r_4 at z_5 implies $\rho_3 = r_4 \nearrow$ on both $(0, x_4)$ and $(x_4, 1)$.

Next, $g_3 \searrow \nearrow$ and $g_3(0+) = 0 < g_3(1-)$ imply the existence of a single root x_3 ; at that, $x_3 > x_4$; similarly, $f_3(0+) = 0 < f_3(1-)$ implies the existence of y_3 . The special-case rules imply $r_3 \nearrow$ on $(0, x_4)$; $g_3(0.64) > 0 > f_3(0.64)$ implies $x_3 < 0.64 < y_3$, or $r_3(x_3-) = \infty$ and $r_3(x_3+) = -\infty$, so that the general rules show that $r_3 \nearrow$ on (x_4, x_3) and $(x_3, 1)$. As r_3 is continuous at x_4 , one has $\rho_2 = r_3 \nearrow$ on $(0, x_3)$ and $(x_3, 1)$.

Next, $g_2 \searrow \nearrow$, along with $g_2(0+) > 0 > g_2(0.5)$ and $g_2(1-) > 0$, implies the existence of two roots $x_{2,1}$ and $x_{2,2}$; similarly, $f_2(0+) > 0 > f_2(0.5)$ and $f_2(1-) > 0$ shows f_2 has two roots $y_{2,1}, y_{2,2}$. Noting that $g_2(0.35) < 0 < f_2(0.35)$ and also $g_2(0.86) > 0 > f_2(0.86)$, we have $x_{2,1} < 0.35 < y_{2,1} < 0.5 < x_{2,2} < 0.86 < y_{2,2}$, whence $r_2(x_{2,1}-) = r_2(x_{2,2}-) = \infty$ and $r_2(x_{2,1}+) = r_2(x_{2,2}+) = -\infty$; the general rules then imply that $r_2 \nearrow$ on each of $(0, x_{2,1})$, $(x_{2,1}, x_3)$, $(x_3, x_{2,2})$ and $(x_{2,2}, 1)$. The continuity of r_2 at x_3 implies $\rho_1 = r_2 \nearrow$ on $(0, x_{2,1})$, $(x_{2,1}, x_{2,2})$ and $(x_{2,2}, 1)$.

Next, $f_1(0+) = g_1(0+) = f_1(1-) = g_1(1-) = 0$ (together with f_2 and g_2 both $+-$) implies the existence of roots x_1 and y_1 . That $r_1 \nearrow$ on $(0, x_{2,1})$ and $(x_{2,2}, 1)$ is implied by the special-case rules; that $r_1 \nearrow$ on $(x_{2,1}, x_1)$ and $(x_1, x_{2,2})$ is implied by the general rules upon noting that $g_1(0.62) < 0 < f_1(0.62)$ (and hence $x_1 < y_1$, or $r_1(x_1-) = \infty$ and $r_1(x_1+) = -\infty$). The continuity of r_1 at $x_{2,1}$ and $x_{2,2}$ implies $\rho_0 = r_1 \nearrow$ on $(0, x_1)$ and $(x_1, 1)$.

Lastly, $f_0(0+) = g_0(0+) = f_0(1-) = g_0(1-) = 0$ shows $g_0 > 0$ on $(0, 1)$ and also, by the special-case rules, $r_0 \nearrow$ on $(0, x_1)$ and $(x_1, 1)$. The continuity of r_0 at x_1 shows $q_{T,S;0} = r_0 \nearrow$ on $(0, 1)$. Further, the l'Hospital rule for limits yields $r_0(0+) = r_2(0+)$ and $r_0(1-) = r_2(1-)$.

As promised in the remarks preceding Lemma 4.3.1, we show that $\sigma_S > 0$ on $(0, 1)$ (and hence on $(-1, 0)$ as σ_S is even). Note $f_0 > 0$ (as $f_0 \nearrow \searrow$ and $f_0(0+) = f_0(1-) = 0$); by (4.17) and (4.16), and recalling that $b = 1$ and $c = 0$, one has $f_0 = \sigma_S^2 - g$, so that $\sigma_S^2 > g$ on $(0, 1)$. As $x^2 g(x) = g_0(x) > 0$, it follows that $\sigma_S^2 > 0$. Further note that there is no circular reasoning here; the above proof stands on its own, regardless of any probabilistic interpretation we give to the functions f or g . \square

Proof of Theorem 4.2.1, (TS1). Adopt the notation of Lemma 4.3.3, with $a = 1$, so that $f(1-) = g(1-) = f'(1-) = g'(1-) = 0$, and repeated application of the l'Hospital rule for limits imply

$$b = \text{ARE}_{T,S}(1-) = \frac{f''(1-)}{g''(1-)} = \frac{9\sqrt{3}(11\sqrt{5} - 15)}{40\pi}$$

and

$$\begin{aligned} c = \text{ARE}'_{T,S}(1-) &= \frac{f'''(1-)g''(1-) - f''(1-)g'''(1-)}{3g''(1-)^2} \\ &= \frac{3(45(3 + \sqrt{3}\pi) - \sqrt{5}(99 + 5\sqrt{3}\pi))}{40\pi^2}. \end{aligned}$$

Then $f_9(0+) = g_9(0+) = 0$ (and $f_{10} > 0$, $g_{10} > 0$, by Lemma 4.3.3) imply that $f_9 > 0$, $g_9 > 0$ and $\rho_8 = r_9 \nearrow$ (by the special-case rules). Also, $f_8(0+) > 0$, $g_8(1-) < 0$ and $\tilde{\rho}_8(0+) < 0$ imply $f_8 > 0$, $g_8 < 0$, and (by the refined general rules) $\rho_7 = r_8 \searrow$ on $(0, 1)$.

Next, $g_7(0+) > 0 > g_7(1-)$ implies the existence of a single root x_7 ; $f_7(0+) > 0$ shows that $f_7 > 0$. The refined general rules imply $\tilde{\rho}_7 \nearrow$ on $(0, x_7)$ and \searrow on $(x_7, 1)$. As $\tilde{\rho}_7(0+) > 0$, we see $r_7 \nearrow$ on $(0, x_7)$; further, $x_7 < 0.2$ (implied by $g_7(0.2) < 0$) yields $\tilde{\rho}_7(x_7+) > \tilde{\rho}_7(0.2) > 0 > \tilde{\rho}_7(1-)$, so that $r_7 \nearrow \searrow$ on $(x_7, 1)$. That is, $\rho_6 = r_7 \nearrow$ on both of $(0, x_7)$ and (x_7, z_7) , and $\rho_6 = r_7 \searrow$ on $(z_7, 1)$.

Next, $g_6(0+) > 0 > g_6(1-)$ implies the existence of x_6 ; $f_6(0+) > 0$ implies $f_6 > 0$ on $(0, 1)$. As $\tilde{\rho}_6(0+) > 0$, the refined general rules imply $r_6 \nearrow$ on $(0, x_7)$. Further, $g_6(0.5) > 0 > r'_7(0.5)$ implies $z_7 < 0.5 < x_6$; as $f_6 > 0$, $f'_6 > 0$, $g_6 > 0$, and $g'_6 < 0$ on the interval (x_7, x_6) , we have $r'_6 = \frac{f'_6 g_6 - f_6 g'_6}{g_6^2} > 0$ and hence $r_6 \nearrow$ on (x_7, x_6) , so that $r_6 \nearrow$ on $(0, x_6)$ (since r_6 is continuous at x_7). Also, $\tilde{\rho}_6 \searrow$ on $(x_6, 1)$ is implied by the refined general rules; then $g_6(0.85) < 0$ implies $x_6 < 0.85$, so that $\tilde{\rho}_6(x_6+) > \tilde{\rho}_6(0.85) > 0 > \tilde{\rho}_6(1-)$ shows that $r_6 \nearrow \searrow$ on $(x_6, 1)$. That is, $\rho_5 = r_6 \nearrow$ on $(0, x_6)$ and (x_6, z_6) and \searrow on $(z_6, 1)$.

Next, $g_5(0+) > 0$ and $g_5(1-) > 0$, along with $g_5 \nearrow \searrow$, imply $g_5 > 0$ on $(0, 1)$; also, $f_5(0+) < 0 < f_5(1-)$ implies f_5 has a single root y_5 . The refined general rules imply $r_5 \nearrow$ on $(0, x_6)$, as $\tilde{\rho}_5(0+) > 0$; also, $f_5(0.5) > 0$ implies $y_5 < 0.5 < x_6$, so that $f_5 > 0$, $f'_5 > 0$, $g_5 > 0$ and $g'_5 < 0$ on $(x_6, 1)$, and hence $r'_5 = \frac{f'_5 g_5 - f_5 g'_5}{g_5^2} > 0$ on $(x_6, 1)$. As r_5 is continuous at x_6 , one has $\rho_4 = r_5 \nearrow$ on $(0, 1)$.

Next, $-\infty = g_4(0+) < 0 < g_4(1-)$ shows g_4 has a single root x_4 ; $f_4(0+) = \infty > 0 > f_4(0.75)$ and $f_4(1-) > 0$ shows f_4 has two roots $y_{4,1}$ and $y_{4,2}$. Also, $g_4(0.75) < 0 < g_4(0.8)$, $f_4(0.75) < 0$, and $f_4(0.8) < 0$ together imply $x_4 \in (0.75, 0.8) \subset (y_{4,1}, y_{4,2})$, so that $r_4(x_4-) = \infty$ and $r_4(x_4+) = -\infty$. The general rules then imply $\rho_3 = r_4 \nearrow$ on each of $(0, x_4)$ and $(x_4, 1)$.

Next, $g_3(0+) > 0 = g_3(1-)$ and $g_3 \searrow \nearrow$ shows g_3 has a single root x_3 ; $f_3(0+) > 0 = f_3(1-)$ and $f_3 \nearrow \searrow \nearrow$ shows f_3 has a single root y_3 . Then $r_3 \nearrow$ on $(x_4, 1)$ by the special-case rules; $g_3(0.5) < 0 < f_3(0.5)$ yields $x_3 < y_3$ (and hence $r_3(x_3-) = \infty$ and $r_3(x_3+) = -\infty$), so that the general rules imply $r_3 \nearrow$ on both of $(0, x_3)$ and (x_3, x_4) . As r_3 is continuous at x_4 , $\rho_2 = r_3 \nearrow$ on $(0, x_3)$ and $(x_3, 1)$.

Next, $g_2(0+) < 0 = g_2(1-)$ and $f_2(0+) < 0 = f_2(1-)$ together yield the existence of roots x_2 and y_2 , along with $r_2 \nearrow$ on $(x_3, 1)$ (via the special-case rules). Also, $g_2(0.1) > 0 > f_2(0.1)$ implies $x_2 < y_2$ (and hence $r_2(x_2-) = \infty$ and $r_2(x_2+) = -\infty$), so that the general rules then imply $r_2 \nearrow$ on $(0, x_2)$ and (x_2, x_3) . Further, r_2 is continuous at x_3 and hence $\rho_1 = r_2 \nearrow$ on $(0, x_2)$ and $(x_2, 1)$.

Next, $g_1(0+) < 0 = g_1(1-)$ and $f_1(0+) < 0 = f_1(1-)$ show that $g_1 < 0$ and $f_1 < 0$ on $(0, 1)$, and also $r_1 \nearrow$ on $(x_2, 1)$ by the special-case rules; $\tilde{\rho}_1(0+) > 0$ implies via the refined general rules that $r_1 \nearrow$ on $(0, x_2)$. The continuity of r_1 at x_2 then shows $\rho_0 = r_1 \nearrow$ on $(0, 1)$.

Lastly, $f_0(1-) = g_0(1-) = 0$ shows that $g_0 > 0$ and further, via the special-case rules, that $q_{T,S;1} = r_0 \nearrow$ on $(0, 1)$. Note $r_0(0+) = \frac{f_0(0+)}{g_0(0+)}$ and, by the l'Hospital rule for limits, $r_0(1-) = r_4(1-)$. \square

Proof of Theorem 4.2.1, (RS0). Set $a = 0$ in the notation of Lemma 4.3.4, so that, in

accordance with (4.15),

$$b = \text{ARE}_{R,S}(0) = \frac{\pi^2}{9} \quad \text{and} \quad c = \text{ARE}'_{R,S}(0) = 0.$$

Then $f_4(0+) = g_4(0+) = 0$, $f_5 > 0$, and $g_5 > 0$ (from Lemma 4.3.4) together imply that $f_4 > 0$, $g_4 > 0$, and also $\rho_3 = r_4 \nearrow$ (via the special-case rules).

Next, $g_3 \nearrow$ and $g_3(0+) < 0 < g_3(1-)$ implies the existence of the root x_3 ; that f_3 has a single root y_3 follows by $f_3 \nearrow$ and $f_3(0+) < 0 < f_3(1-)$. From $x_3 < y_3$ (implied by $g_3(0.64) > 0 > f_3(0.64)$) follows $r_3(x_3-) = \infty$ and $r_3(x_3+) = -\infty$; the general rules then imply $\rho_2 = r_3 \nearrow$ on both $(0, x_3)$ and $(x_3, 1)$.

That g_2 has two distinct roots $x_{2,1}$ and $x_{2,2}$ follows from $g_2(0+) > 0 > g_2(0.5)$ and $g_2(1-) > 0$ (along with $g_2 \searrow \nearrow$); similarly, f_2 has two roots $y_{2,1}$ and $y_{2,2}$, which follows from $f_2(0+) > 0 > f_2(0.5)$ and $f_2(1-) > 0$. Then $g_2(0.33) < 0 < f_2(0.33)$ shows that $x_{2,1} < y_{2,1}$, and $g_2(0.86) > 0 > f_2(0.86)$ (together with $0 > g_2(0.5)$ and $0 > f_2(0.5)$) show that $y_{2,1} < 0.5 < x_{2,2} < y_{2,2}$. The general rules then imply (since $r_2(x_{2,1}-) = r_2(x_{2,2}-) = \infty$ and $r_2(x_{2,1}+) = r_2(x_{2,2}+) = -\infty$) that $\rho_1 = r_2 \nearrow$ on the four intervals $(0, x_{2,1})$, $(x_{2,1}, x_3)$, $(x_3, x_{2,2})$ and $(x_{2,2}, 1)$; the continuity of r_2 at x_3 implies $\rho_1 = r_2 \nearrow$ on $(x_{2,1}, x_{2,2})$.

As $f_1(0+) = g_1(0+) = f_1(1-) = g_1(1-) = 0$, one finds the existence of roots x_1 and y_1 (since g_2 and f_2 are both $+-+$), as well as $r_1 \nearrow$ on $(0, x_{2,1})$ and $(x_{2,2}, 1)$ via the special-case rules. Further, $g_1(0.6) < 0 < f_1(0.6)$ shows $x_1 < y_1$ (and hence $r_1(x_1-) = \infty$ and $r_1(x_1+) = -\infty$), so that the general rules imply $r_1 \nearrow$ on $(x_{2,1}, x_1)$ and $(x_1, x_{2,2})$. The continuity of r_1 at $x_{2,1}$ and $x_{2,2}$ then implies $\rho_0 = r_1 \nearrow$ on $(0, x_1)$ and $(x_1, 1)$.

Lastly, $f_0(0+) = g_0(0+) = f_0(1-) = g_0(1-) = 0$ and $g_0 \nearrow \searrow$ imply that $g_0 > 0$ and also (by the special-case rules) that $r_0 \nearrow$ on both $(0, x_1)$ and $(x_1, 1)$. The continuity of r_0 at x_1 then implies $q_{R,S;0} = r_0 \nearrow$ on $(0, 1)$. The l'Hospital rule for limits implies $r_0(0+) = r_2(0+)$ and $r_0(1-) = r_2(1-)$. \square

Proof of Theorem 4.2.1, (RS1). Adopt the notation of Lemma 4.3.4, with $a = 1$, so that $f(1-) = g(1-) = f'(1-) = g'(1-) = 0$ and repeated application of the l'Hospital rule for limits together yield

$$b = \text{ARE}_{R,S}(1-) = \frac{f''(1-)}{g''(1-)} = \frac{3(11\sqrt{5} - 15)}{20}$$

and

$$c = \text{ARE}'_{R,S}(1-) = \frac{f'''(1-)g''(1-) - f''(1-)g'''(1-)}{3g''(1-)^2} = \frac{15 - 4\sqrt{5}}{5}.$$

From $g_4(0+) < 0 < g_4(1-)$ and $g_5 > 0$ follows the existence of x_4 ; similarly, $f_4(0+) < 0 < f_4(1-)$ and $f_5 > 0$ imply the existence of y_4 . Then $g_4(0.8) > 0 > f_4(0.8)$ shows $x_4 < 0.8 < y_4$, or hence $r_4(x_4-) = \infty$ and $r_4(x_4+) = -\infty$, and so the general rules imply $\rho_3 = r_4 \nearrow$ on both $(0, x_4)$ and $(x_4, 1)$.

Next, $g_3 \searrow \nearrow$ (as g_4 is $+-$) and $g_3(0+) = \infty > 0 = g_3(1-)$ yield the existence of x_3 ; that f_3 has a single root y_3 also follows by $f_3 \searrow \nearrow$ and $f_3(0+) = \infty > 0 = f_3(1-)$. The special-case rules imply $r_3 \nearrow$ on $(x_4, 1)$; also $x_3 < y_3$ follows from $g_3(0.5) < 0 < f_3(0.5)$ (whence $r_3(x_3-) = \infty$ and $r_3(x_3+) = -\infty$), and so, the general rules imply $r_3 \nearrow$ on both

of $(0, x_3)$ and (x_3, x_4) . Also, r_3 is continuous at x_4 and hence $\rho_2 = r_3 \nearrow$ on $(0, x_3)$ and $(x_3, 1)$.

As $g_2(0+) < 0 = g_2(1-)$ and $f_2(0+) < 0 = f_2(1-)$ (and g_3 and f_3 are both $+-$), there exist roots x_2 and y_2 ; the special-case rules imply $r_2 \nearrow$ on $(x_3, 1)$. Further, $g_2(0.1) > 0 > f_2(0.1)$ shows $x_2 < 0.1 < y_2$ and hence $r_2(x_2-) = \infty$ and $r_2(x_2+) = -\infty$. The general rules then imply $r_2 \nearrow$ on $(0, x_2)$ and (x_2, x_3) ; the continuity of r_2 at x_3 then implies $\rho_1 = r_2 \nearrow$ on $(0, x_2)$ and $(x_2, 1)$.

One finds that $g_1 < 0$ and $f_1 < 0$ on $(0, 1)$, as $g_1(0+) < 0 = g_1(1-)$ (with $g_1 \searrow \nearrow$) and $f_1(0+) < 0 = f_1(1-)$ (with $f_1 \searrow \nearrow$), which further imply by the special-case rules that $r_1 \nearrow$ on $(x_2, 1)$. Also, $\tilde{\rho}_1(0+) > 0$ implies via the refined general rules that $\tilde{\rho}_1 > 0$, or $r_1 \nearrow$, on $(0, x_2)$; as r_1 is continuous on $(0, 1)$, one sees $\rho_0 = r_1 \nearrow$ on $(0, 1)$.

Lastly, $f_0(1-) = g_0(1-) = 0$ imply in the first place that $g_0 > 0$ (as $g_0 \searrow$), and in the second place that $q_{R,S;1} = r_0 \nearrow$ on $(0, 1)$ (via the special-case rules). The l'Hospital rule for limits implies $r_0(1-) = r_4(1-)$, and $g_0(0+) > 0$ implies $r_0(0+) = \frac{f_0(0+)}{g_0(0+)}$. \square

Proof of Corollary 4.2.2. As the ARE's are even functions here, one has $\text{ARE}'(0) = 0$, and hence $\text{ARE}(x) = \text{ARE}(0) + x^2 q_0(x)$ for $x \in (0, 1)$. Theorem 4.2.1 shows $q_0 \nearrow$ and $q_0(0+) > 0$, which imply $q_0 > 0$ on $(0, 1)$; hence $\text{ARE} \nearrow$ on $(0, 1)$ as well. The values $\text{ARE}(0+)$ and $\text{ARE}(1-)$ are exactly those values of b given at the beginning of the proof of each of the six parts of Theorem 4.2.1. \square

Proof of Corollary 4.2.3. The result immediately follows from Theorem 4.2.1:

$$\begin{aligned} (x-a)^2 q_a(0+) &< \text{ARE}(x) - \text{ARE}(a) - \text{ARE}'(a)(x-a) < (x-a)^2 q_a(1-) \\ \Rightarrow L_a(x) &< \text{ARE}(x) < U_a(x) \end{aligned}$$

for all $x \in (0, 1)$ and $a \in \{0, 1\}$. Replacing “ x ” with “ $-x$ ” in the above inequality when $x \in (-1, 0)$ and recalling the ARE is even yields the desired results. \square

Proof of Corollary 4.2.7. Noting that $\text{ARE}_{R,S} = \text{ARE}_{R,T} \cdot \text{ARE}_{T,S}$, one has $L_{R,T} \cdot L_{T,S} < \text{ARE}_{R,S} < U_{R,T} \cdot U_{T,S}$ on $(-1, 1) \setminus \{0\}$. That $\tilde{L}_{R,S} > L_{R,S}$ and $\tilde{U}_{R,S} < U_{R,S}$ is easily verified by noting $\tilde{L}_{R,S} - L_{R,S}$ and $\tilde{U}_{R,S} - U_{R,S}$ have no roots on $(-1, 1) \setminus \{0\}$ and verifying their appropriate signs. \square

Future work

In the course of researching the various results presented in the previous four chapters, several questions arose which as of yet remain unanswered and hence can serve as starting points for further investigations. We consider some of these problems below:

- In a previous version of the paper presented in Chapter 1, we used a nonuniform BE bound found in the paper [89] by Nefedova and Shevtsova. The inequality stated there allowed for us to replace the constant 30.2211 in (1.26) with an expression $\mathbf{NS}(z)$, with $\mathbf{NS}(0+) = 18.1139$, $\mathbf{NS}(\infty) = 1$, and \mathbf{NS} decreasing on $(0, \infty)$. Using this inequality allowed for some significant improvements to the constants in our nonuniform bound on the self-normalized sum. Namely, assuming the validity of the result from [89], we could prove that the tables 1.1 and 1.2 could be replaced by

	$\omega = 0.1$			$\omega = 0.5$		
	\hat{A}_3	\hat{A}_4	\hat{A}_6	\hat{A}_3	\hat{A}_4	\hat{A}_6
$w_g = 1$	24	24	24	30	29	29
	27	14	6	60	29	13
$w_g = 0$	142	139	138	161	157	157
	161	82	27	225	113	44

and

	$\omega = 0.1$			$\omega = 0.5$		
	\hat{A}_3	\hat{A}_4	\hat{A}_6	\hat{A}_3	\hat{A}_4	\hat{A}_6
$w_g = 1$	21	20	20	25	25	25
	23	12	3	52	26	11
$w_g = 0$	113	113	111	134	133	131
	132	67	16	193	98	39

respectively. Unfortunately, the proof of the inequality in [89] was broken into three cases which did not exhaust all possibilities. It is plausible that an inequality similar to Nefedova and Shevtsova's could be obtained by considering the missing case; while fixing their proof would likely result in an expression $\mathbf{NS}(z)$ which is greater than the one they report, it seems reasonable to suppose that a corrected bound would still be smaller than Michel's bound in [81].

- After finding the aforementioned error in [89], it was deemed prudent to check the uniform BE bound in Shevtsova's paper [136], upon which the constants reported in Corollaries 1.4.10, 1.4.11, and 1.4.20 relied. Our check of Shevtsova's result suggested

that the bound of $0.4748 \mathbb{E}|X_1|^3/\sqrt{n}$ (on the distance to normality of $\sum_j X_j/\sqrt{n}$ when the X_j 's are i.i.d., zero-mean, and unit-variance) could be improved by lowering the constant 0.4748. This potential for a smaller constant was discovered by replacing $\min\{\int f, \int g\}$ with an upper bound on $\int \min\{f, g\}$ in calculations suggested by the proof in [136], where f and g are somewhat complicated functions whose integrals are calculated with the aid of a computer algebra system (CAS) such as Mathematica. At the time of this writing, further calculations are needed to see how significant of a reduction in the absolute constant 0.4748 can be achieved.

- There is also potential improvement to the structure of the most recent uniform BE bounds proved by Tyurin [141] and Shevtsova [136]. Both of these papers rely in part on some inequalities on characteristic functions proved by Prawitz in [123, 124]. Namely, Prawitz bounds the absolute value of a characteristic function of a r.v. (and also of the sum of independent r.v.'s) by several functions which depend on the first three absolute moments of the r.v. (or the sums of these moments when considering the sum of independent r.v.'s). The bounds found in [141, 136] use only the inequalities of Prawitz which contain the second and third absolute moments of a distribution. Work is currently being performed to see if the methods employed there can be adapted so as to result in a bound which also depends on the first absolute moments of the random summands, with the hopes that such a bound would yield improvements to the constants found in [141, 136].
- As the subject of BE bounds on the F -statistic used in linear models appears to be fairly unexplored in the literature, there are several options available in extending the work found in Chapter 2. For instance, a bound which assumes only the existence of the $2p^{\text{th}}$ absolute moments of the error terms in the model, for $p \in (2, 3)$, could likely be obtained using the existing framework given in that paper. Considerably different techniques would be needed to obtain a nonuniform bound, a bound on the non-null distribution of F , or a bound where the components of the error vector ε are not assumed to be independent, though all three of these generalizations would be valuable additions to the literature.

As became evident in the work that went into creating this dissertation, following one line of inquiry will invariably lead to questions in seemingly unrelated areas of research. The search for answers to these questions of course leads to more questions, and this proceeds along an often unpredictable route. The above-mentioned work will undoubtedly provide many years' worth of food for thought to this researcher, and then hopefully to a larger research community as well.

References

- [1] M. M. Ali and S. C. Sharma. Robustness to nonnormality of regression F -tests. 71:175–205, 1996.
- [2] G. D. Anderson, M. K. Vamanamurthy, and M. Vuorinen. Inequalities for quasiconformal mappings in space. *Pacific Journal of Mathematics*, 160(1):1–18, 1993.
- [3] M. Atiqullah. The estimation of residual variance in quadratically balanced least-squares problems and the robustness of the f -test. 49(1/2):83–91, 1962.
- [4] G. Bachman and L. Narici. *Functional analysis*. Academic Press, New York, 1966.
- [5] R. R. Bahadur. Stochastic comparison of tests. *Annals of Mathematical Statistics*, 31:276–295, 1960.
- [6] A. D. Barbour and L. H. Y. Chen, editors. *An introduction to Stein's method*, volume 4 of *Lecture Notes Series. Institute for Mathematical Sciences. National University of Singapore*. Singapore University Press, Singapore, 2005. Lectures from the Meeting on Stein's Method and Applications: a Program in Honor of Charles Stein held at the National University of Singapore, Singapore, July 28–August 31, 2003.
- [7] G. Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.
- [8] V. Bentkus, M. Bloznelis, and F. Götze. A Berry-Esséen bound for Student's statistic in the non-i.i.d. case. *Journal of Theoretical Probability*, 9(3):765–796, 1996.
- [9] V. Bentkus and F. Götze. On smoothness conditions and convergence rates in the CLT in Banach spaces. *Probability Theory and Related Fields*, 96(2):137–151, 1993.
- [10] V. Bentkus and F. Götze. The Berry-Esseen bound for Student's statistic. *The Annals of Probability*, 24(1):491–503, 1996.
- [11] V. Bentkus, B.-Y. Jing, Q.-M. Shao, and W. Zhou. Limiting distributions of the non-central t -statistic and their applications to the power of t -tests under non-normality. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 13(2):346–364, 2007.
- [12] K. J. Berry and P. W. Mielke. A Monte Carlo investigation of the Fisher Z transformation for normal and nonnormal distributions. *Psychological Reports*, 87:1101–1114, 2000.

- [13] R. Bhattacharya and S. Holmes. An exposition of Götze's estimation of the rate of convergence in the multivariate central limit theorem (preprint). <http://arxiv.org/abs/1003.4254>, 2010.
- [14] R. N. Bhattacharya and J. K. Ghosh. On the validity of the formal Edgeworth expansion. *The Annals of Statistics*, 6(2):434–451, 1978.
- [15] C. B. Borkowf. Computing the nonnull asymptotic variance and the asymptotic relative efficiency of Spearman's rank correlation. *Computational Statistics & Data Analysis*, 39(3):271–286, 2002.
- [16] G. E. P. Box and G. S. Watson. Robustness to non-normality of regression tests. *Biometrika*, 49(1/2):93–106, 1962.
- [17] L. H. Chen and X. Fang. Multivariate normal approximation by Stein's method: The concentration inequality approach. <http://arxiv.org/abs/1111.4073>, 2011.
- [18] L. H. Chen and X. Fang. Multivariate normal approximation by Stein's method: the concentration inequality approach (preprint). <http://arxiv.org/abs/1111.4073>, 2011.
- [19] L. H. Y. Chen and Q.-M. Shao. Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 13(2):581–599, 2007.
- [20] G. P. Chistyakov and F. Götze. Moderate deviations for Student's statistic. *Theory of Probability and Its Applications*, 47(3):415–428, 2003.
- [21] G. P. Chistyakov and F. Götze. On bounds for moderate deviations for Student's statistic. *Theory of Probability and Its Applications*, 48(3):528–535, 2004.
- [22] G. E. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In *Quantifier elimination and cylindrical algebraic decomposition (Linz, 1993)*, Texts Monogr. Symbol. Comput., pages 85–121. Springer, Vienna, 1998.
- [23] H. Cramér. *Mathematical methods of statistics*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1999. Reprint of the 1946 original.
- [24] H. E. Daniels. The relation between measures of correlation in the universe of sample permutations. *Biometrika*, 33:129–135, 1944.
- [25] H. E. Daniels. Rank correlation and population models. *Journal of the Royal Statistical Society. Series B. Methodological*, 12:171–181, 1950.
- [26] F. N. David. *Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples*. Cambridge University Press, New York, 1938.
- [27] F. N. David and C. L. Mallows. The variance of Spearman's rho in normal samples. *Biometrika*, 48:19–28, 1961.

- [28] A. de Acosta and J. D. Samur. Infinitely divisible probability measures and the converse Kolmogorov inequality in Banach spaces. *Polska Akademia Nauk. Institut Matematyczny. Studia Mathematica*, 66(2):143–160, 1979.
- [29] A. Dembo and Q.-M. Shao. Large and moderate deviations for Hotelling’s T^2 -statistic. *Electronic Communications in Probability*, 11:149–159 (electronic), 2006.
- [30] J. Durbin and A. Stuart. Inversions and rank correlation coefficients. *Journal of the Royal Statistical Society. Series B. Methodological*, 13:303–309, 1951.
- [31] M. L. Eaton. *Multivariate statistics: a vector space approach*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1983.
- [32] H. H. E.C. Fieller and E. Pearson. Tests for rank correlation coefficients. i. 44(3/4):470–481, 1957.
- [33] D. J. G. Farlie. The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47:307–323, 1960.
- [34] R. A. Fisher. On the probable error of a coefficient of correlation deduced from a small sample. 1:3–32, 1921.
- [35] M. J. Frank. On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematicae*, 19(2-3):194–226, 1979.
- [36] M. Fréchet. Sur les tableaux de corrélation dont les marges sont données. 14:53–77, 1951.
- [37] Y. Fujikoshi. An asymptotic expansion for the distribution of Hotelling’s T^2 -statistic under nonnormality. *Journal of Multivariate Analysis*, 61(2):187–193, 1997.
- [38] A. K. Gayen. The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika*, 38:219–247, 1951.
- [39] C. Genest. Frank’s family of bivariate distributions. *Biometrika*, 74(3):549–555, 1987.
- [40] B. K. Ghosh. Asymptotic expansions for the moments of the distribution of correlation coefficient. *Biometrika*, 53:258–262, 1966.
- [41] E. Giné, F. Götze, and D. M. Mason. When is the Student t -statistic asymptotically standard normal? *The Annals of Probability*, 25(3):1514–1531, 1997.
- [42] E. Giné and D. M. Mason. On the LIL for self-normalized sums of IID random variables. *Journal of Theoretical Probability*, 11(2):351–370, 1998.
- [43] F. Götze. On the rate of convergence in the central limit theorem in Banach spaces. *The Annals of Probability*, 14(3):922–942, 1986.

- [44] F. Götze. On the rate of convergence in the multivariate CLT. *The Annals of Probability*, 19(2):724–739, 1991.
- [45] P. Hall. Edgeworth expansion for Student’s t statistic under minimal moment conditions. *The Annals of Probability*, 15(3):920–931, 1987.
- [46] P. R. Halmos. *Finite-dimensional vector spaces*. The University Series in Undergraduate Mathematics. D. Van Nostrand Co., Inc., Princeton-Toronto-New York-London, 1958. 2nd ed.
- [47] S. W. Harrar and A. K. Gupta. Asymptotic expansion for the null distribution of the F -statistic in one-way anova under non-normality. *Annals of the Institute of Statistical Mathematics*, 59(3):531–556, 2007.
- [48] D. G. Herr. On the history of the use of geometry in the general linear model. *The American Statistician*, 34(1):43–47, 1980.
- [49] J. L. Hodges, Jr. and E. L. Lehmann. The efficiency of some nonparametric competitors of the t -test. *Annals of Mathematical Statistics*, 27:324–335, 1956.
- [50] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19:293–325, 1948.
- [51] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.
- [52] W. Hoeffding and J. R. Rosenblatt. The efficiency of tests. *Annals of Mathematical Statistics*, 26:52–63, 1955.
- [53] J. Hoffmann-Jørgensen and G. Pisier. The law of large numbers and the central limit theorem in Banach spaces. *The Annals of Probability*, 4(4):587–599, 1976.
- [54] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1985.
- [55] H. Hotelling. New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society. Series B. Methodological*, 15:193–225; discussion, 225–232, 1953.
- [56] H. Hotelling and M. R. Pabst. Rank correlation and tests of significance involving no assumption of normality. *Annals of Mathematical Statistics*, 7:29–43, 1936.
- [57] R. Ibragimov and S. Sharakhmetov. On an exact constant for the Rosenthal inequality. *Rossiiskaya Akademiya Nauk. Teoriya Veroyatnostei i ee Primeneniya*, 42(2):341–350, 1997.
- [58] K. Ito. Asymptotic formulae for the distribution of Hotelling’s generalized T_0^2 statistic. *Annals of Mathematical Statistics*, 27(4):1091–1105, 1956.
- [59] K. Ito. Asymptotic formulae for the distribution of Hotelling’s generalized T_0^2 statistic. II. *Annals of Mathematical Statistics*, 31(4):1148–1153, 1960.

- [60] N. C. Jain and M. B. Marcus. Integrability of infinite sums of independent vector-valued random variables. *Transactions of the American Mathematical Society*, 212:1–36, 1975.
- [61] B.-Y. Jing, Q.-M. Shao, and Q. Wang. Self-normalized Cramér-type large deviations for independent random variables. *The Annals of Probability*, 31(4):2167–2215, 2003.
- [62] B.-Y. Jing, Q.-M. Shao, and W. Zhou. Towards a universal self-normalized moderate deviation. *Transactions of the American Mathematical Society*, 360(8):4263–4285, 2008.
- [63] M. E. Johnson and A. Tenenbein. Bivariate distributions with given marginals and fixed measures of dependence. Technical Report LA-7700-MS, Los Alamos Scientific Lab., 1979.
- [64] M. E. Johnson and A. Tenenbein. A bivariate distribution family with specified marginals. *Journal of the American Statistical Association*, 76(373):198–201, 1981.
- [65] Y. Kakizawa and T. Iwashita. Hotelling’s one-sample and two-sample T^2 tests and the multivariate Behrens-Fisher problem under nonnormality. *Journal of Statistical Planning and Inference*, 138(11):3379–3404, 2008.
- [66] Y. Kano. An asymptotic expansion of the distribution of Hotelling’s T^2 -statistic under general distributions. *American Journal of Mathematical and Management Sciences*, 15(3-4):317–341, 1995.
- [67] M. Kendall and J. D. Gibbons. *Rank correlation methods*. A Charles Griffin Title. Edward Arnold, London, fifth edition, 1990.
- [68] M. G. Kendall. A new measure of rank correlation. 30(1/2):81–93, June 1938.
- [69] G. Kimeldorf and A. Sampson. One-parameter families of bivariate distributions with fixed marginals. 4:293–301, 1975.
- [70] T. Kohno. The volume of a hyperbolic simplex and iterated integrals. In *Intelligence of low dimensional topology 2006*, volume 40 of *Ser. Knots Everything*, pages 179–188. World Sci. Publ., Hackensack, NJ, 2007.
- [71] T. Kollo and K. Ruul. Approximations to the distribution of the sample correlation matrix. *Journal of Multivariate Analysis*, 85(2):318–334, 2003.
- [72] H. S. Konijn. On the power of certain tests for independence in bivariate populations. *Annals of Mathematical Statistics*, 27:300–323, 1956.
- [73] V. S. Koroljuk and Y. V. Borovskich. *Theory of U-statistics*, volume 273 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994. Translated from the 1989 Russian original by P. V. Malyshev and D. V. Malyshev and revised by the authors.

- [74] C. J. Kowalski. On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 21(1):1–12, 1972.
- [75] H. C. Kraemer. On estimation and hypothesis testing problems for correlation coefficients. 40(4):473–485, 1975.
- [76] W. Kruskal. When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. *Annals of Mathematical Statistics*, 39(1):70–75, 1968.
- [77] S. F. H. K. M. G. Kendall and B. B. Smith. The distribution of Spearman’s coefficient of rank correlation in a universe in which all rankings occur an equal number of times. 30(3/4):251–273, Jan. 1939.
- [78] K. V. Mardia. Some contributions to contingency-type bivariate distributions. *Biometrika*, 54:235–249, 1967.
- [79] K. V. Mardia. The performance of some tests of independence for contingency-type bivariate distributions. *Biometrika*, 56(2):449–451, 1969.
- [80] K. V. Mardia. *Families of bivariate distributions*. Hafner Publishing Co., Darien, Conn., 1970. Griffin’s Statistical Monographs and Courses, No. 27.
- [81] R. Michel. On the constant in the nonuniform version of the Berry-Esseen theorem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 55(1):109–117, 1981.
- [82] R. Molzon and I. Pinelis. Monotonicity properties of the asymptotic relative efficiency between common correlation statistics in the bivariate normal model (preprint). <http://arxiv.org/abs/0907.5448>, 2009.
- [83] P. A. P. Moran. Partial and multiple rank correlations. *Biometrika*, 38:26–32, 1951.
- [84] P. A. P. Moran. The use of correlation in large samples. *The Australian Journal of Statistics*, 21(3):293–300, 1979.
- [85] R. J. Muirhead. The asymptotic noncentral distribution of Hotelling’s generalized T_0^2 . *Annals of Mathematical Statistics*, 43(5):1671–1677, 1972.
- [86] J. Murakami and M. Yano. On the volume of a hyperbolic and spherical tetrahedron. *Communications in Analysis and Geometry*, 13(2):379–400, 2005.
- [87] S. V. Nagaev. The Berry-Esseen bound for self-normalized sums. *Siberian Advances in Mathematics*, 12(3):79–125, 2002.
- [88] S. V. Nagaev. On large deviations of a self-normalized sum. *Theory of Probability and Its Applications*, 49(4):704–713, 2005.
- [89] Y. S. Nefedova and I. G. Shevtsova. Nonuniform estimates of convergence rate in the central limit theorem (in Russian). 57(1):62–97, 2012.

- [90] R. B. Nelsen. Properties of a one-parameter family of bivariate distributions with specified marginals. *Communications in Statistics. A. Theory and Methods*, 15(11):3277–3285, 1986.
- [91] R. B. Nelsen. *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition, 2006.
- [92] G. E. Noether. On a theorem of Pitman. *Annals of Mathematical Statistics*, 26(1):64–68, 1955.
- [93] S. Y. Novak. On self-normalized sums. *Mathematical Methods of Statistics*, 9(4):415–436, 2000.
- [94] S. Y. Novak. On self-normalized sums and Student’s statistic. *Theory of Probability and Its Applications*, 49(2):336–344, 2005.
- [95] L. Paditz. Abschätzungen der Konvergenzgeschwindigkeit zur Normalverteilung unter Voraussetzung einseitiger Momente. *Mathematische Nachrichten*, 82:131–156, 1978.
- [96] T.-x. Pang, L.-x. Zhang, and J.-f. Wang. Precise asymptotics in the self-normalized law of the iterated logarithm. *Journal of Mathematical Analysis and Applications*, 340(2):1249–1262, 2008.
- [97] V. V. Petrov. *Sums of independent random variables*. Springer-Verlag, New York, 1975. Translated from the Russian by A. A. Brown, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Band 82.
- [98] I. Pinelis. Comparing the asymptotic closeness of the distributions of Pearson’s and Fisher’s correlation statistics to normality (draft).
- [99] I. Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *The Annals of Probability*, 22(4):1679–1706, 1994.
- [100] I. Pinelis. Optimum bounds on moments of sums of independent random vectors. *Siberian Advances in Mathematics*, 5(3):141–150, 1995.
- [101] I. Pinelis. L’Hospital type rules for oscillation, with applications. *JIPAM. Journal of Inequalities in Pure and Applied Mathematics*, 2(3):Article 33, 24 pp. (electronic), 2001.
- [102] I. Pinelis. L’Hospital type results for monotonicity, with applications. *JIPAM. Journal of Inequalities in Pure and Applied Mathematics*, 3(1):Article 5, 5 pp. (electronic), 2002.
- [103] I. Pinelis. On l’Hospital-type rules for monotonicity. *JIPAM. Journal of Inequalities in Pure and Applied Mathematics*, 7(2):Article 40, 19 pp. (electronic), 2006.
- [104] I. Pinelis. Toward the best constant factor for the Rademacher-Gaussian tail comparison. *ESAIM. Probability and Statistics*, 11:412–426, 2007.

- [105] I. Pinelis. Optimal two-value zero-mean disintegration of zero-mean random variables. *Electronic Journal of Probability*, 14:no. 26, 663–727, 2009.
- [106] I. Pinelis. On the non-degeneracy of Kendall’s and Spearman’s correlation coefficients. <http://arxiv.org/abs/0711.3400>, 2010.
- [107] I. Pinelis. Exact bounds on the truncated-tilted mean, with applications (preprint). <http://arxiv.org/abs/1103.3683>, 2011.
- [108] I. Pinelis. Exact lower bounds on the exponential moments of truncated random variables. *Journal of Applied Probability*, 48:547–60, 2011.
- [109] I. Pinelis. Improved nonuniform Berry–Esseen-type bounds (preprint). <http://arxiv.org/abs/1109.0741>, 2011.
- [110] I. Pinelis. Monotone tail and moment ratio properties of Student’s family of distributions. <http://arxiv.org/abs/1101.3289>, 2011.
- [111] I. Pinelis. On the Berry–Esseen bound for the Student statistic (preprint). <http://arxiv.org/abs/1101.3286>, 2011.
- [112] I. Pinelis. Optimal re-centering bounds, with applications to Rosenthal-type concentration of measure inequalities (preprint). <http://arxiv.org/abs/1111.2622>, 2011.
- [113] I. Pinelis. An exact bound on the truncated-tilted mean for symmetric distributions (preprint). <http://arxiv.org/abs/1205.5234>, 2012.
- [114] I. Pinelis. On the Bennett–Hoeffding inequality, *a shorter version to appear in Annales de l’Institut Henri Poincaré*. <http://arxiv.org/abs/0902.4058>, 2012.
- [115] I. Pinelis and R. Molzon. Relative efficiency between Kendall’s, Spearman’s, and Pearson’s correlation statistics (preprint).
- [116] I. Pinelis and R. Molzon. Berry–Esseen bounds for general nonlinear statistics, with applications to Pearson’s and non-central Student’s and Hotelling’s (preprint). <http://arxiv.org/abs/0906.0177v2>, 2011.
- [117] I. F. Pinelis. Estimates of moments of infinite-dimensional martingales. *Mathematical Notes*, 27(6):459–462, 1980.
- [118] I. F. Pinelis. Probability inequalities for sums of independent random variables with values in a Banach space. *Mathematical Notes*, 39(3):241–244, 1986.
- [119] I. F. Pinelis and A. I. Sakhanenko. Remarks on inequalities for large deviation probabilities. *Theory of probability and its applications*, 30(1):143–148, 1986.
- [120] I. F. Pinelis and S. A. Utev. Exact exponential bounds for sums of independent random variables. *Theory of Probability and Its Applications*, 34(2):340–346, 1989.

- [121] R. L. Plackett. A reduction formula for normal multivariate integrals. *Biometrika*, 41:351–360, 1954.
- [122] R. L. Plackett. A class of bivariate distributions. *Journal of the American Statistical Association*, 60:516–522, 1965.
- [123] H. Prawitz. Ungleichungen für den absoluten Betrag einer charakteristischen Funktion. pages 11–16, 1973.
- [124] H. Prawitz. Weitere Ungleichungen für den absoluten Betrag einer charakteristischen Funktion. *Scandinavian Actuarial Journal*, pages 21–28, 1975.
- [125] Y. Rinott. A multivariate clt for local dependence with $n^{-1/2} \log n$ rate and applications to multivariate graph related statistics. *Journal of Multivariate Analysis*, 56:333–350, 1996.
- [126] J. Robinson and Q. Wang. On the self-normalized Cramér-type large deviation. *Journal of Theoretical Probability*, 18(4):891–909, 2005.
- [127] H. P. Rosenthal. On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8:273–303, 1970.
- [128] G. Rothe. Some properties of the asymptotic relative Pitman efficiency. *The Annals of Statistics*, 9(3):663–669, 1981.
- [129] H. Scheffé. A useful convergence theorem for probability distributions. *Annals of Mathematical Statistics*, 18:434–438, 1947.
- [130] H. Scheffé. *The Analysis of Variance*. Wiley Classics Library. John Wiley & Sons, 1999.
- [131] L. Schläfli. On the multiple integral $\int^n dx dy \dots dz$, whose limits are $p_1 = a_1 x + b_1 y + \dots + h_1 z > 0$, $p_2 > 0, \dots p_n > 0$, and $x^2 + y^2 + \dots + z^2 < 1$. *The quarterly journal of pure and applied mathematics*, 2:269–301, 1858.
- [132] Q.-M. Shao. Self-normalized large deviations. *The Annals of Probability*, 25(1):285–328, 1997.
- [133] Q.-M. Shao. A Cramér type large deviation result for Student’s t -statistic. *Journal of Theoretical Probability*, 12(2):385–398, 1999.
- [134] Q.-M. Shao. An explicit Berry-Esseen bound for Student’s t -statistic via Stein’s method. In *Stein’s method and applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 143–155. Singapore Univ. Press, Singapore, 2005.
- [135] S. Shen. Large deviation for the empirical correlation coefficient of two Gaussian random variables. *Acta Mathematica Scientia. Series B. English Edition*, 27(4):821–828, 2007.
- [136] I. Shevtsova. On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands. <http://arxiv.org/abs/1111.6554>, 2011.

- [137] M. Siotani. An asymptotic expansion of the non-null distribution of Hotelling's generalized T_0^2 -statistic. *Annals of Mathematical Statistics*, 42:560–571, 1971.
- [138] M. Sklar. Fonctions de répartition à n dimensions et leurs marges. 8:229–231, 1959.
- [139] V. V. Slavova. On the Berry-Esseen bound for Student's statistic. In *Stability problems for stochastic models (Uzhgorod, 1984)*, volume 1155 of *Lecture Notes in Math.*, pages 355–390. Springer, Berlin, 1985.
- [140] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. RAND Corporation, Santa Monica, Calif., 1948.
- [141] I. Tyurin. New estimates of the convergence rate in the Lyapunov theorem (preprint). <http://arxiv.org/abs/0912.0726>, 2009.
- [142] Q. Wang and P. Hall. Relative errors in central limit theorems for Student's t statistic, with applications. *Statistica Sinica*, 19(1):343–354, 2009.
- [143] Q. Wang and B.-Y. Jing. An exponential nonuniform Berry-Esseen bound for self-normalized sums. *The Annals of Probability*, 27(4):2068–2088, 1999.
- [144] M. Wichura. *The Coordinate-Free Approach to Linear Models*. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, 2006.
- [145] H. Yanagihara. Asymptotic expansion of the null distribution of one-way anova test statistic for heteroscedastic case under nonnormality. *Communications in Statistics. A. Theory and Methods*, 29(2):463–476, 2000.
- [146] H. Yanagihara. Asymptotic expansion of the null distribution of test statistic for linear hypothesis in nonnormal linear model. *Journal of Multivariate Analysis*, 84:222–246, 2003.
- [147] B. A. Zaleskii. On the accuracy of normal approximation in Banach spaces. *Theory of Probability and Its Applications*, 33(2):239–247, 1988.
- [148] B. A. Zaleskii. The accuracy of Gaussian approximation in Banach spaces. *Theory of Probability and Its Applications*, 34(4):747–748, 1990.