Approximate Central Limit Theorems

Peer-reviewed author version

BERCKMOES, Ben & MOLENBERGHS, Geert (2018) Approximate Central Limit Theorems. In: JOURNAL OF THEORETICAL PROBABILITY, 31(3), p. 1590-1605.

DOI: 10.1007/s10959-017-0744-6
Handle: http://hdl.handle.net/1942/27696

APPROXIMATE CENTRAL LIMIT THEORY AND STATISTICAL INFERENCE WITH CONTAMINATED DATA

BEN BERCKMOES AND GEERT MOLENBERGHS

Abstract. We refine the classical Lindeberg-Feller central limit theorem by obtaining asymptotic bounds on the Kolmogorov distance, the Wasserstein distance, and the parametrized Prokhorov distances in terms of a Lindeberg index. We thus obtain more general approximate central limit theorems, which roughly state that the row-wise sums of a triangular array are approximately asymptotically normal if the array approximately satisfies Lindeberg's condition. Stein's method plays a key role in the development of this theory. We use the theoretical results to study the asymptotic behavior of the sample average estimator in the presence of data contamination.

1. Introduction

Throughout, we assume that all random variables are defined on a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Let ξ be a standard normal random variable, that is, a normally distributed random variable with $\mathbb{E}[\xi] = 0$ and $\mathbb{E}[\xi^2] = 1$, and $\{\xi_{n,k}\}$ a standard triangular array (STA) of random variables, that is, a triangular array

$$\begin{array}{lll} \xi_{1,1} & & \\ \xi_{2,1} & \xi_{2,2} & \\ \xi_{3,1} & \xi_{3,2} & \xi_{3,3} \\ & \vdots & \end{array}$$

of random variables with $\xi_{n,1}, \ldots, \xi_{n,n}$ independent for all n, $\mathbb{E}[\xi_{n,k}] = 0$ for all n, k, and $\sum_{k=1}^{n} \mathbb{E}[\xi_{n,k}^2] = 1$ for all n.

Recall that the sequence $\left(\sum_{k=1}^{n} \xi_{n,k}\right)_n$ is said to converge weakly to ξ iff

$$\lim_{n\to\infty} \mathbb{P}\left[\sum_{k=1}^{n} \xi_{n,k} \leq x_0\right] = \mathbb{P}[\xi \leq x_0]$$

Ben Berckmoes is post doctoral fellow at the Fund for Scientific Research of Flanders (FWO).

Geert Molenberghs gratefully acknowledges financial support from the IAP research network #P7/06 of the Belgian Government (Belgian Science Policy).

for all x_0 at which the map $x \mapsto \mathbb{P}[\xi \leq x]$ is continuous, or, equivalently, iff

$$\lim_{n\to\infty} \mathbb{E}\left[h\left(\sum_{k=1}^{n} \xi_{n,k}\right)\right] = \mathbb{E}[h(\xi)]$$

for all $h : \mathbb{R} \to \mathbb{R}$ bounded and continuous.

We say that $\{\xi_{n,k}\}$ satisfies Feller's condition iff

$$\lim_{n\to\infty} \max_{k=1}^{n} \mathbb{E}\left[\xi_{n,k}^2\right] = 0, \tag{1}$$

and Lindeberg's condition iff

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[\xi_{n,k}^2; |\xi_{n,k}| > \epsilon\right] = 0$$

for all $\epsilon > 0$. It is easily seen that Lindeberg's condition implies Feller's, but that the converse does not hold.

The above language allows us to formulate the following result, which belongs to the heart of classical probability theory.

Theorem 1.1 (Lindeberg-Feller Central Limit Theorem)**.** *Let ξ and $\{\xi_{n,k}\}$ be as above. If $\{\xi_{n,k}\}$ satisfies Lindeberg's condition, then the sequence $(\sum_{k=1}^{n} \xi_{n,k})_n$ converges weakly to ξ. The converse holds if $\{\xi_{n,k}\}$ satisfies Feller's condition.*

The number

$$\mathrm{Lin}\left(\{\xi_{n,k}\}\right) = \sup_{\epsilon>0} \limsup_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[\xi_{n,k}^2; |\xi_{n,k}| > \epsilon\right] \tag{2}$$

was introduced in [BLV13] as the Lindeberg index. Notice that it produces for each STA a number between 0 and 1, and that it is 0 if and only if Lindeberg's condition is satisfied. So it could be thought of as a number which measures how far a given STA deviates from satisfying Lindeberg's condition.

Furthermore, let $d(\eta, \eta')$ be a metric on random variables with the property that $\lim_{n\to\infty} d(\eta, \eta_n) = 0$ is equivalent with weak convergence of $(\eta_n)_n$ to η, and define the quantity

$$\lambda_d\left(\sum_{k=1}^{n} \xi_{n,k} \to \xi\right) = \limsup_{n\to\infty} d\left(\xi, \sum_{k=1}^{n} \xi_{n,k}\right). \tag{3}$$

Clearly, (3) assigns a positive number to each STA which is 0 if and only if the row-wise sums of the STA are asymptotically normal. Thus this number measures how far a given STA deviates from having an asymptotically normal sequence of row-wise sums.

Now, using the numbers (2) and (3), the first part of Theorem 1.1 boils down to the implication

$$\mathrm{Lin}(\{\xi_{n,k}\}) = 0 \Rightarrow \lambda_d \left(\sum_{k=1}^{n} \xi_{n,k} \to \xi \right) = 0.$$

Observe that Theorem 1.1 fails to provide any information for the large class of STA's which fail to satisfy Lindeberg's condition, regardless of whether $\mathrm{Lin}(\{\xi_{n,k}\})$ is large or small. Thus the following natural question arises.

Question 1.2. *Suppose that we are given an STA $\{\xi_{n,k}\}$ which is close to satisfying Lindeberg's condition in the sense that $\mathrm{Lin}(\{\xi_{n,k}\})$ is non-zero but small. Is it still possible to conclude that the row-wise sums of $\{\xi_{n,k}\}$ are close to being asymptotically normal in the sense that $\lambda_d \left(\sum_{k=1}^{n} \xi_{n,k} \to \xi\right)$ is small?*

Let us briefly describe how in the case where d is the Kolmogorov metric

$$K(\eta, \eta') = \sup_{x \in \mathbb{R}} \left| \mathbb{P}[\eta \le x] - \mathbb{P}[\eta' \le x] \right|,$$

a positive answer to Question 1.2 can be derived from the existing literature.

The following refinement of the sufficiency of Lindeberg's condition in Theorem 1.1 was obtained in terms of the Kolmogorov distance in [O66] and [F68].

Theorem 1.3. *Let ξ be as above. Then there exists a universal constant $C > 0$ such that*

$$K\left(\xi, \sum_{k=1}^{n} \xi_{n,k}\right) \le C \left(\sum_{k=1}^{n} \mathbb{E}\left[\xi_{n,k}^2; |\xi_{n,k}| > 1\right] + \sum_{k=1}^{n} \mathbb{E}\left[|\xi_{n,k}|^3; |\xi_{n,k}| \le 1\right] \right)$$

for all STA's $\{\xi_{n,k}\}$ and all n.

It was shown in [F68] that the constant C in Theorem 1.3 can be taken equal to 6. A proof of Theorem 1.3 based on Stein's method was given in [BH84], and in [CS01], combining Stein's method with Chen's concentration inequality approach, it was established that C can be taken equal to 4.1, the best value known so far up to our knowledge.

We will infer a corollary from Theorem 1.3 which is related to Question 1.2. To this end, we remark that it was pointed out in [L75] that the truncation at 1 in Theorem 1.3 is optimal in the sense that

$$\sum_{k=1}^{n} \mathbb{E}\left[\xi_{n,k}^2; |\xi_{n,k}| > 1\right] + \sum_{k=1}^{n} \mathbb{E}\left[|\xi_{n,k}|^3; |\xi_{n,k}| \le 1\right]$$

is dominated by

$$\sum_{k=1}^{n} \mathbb{E}\left[\xi_{n,k}^2; \xi_{n,k} \in A\right] + \sum_{k=1}^{n} \mathbb{E}\left[|\xi_{n,k}|^3; \xi_{n,k} \in \mathbb{R} \setminus A\right]$$

for each Borel set $A \subset \mathbb{R}$. Therefore, we easily derive from Theorem 1.3 that

$$K\left(\xi, \sum_{k=1}^{n} \xi_{n,k}\right) \leq C\left(\sum_{k=1}^{n} \mathbb{E}\left[\xi_{n,k}^{2}; |\xi_{n,k}| > \epsilon\right] + \epsilon\right)$$

for all $\epsilon > 0$, which, calculating the superior limit of both sides and letting $\epsilon \downarrow 0$, yields

$$\limsup_{n\to\infty} K\left(\xi, \sum_{k=1}^{n} \xi_{n,k}\right) \leq C \sup_{\epsilon>0} \limsup_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[\xi_{n,k}^{2}; |\xi_{n,k}| > \epsilon\right].$$

Using the numbers defined in (2) and (3), we now derive the following result as a corollary of Theorem 1.3.

Theorem 1.4. *Let ξ be as above. Then there exists a universal constant $C > 0$ such that*

$$\lambda_K\left(\sum_{k=1}^{n} \xi_{n,k} \to \xi\right) \leq C\mathrm{Lin}\left(\{\xi_{n,k}\}\right)$$

for all STA's $\{\xi_{n,k}\}$.

Remark 1.5. *In* [BLV13]*, combining Stein's method with an asymptotic smoothing technique, it was established that the constant C in Theorem 1.4 can be taken equal to 1 if $\{\xi_{n,k}\}$ satisfies Feller's condition.*

Notice that Theorem 1.4 gives a positive answer to Question 1.2 in the case where $d = K$. It strictly generalizes the sufficiency of Lindeberg's condition in Theorem 1.1, and, contrary to Theorem 1.1, it continues to provide useful information for STA's which have a low Lindeberg index, but which fail to satisfy Lindeberg's condition. More precisely, it allows us to conclude that $\left(\sum_{k=1}^{n} \xi_{n,k}\right)_n$ is approximately convergent to ξ if $\{\xi_{n,k}\}$ approximately satisfies Lindeberg's condition. Therefore, it seems plausible to refer to Theorem 1.4 as an approximate central limit theorem.

The problem of generalizing Theorem 1.3 to the multivariate setting is hard, and remains open. Notice however that recently, combining a multivariate version of Stein's method as outlined in e.g. [M09] and [NPR10] with the establishment of an explicit integral representation of a solution to the Stein PDE with a character function as test function, a partial extension of Theorem 1.4 for the Fourier transforms of random vectors has been obtained in [BLV16].

In this paper, we will focus on the following two questions concerning Theorem 1.4.

1) Can we widen the scope of applicability of Theorem 1.4 by extending it to other probability metrics d?

2) Are there canonical situations in mathematical statistics in which it is natural to invoke (an extension of) Theorem 1.4?

In Section 2, we will address the first question. More precisely, we will show that it is possible to extend the techniques used in [BLV13] to a large class of test functions, leading to a general inequality. This inequality will in turn entail approximate central limit theorems, similar to Theorem 1.4, for the Wasserstein distance, and even for the class of parametrized Prokhorov distances. A natural example will show that it is impossible to obtain a result of the same flavor for the total variation distance.

In Section 3, we will provide an answer to the second question. More precisely, we will investigate the asymptotic behavior of the sample average estimator, based on independent data which are contaminated according to what is called the inflated variance model. We will establish the weak consistency of this estimator under a fairly general condition. Moreover, it is shown that in many cases, the asymptotic normality of this estimator can only be studied by relying on an approximate central limit theorem such as Theorem 1.4. The theoretical results obtained in this section will be illustrated by an example and a simulation study.

Some open questions are formulated in Section 4.

2. Approximate central limit theory

2.1. **Some probability metrics.** Let us start by giving a short description of the probability metrics that will play an important role in the remainder of this paper.

Let $\mathcal{P}(\mathbb{R})$ be the collection of Borel probability measures on $\mathbb{R}$. Furthermore, let $\mathcal{P}_1(\mathbb{R})$ be the set of all $P \in \mathcal{P}(\mathbb{R})$ with finite absolute first moment, i.e. for which $\int_{-\infty}^{\infty} |x|\, dP(x) < \infty$.

The Wasserstein distance on $\mathcal{P}_1(\mathbb{R})$, see e.g. [V03], is defined by the formula

$$W(P,Q) = \inf_{\pi} \int_{\mathbb{R}\times\mathbb{R}} d(x,y) d\pi(x,y),$$

where the infimum is taken over all Borel probability measures π on $\mathbb{R} \times \mathbb{R}$ with first marginal P and second marginal Q. Kantorovich duality theory implies that the metric W can also be written as

$$W(P,Q) = \sup_{h\in\mathcal{K}(\mathbb{R})} \left| \int_{\mathbb{R}} hdP - \int_{\mathbb{R}} hdQ \right|, \tag{4}$$

where $\mathcal{K}(\mathbb{R})$ stands for the set of all contractions $h : \mathbb{R} \to \mathbb{R}$, where h is called a contraction iff $|h(x) - h(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}$. Also, we have

$$W(P,Q) = \int_{-\infty}^{\infty} |F_P(x) - F_Q(x)|\, dx = \int_0^1 \left|F_P^{-1}(t) - F_Q^{-1}(t)\right| dt,$$

with F_P (respectively F_Q) the cumulative distribution function associated with P (respectively Q), and F_P^{-1} (respectively F_Q^{-1}) its generalized inverse.

The topology underlying the Wasserstein distance is slightly stronger than the weak topology. More precisely, for P and $(P_n)_n$ in $\mathcal{P}_1(\mathbb{R})$, it holds that

$$\lim_{n\to\infty} W(P, P_n) = 0$$

is equivalent with weak convergence of $(P_n)_n$ to P in addition to convergence of $\left(\int_{-\infty}^{\infty} |x|\, dP_n(x)\right)_n$ to $\int_{-\infty}^{\infty} |x|\, dP(x)$. Also, the Wasserstein distance is separable and complete, see [B08].

Furthermore, for $\lambda \in \mathbb{R}_0^+$, the (parametrized) Prokhorov distance $\rho_\lambda(P, Q)$ between probability measures P and Q in $\mathcal{P}(\mathbb{R})$ is defined to be the infimum of all positive numbers $\alpha \in \mathbb{R}_0^+$ for which the inequality

$$P[A] \leq Q\left[A^{(\lambda\alpha)}\right] + \alpha,$$

with

$$A^{(\lambda\alpha)} = \left\{x \in \mathbb{R} \mid \inf_{a\in A} |x - a| \leq \lambda\alpha\right\},$$

holds for every Borel set $A \subset \mathbb{R}$. One easily establishes that

$$\rho_{\lambda_1}(P, Q) \leq \rho_{\lambda_2}(P, Q)$$

whenever $\lambda_2 \leq \lambda_1$. In [B99] it is shown that, for each $\lambda \in \mathbb{R}_0^+$, ρ_λ is a separable and complete metric which metrizes weak convergence of probability measures.

Finally, the total variation distance $d_{TV}(P, Q)$ between probability measures P and Q in $\mathcal{P}(\mathbb{R})$ is defined by the number

$$d_{TV}(P, Q) = \sup_A |P[A] - Q[A]|,$$

the supremum of course taken over all Borel sets $A \subset \mathbb{R}$. One easily verifies that d_{TV} is a complete metric, that, for each $\lambda \in \mathbb{R}_0^+$,

$$\rho_\lambda(P, Q) \leq d_{TV}(P, Q),$$

and that the limit relation

$$\lim_{\lambda\downarrow 0} \rho_\lambda(P, Q) = d_{TV}(P, Q) \tag{5}$$

holds true. Note however that d_{TV} is not separable and that its underlying topology is strictly stronger than the weak topology.

For a general and systematic treatment of the theory of probability metrics, we refer the reader to the excellent expositions [Z83] and [R91].

2.2. **A general inequality.** Let ξ be as in Section 1 and $h : \mathbb{R} \to \mathbb{R}$ a continuous map for which $\mathbb{E}\,|h(\xi)| < \infty$. Then the Stein transform of h is the map $f_h : \mathbb{R} \to \mathbb{R}$ defined by the formula

$$f_h(x) = e^{x^2/2} \int_{-\infty}^{x} (h(t) - \mathbb{E}[h(\xi)])\, e^{-t^2/2} dt. \tag{6}$$

The crux of Stein's method is that, for any random variable η, we have

$$\mathbb{E}\,[h(\xi) - h(\eta)] = \mathbb{E}[\eta f_h(\eta) - f_h'(\eta)],$$

and that, in many cases, it is easier to find upper bounds for the derivatives of f_h than for the derivatives of h, see e.g. [BC05] and [CGS11].

We will now establish a general inequality in terms of the Stein transform, which will allow us to extend Theorem 1.4 to many of the above described probability metrics. For the proof, it basically suffices to notice that the techniques developed in [BLV13] can be extended to a very general collection of test functions. For the sake of completeness, we present the proof in Appendix A.

Theorem 2.1. *Let ξ and $\{\xi_{n,k}\}$ be as in Section 1, and let $h : \mathbb{R} \to \mathbb{R}$ be any continuously differentiable map with a bounded derivative. Then the Stein transform f_h, defined by (6), is twice continuously differentiable, has bounded first and second derivatives, and the inequality*

$$\left|\mathbb{E}\left[h(\xi) - h\left(\sum_{k=1}^{n} \xi_{n,k}\right)\right]\right| \tag{7}$$

$$\leq \frac{1}{2}\|f_h''\|_\infty \epsilon + \left(\sup_{x_1,x_2\in\mathbb{R}} |f_h'(x_1) - f_h'(x_2)|\right) \sum_{k=1}^{n} \mathbb{E}\left[\xi_{n,k}^2; |\xi_{n,k}| \geq \epsilon\right]$$

$$+ \left(\sup_{x_1,x_2\in\mathbb{R}} |f_h''(x_1) - f_h''(x_2)|\right) \max_{k=1}^{n} \mathbb{E}\left[|\xi_{n,k}|\right]$$

holds for all n and all $\epsilon > 0$.

2.3. **Approximate central limit theorems.** We will apply Theorem 2.1 to obtain results similar to Theorem 1.4 for the Wasserstein distance (Theorem 2.4) and the parametrized Prokhorov distances (Theorem 2.8). Where needed, we tacitly transport these probability metrics to random variables via their image measures.

The following lemma guarantees that we can capture the Wasserstein distance with continuously differentiable contractions.

Lemma 2.2. *The Wasserstein distance on $\mathcal{P}_1(\mathbb{R})$ is given by*

$$W(P,Q) = \sup_{h\in\mathcal{K}_c(\mathbb{R})} \left|\int h dP - \int h dQ\right|, \tag{8}$$

where $\mathcal{K}_c(\mathbb{R})$ stands for the set of all continuously differentiable contractions $h : \mathbb{R} \to \mathbb{R}$.

Proof. Let $h : \mathbb{R} \to \mathbb{R}$ be a contraction and fix $\epsilon > 0$. We will show that there exists a smooth contraction which is closer than ϵ to h for the $\|\cdot\|_\infty$-norm. Once this is established, the lemma will follow from formula (4).

Let

$$\psi_\epsilon : \mathbb{R} \to \mathbb{R}$$

be positive and smooth, with support contained in the interval $[-\epsilon, \epsilon]$, and such that

$$\int_{\mathbb{R}} \psi_\epsilon(y)dy = 1.$$

Put

$$h_\epsilon(x) = (h \star \psi_\epsilon)(x) = \int_{\mathbb{R}} h(x-y)\psi_\epsilon(y)dy = \int_{\mathbb{R}} \psi_\epsilon(x-y)h(y)dy.$$

Then h_ϵ is smooth. Furthermore, for $x_1, x_2 \in \mathbb{R}$,

$$\begin{aligned} &|h_\epsilon(x_1) - h_\epsilon(x_2)| \\ &= \left| \int_{\mathbb{R}} h(x_1 - y)\psi_\epsilon(y)dy - \int_{\mathbb{R}} h(x_2 - y)\psi_\epsilon(y)dy \right| \\ &\leq \int_{\mathbb{R}} |h(x_1 - y) - h(x_2 - y)|\, \psi_\epsilon(y)dy, \end{aligned}$$

which is, h being a contraction, bounded by $\int_{\mathbb{R}} \psi_\epsilon(y)dy = 1$, and we infer that h_ϵ is also a contraction. Finally, for $x \in \mathbb{R}$,

$$\begin{aligned} &|h(x) - h_\epsilon(x)| \\ &= \left| \int_{\mathbb{R}} (h(x) - h(x-y))\, \psi_\epsilon(y)dy \right| \\ &= \left| \int_{-\epsilon}^{\epsilon} (h(x) - h(x-y))\, \psi_\epsilon(y)dy \right|, \end{aligned} \tag{9}$$

(10)

the last equality following from the fact that the support of ψ_ϵ is contained in $[-\epsilon, \epsilon]$. Now, h being a contraction, it follows that the expression in (9) is bounded by ϵ, whence

$$\|h - h_\epsilon\|_\infty < \epsilon.$$

This concludes the proof. □

The following lemma belongs to the folklore of Stein's method, see e.g. [BC05], p.10-11.

Lemma 2.3. *Let h and f_h be as in Theorem 2.1. Then*

$$\|f_h'\|_\infty \leq 4\|h'\|_\infty \tag{11}$$

and

$$\|f_h'\|_\infty \leq 2\|\mathbb{E}[h(\xi)] - h\|_\infty \tag{12}$$

and

$$\|f_h''\|_\infty \leq 2\|h'\|_\infty. \tag{13}$$

Theorem 2.4. *Let ξ be as in Section 1. Then there exists a universal constant $C_W > 0$ such that*

$$\lambda_W\left(\sum_{k=1}^n \xi_{n,k} \to \xi\right) \leq C_W \mathrm{Lin}\left(\{\xi_{n,k}\}\right)$$

for all STA's $\{\xi_{n,k}\}$ which satisfy Feller's condition (1). Moreover, C_W can be taken equal to 8.

Proof. Let $h : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable contraction. Then

$$\sup_{x_1,x_2\in\mathbb{R}} |f_h'(x_1) - f_h'(x_2)| \leq 2\|f_h'\|_\infty \leq 8\|h'\|_\infty \leq 8, \tag{14}$$

by (11), and

$$\sup_{x_1,x_2\in\mathbb{R}} |f_h''(x_1) - f_h''(x_2)| \leq 2\|f_h''\|_\infty \leq 4\|h'\|_\infty \leq 4, \tag{15}$$

by (13). Furthermore, combining (7) with (13), (14), and (15), yields

$$\begin{aligned} &\left|\mathbb{E}\left[h(\xi) - h\left(\sum_{k=1}^n \xi_{n,k}\right)\right]\right| \qquad (16)\\ &\leq \epsilon + 8\sum_{k=1}^n \mathbb{E}\left[\xi_{n,k}^2; |\xi_{n,k}| \geq \epsilon\right] + 4\max_{k=1}^n \mathbb{E}\left[|\xi_{n,k}|\right] \end{aligned}$$

for all STA's $\{\xi_{n,k}\}$, all n, and all $\epsilon > 0$. Finally, assuming that $\{\xi_{n,k}\}$ satisfies Feller's condition, taking the supremum over all $h \in \mathcal{K}_c(\mathbb{R})$, calculating the superior limits, and letting $\epsilon \downarrow 0$, we see that that (8) and (16) lead to the desired result. □

Lemma 2.7 reveals that we can capture all parametrized Prokhorov distances by one collection of smooth test functions. Its proof will be based on two additional lemma's.

Lemma 2.5. *We have, for P and $(P_n)_n$ in $\mathcal{P}(\mathbb{R})$,*

$$\sup_{\lambda>0}\limsup_{n\to\infty}\rho_\lambda(P, P_n) = \sup_{\alpha>0}\limsup_{n\to\infty}\sup_A\left(P[A] - P_n\left[A^{(\alpha)}\right]\right), \tag{17}$$

the second supremum of the right-hand side taken over all Borel sets $A \subset \mathbb{R}$.

Proof. Suppose that the right-hand side of (17) is dominated by $\theta > 0$. Then, for any $\lambda > 0$, there exists n such that for all $k \geq n$ and all Borel sets $A \subset \mathbb{R}$, $P[A] \leq P_k[A^{(\lambda\theta)}] + \theta$, that is, $\rho_\lambda(P, P_k) \leq \theta$. We conclude that also the left-hand side of (17) is dominated by θ.

Conversely, suppose that the right-hand side of (17) is larger than $\theta > 0$. Then there exists $\alpha > 0$ such that for each n there exist $k \geq n$ and $A \subset \mathbb{R}$ Borel such that, putting $\lambda = \alpha\theta^{-1}$, $P[A] \geq P_k[A^{(\lambda\theta)}] + \theta$, that is, $\rho_\lambda(P, P_k) > \theta$. We infer that also the left-hand side of (17) is larger than θ. □

The following lemma was established in full generality in [BLV11] (Section 2, Lemma 2.2) to study the so-called weak approach structure on the set of probability measures on a separable metric space, see also [L15]. It is a crucial step to become an approximate central limit theorem for the parametrized Prokhorov distances. The fact that we can take sufficiently smooth maps is easily seen in the proof given there.

Lemma 2.6. *For each $\alpha > 0$ and each $\epsilon > 0$ there exists a finite collection $\mathcal{H}_0$ of continuously differentiable maps $h : \mathbb{R} \to [0,1]$ with bounded first derivative, such that for all $Q \in \mathcal{P}(\mathbb{R})$*

$$\sup_A \left(P[A] - Q[A^{(\alpha)}]\right) \leq \max_{h\in\mathcal{H}_0} \left| \int_{\mathbb{R}} hdP - \int_{\mathbb{R}} hdQ \right| + \epsilon, \tag{18}$$

the first supremum taken over all Borel sets $A \subset \mathbb{R}$.

Conversely, for each continuous map $h : \mathbb{R} \to [0,1]$ and each $\epsilon > 0$ there exists $\alpha > 0$ such that for all $Q \in \mathcal{P}(\mathbb{R})$

$$\left| \int_{\mathbb{R}} hdP - \int_{\mathbb{R}} hdQ \right| \leq \sup_A \left(P[A] - Q[A^{(\alpha)}]\right) + \epsilon,$$

the supremum again taken over all Borel sets $A \subset \mathbb{R}$.

Lemma 2.7. *Let $\mathcal{H}$ be the collection of continuously differentiable maps $h : \mathbb{R} \to [0,1]$ with a bounded derivative. Then, for P and $(P_n)_n$ in $\mathcal{P}(\mathbb{R})$,*

$$\sup_{\lambda>0} \limsup_{n\to\infty} \rho_\lambda(P, P_n) = \sup_{h\in\mathcal{H}} \limsup_{n\to\infty} \left| \int_{\mathbb{R}} hdP - \int_{\mathbb{R}} hdP_n \right|. \tag{19}$$

Proof. By Lemma 2.5, it suffices to show that the right-hand side of (19) equals the right-hand side of (17).

Fix α and $\epsilon > 0$, and choose, according to the first assertion in Lemma 2.6, a finite collection $\mathcal{H}_0$ of continuously differentiable maps $h : \mathbb{R} \to [0,1]$ with bounded derivative, such that (18) holds. But then

$$\begin{aligned} &\limsup_{n\to\infty} \sup_A \left(P[A] - P_n\left[A^{(\alpha)}\right]\right) \\ &\leq \limsup_{n\to\infty} \max_{h\in\mathcal{H}_0} \left| \int_{\mathbb{R}} hdP - \int_{\mathbb{R}} hdP_n \right| + \epsilon \\ &= \max_{h\in\mathcal{H}_0} \limsup_{n\to\infty} \left| \int_{\mathbb{R}} hdP - \int_{\mathbb{R}} hdP_n \right| + \epsilon, \end{aligned}$$

the last equality following from the finiteness of $\mathcal{H}_0$. This shows that the right-hand side of (19) dominates the right-hand side of (17).

The converse inequality follows analogously from the second assertion in Lemma 2.6. □

Theorem 2.8. *Let ξ be as in Section 1. Then there exists a universal constant $C_P > 0$ such that*

$$\lambda_{\rho_\lambda}\left(\sum_{k=1}^{n} \xi_{n,k} \to \xi\right) \leq C_P \mathrm{Lin}\left(\{\xi_{n,k}\}\right)$$

for all $\lambda > 0$ and all STA's $\{\xi_{n,k}\}$ which satisfy Feller's condition (1). Moreover, C_P can be taken equal to 4.

Proof. Let $h : \mathbb{R} \to [0,1]$ be a continuously differentiable map with a bounded derivative. Then

$$\sup_{x_1,x_2\in\mathbb{R}} |f_h'(x_1) - f_h'(x_2)| \leq 2\|f_h'\|_\infty \leq 4\|\mathbb{E}[h(\xi)] - h\|_\infty \leq 4, \tag{20}$$

by (12), and

$$\sup_{x_1,x_2\in\mathbb{R}} |f_h''(x_1) - f_h''(x_2)| \leq 2\|f_h''\|_\infty \leq 4\|h'\|_\infty, \tag{21}$$

by (13). Furthermore, combining (7) with (13), (20), and (21), yields

$$\begin{aligned} &\left|\mathbb{E}\left[h(\xi) - h\left(\sum_{k=1}^{n} \xi_{n,k}\right)\right]\right| \qquad (22) \\ \leq\ & \|h'\|_\infty \epsilon + 4\sum_{k=1}^{n} \mathbb{E}\left[\xi_{n,k}^2; |\xi_{n,k}| \geq \epsilon\right] + 4\|h'\|_\infty \max_{k=1}^{n} \mathbb{E}\left[|\xi_{n,k}|\right] \end{aligned}$$

for all STA's $\{\xi_{n,k}\}$, all n, and all $\epsilon > 0$. Finally, assuming that $\{\xi_{n,k}\}$ satisfies Feller's condition, calculating the superior limits, and letting $\epsilon \downarrow 0$, we see that that (19) and (22) lead to the desired result. □

Notice the remarkable fact that the constant C_P in Theorem 2.8 does not depend on the parameter λ. This, in light of relation (5), suggests that an approximate central limit theorem in the spirit of Theorem 1.4 for the total variation distance d_{TV} might be derived from Theorem 2.8. However, the following example shows that this is not the case.

Example 2.9. *Let ξ and $\{\xi_{n,k}\}$ be as in Section 1, and assume that $\{\xi_{n,k}\}$ consists of discrete random variables and satisfies Lindeberg's condition. Then*

$$\mathrm{Lin}(\{\xi_{n,k}\}) = 0$$

and, each $\sum_{k=1}^{n} \xi_{n,k}$ also being discrete,

$$d_{TV}\left(\xi, \sum_{k=1}^{n} \xi_{n,k}\right) = 1.$$

We conclude that there does not exist a constant $C > 0$ such that

$$\lambda_{d_{TV}}\left(\sum_{k=1}^{n} \xi_{n,k} \to \xi\right) \leq C\mathrm{Lin}(\{\xi_{n,k}\})$$

for all STA's $\{\xi_{n,k}\}$ satisfying Feller's condition.

We summarize the information obtained in Theorem 1.4 and Remark 1.5, Theorem 2.4, and Theorem 2.8, in the following result. We put

$$\lambda_P\left(\sum_{k=1}^{n}\xi_{n,k}\to\xi\right)=\sup_{\lambda\in\mathbb{R}_0^+}\lambda_{\rho_\lambda}\left(\sum_{k=1}^{n}\xi_{n,k}\to\xi\right).$$

Theorem 2.10. *Let ξ be as in Section 1. Then, for each $\delta\in\{K,W,P\}$, there exists a universal constant $C_\delta>0$ such that*

$$\lambda_\delta\left(\sum_{k=1}^{n}\xi_{n,k}\to\xi\right)\le C_\delta\mathrm{Lin}\left(\{\xi_{n,k}\}\right)$$

for all STA's $\{\xi_{n,k}\}$ satisfying Feller's condition (1). Moreover, C_K can be taken equal to 1, C_W equal to 8, and C_P equal to 4.

3. Statistical inference with contaminated data

3.1. **Motivation and framework.** We will now consider at a statistical situation in which it is beneficial to rely on the approximate central limit theory developed in the previous section.

Let F be a cumulative distribution function on the real line with

$$\int_{-\infty}^{\infty}xdF(x)=0$$

and

$$\int_{-\infty}^{\infty}x^2dF(x)=1.$$

Fix $\mu\in\mathbb{R}$ and let X_1, X_2, …, X_k, … be independent observations of $F(\cdot-\mu)$ which are contaminated according to the inflated variance model (see [TSM85], p.108), that is

$$X_k\sim(1-p_k)F(\cdot-\mu)+p_kF\left(\frac{\cdot-\mu}{\sigma_k}\right),$$

where $p_k\in[0,1]$ and $\sigma_k\in[1,\infty[$. Notice that this is a natural model to express the fact that for each observation X_k there is a large probability $1-p_k$ of observing the correct target distribution $F(\cdot-\mu)$, and a small probability p_k of observing a contaminant with distribution $F(\frac{\cdot-\mu}{\sigma_k})$. Observe that

$$\mathbb{E}[X_k]=\mu$$

and

$$\mathrm{Var}[X_k]=(1-p_k)+p_k\sigma_k^2.$$

Now define the sample mean in the usual way as

$$\overline{X}_n=\frac{1}{n}\sum_{k=1}^{n}X_k.$$

Notice that

$$\mathbb{E}[\overline{X}_n]=\mu$$

and

$$\mathrm{Var}[\overline{X}_n] = \left(\frac{s_n}{n}\right)^2$$

where

$$s_n^2 = \sum_{k=1}^{n} [(1-p_k) + p_k \sigma_k^2].$$

Also,

$$s_n^2 \geq n \tag{23}$$

because $\sigma_k^2 \geq 1$ for all k. Here we will investigate up to what extent the estimator $\overline{X}_n$ is weakly consistent for μ in the sense that

$$\overline{X}_n \xrightarrow{\mathbb{P}} \mu, \tag{24}$$

i.e.

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\mu - \overline{X}_n\right| \geq \epsilon\right] = 0$$

for all $\epsilon > 0$, and asymptotically normal in the sense that

$$\frac{n}{s_n}\left(\overline{X}_n - \mu\right) \xrightarrow{w} \xi, \tag{25}$$

with ξ as in Section 1, i.e. $\left(\frac{n}{s_n}\left(\overline{X}_n - \mu\right)\right)_n$ converges weakly to ξ. Notice that in the uncontaminated case where $\sigma_k = 1$ for all k, the Weak Law of Large Numbers implies the truth of (24) and the Central Limit Theorem entails the validity of (25). It will turn out that in the contaminated case, the asymptotic normality is often only tractable by an approximate central limit theorem such as Theorem 2.10.

3.2. **Consistency and asymptotic normality.** We keep the terminology and the notation from above.

The following relatively straightforward result shows that the contaminated sample mean is weakly consistent under a fairly mild condition.

Theorem 3.1. *Suppose that*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^{n} p_k \sigma_k^2 = 0. \tag{26}$$

Then

$$\overline{X}_n \xrightarrow{\mathbb{P}} \mu.$$

Proof. Assume without loss of generality that $\mu = 0$. For $\epsilon > 0$, Chebyshev's inequality gives

$$\begin{aligned} \mathbb{P}\left[\left|\overline{X}_n\right| \geq \epsilon\right] &\leq \frac{1}{\epsilon^2}\mathrm{Var}[\overline{X}_n] \\ &= \frac{1}{\epsilon^2}\left[\frac{1}{n^2}\sum_{k=1}^{n}(1-p_k) + \frac{1}{n^2}\sum_{k=1}^{n} p_k\sigma_k^2\right], \end{aligned}$$

which easily implies that

$$\limsup_{n\to\infty} \mathbb{P}\left[\left|\overline{X}_n\right| \geq \epsilon\right] \leq \frac{1}{\epsilon^2} \limsup_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^{n} p_k \sigma_k^2.$$

This finishes the proof. □

We now turn to the asymptotic normality of $\overline{X}_n$. It turns out that the STA $\left\{\frac{1}{s_n}(X_k-\mu)\right\}$, which is of crucial importance, satisfies Lindeberg's condition if the sequence of contaminating variances $(\sigma_k)_k$ is controllable in a sense made precise in the following theorem.

Theorem 3.2. *Suppose that*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \max_{k=1}^{n} \sigma_k^2 \to 0. \tag{27}$$

Then the STA $\left\{\frac{1}{s_n}(X_k-\mu)\right\}$ *satisfies Lindeberg's condition, i.e.*

$$\mathrm{Lin}\left(\left\{\frac{1}{s_n}(X_k-\mu)\right\}\right) = 0.$$

Proof. Assume without loss of generality that $\mu = 0$ and let X be a random variable with cumulative distribution function F. Then, for $\epsilon > 0$,

$$\begin{aligned}
&\sum_{k=1}^{n} \mathbb{E}\left[\left(\frac{1}{s_n}X_k\right)^2; \left|\frac{1}{s_n}X_k\right| \geq \epsilon\right] \\
&= \frac{1}{s_n^2}\sum_{k=1}^{n}(1-p_k)\mathbb{E}\left[X^2; |X| \geq \epsilon s_n\right] + \frac{1}{s_n^2}\sum_{k=1}^{n} p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right],
\end{aligned}$$

which is

$$\begin{aligned}
\leq\ & \frac{1}{s_n^2}\sum_{k=1}^{n}(1-p_k)\mathbb{E}\left[X^2; |X| \geq \epsilon s_n\right] \\
& + \frac{1}{s_n^2}\sum_{k=1}^{n} p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \epsilon\sqrt{\frac{s_n^2}{\max_{k=1}^{n}\sigma_k^2}}\right] \\
\leq\ & \mathbb{E}[X^2; |X| \geq \epsilon s_n] + \mathbb{E}\left[X^2; |X| \geq \epsilon\sqrt{\frac{s_n^2}{\max_{k=1}^{n}\sigma_k^2}}\right].
\end{aligned}$$

The latter quantity converges to 0 as n tends to ∞ by (23) and (27). This finishes the proof. □

Remark 3.3. *Observe that (27) implies (26).*

The classical central limit theory now leads to the following result.

Theorem 3.4. *Let ξ be as in Section 1 and suppose that*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \max_{k=1}^{n} \sigma_k^2 \to 0.$$

Then

$$\frac{n}{s_n}\left(\overline{X}_n - \mu\right) \xrightarrow{w} \xi.$$

Proof. Notice that the n-th rowwise sum of the STA $\left\{\frac{1}{s_n}\left(X_k - \mu\right)\right\}$ coincides with $\frac{n}{s_n}\left(\overline{X}_n - \mu\right)$. Now apply Theorem 3.2 and Theorem 1.1. □

If the sequence $(\sigma_k)_k$ cannot be controlled by condition (27), then it is more appropriate to make use of the approximate central limit theory of Section 2. As Feller's condition plays an important role in this theory, we start with the following characterization.

Theorem 3.5. *The STA $\left\{\frac{1}{s_n}(X_k - \mu)\right\}$ satisfies Feller's condition if and only if*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \max_{k=1}^{n} p_k\sigma_k^2 = 0. \tag{28}$$

Proof. Assume without loss of generality that $\mu = 0$. Now

$$\max_{k=1}^{n} \mathbb{E}\left[\frac{1}{s_n^2} X_k^2\right] = \frac{1}{s_n^2} \max_{k=1}^{n} (1 - p_k) + \frac{1}{s_n^2} \max_{k=1}^{n} p_k\sigma_k^2$$

whence, by (23),

$$\limsup_{n\to\infty} \max_{k=1}^{n} \mathbb{E}\left[\frac{1}{s_n^2} X_k^2\right] = \limsup_{n\to\infty} \frac{1}{s_n^2} \max_{k=1}^{n} p_k\sigma_k^2.$$

This finishes the proof. □

Remark 3.6. *Observe that (27) implies (28), which in turn implies (26).*

Theorems 3.8 and 3.9 will reveal that even in the absence of condition (27), the Lindeberg index of the STA $\left\{\frac{1}{s_n}\left(X_k - \mu\right)\right\}$ can still be bounded from above. Moreover, it can be explicitly computed under a fairly easy set of conditions.

We first need the following lemma.

Lemma 3.7. *Suppose that the sequence $\left(\frac{1}{n}\sum_{k=1}^{n} p_k\sigma_k^2\right)_n$ is bounded and let X be a random variable with cumulative distribution function F. Then*

$$\begin{aligned} &\operatorname{Lin}\left(\left\{\frac{1}{s_n}(X_k - \mu)\right\}\right) \\ &= \sup_{\gamma>0} \sup_{\epsilon>0} \limsup_{n\to\infty} \frac{1}{s_n^2} \sum_{k=\lceil \gamma n\rceil}^{n} p_k\sigma_k^2 \mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right], \end{aligned}$$

where $\lceil\cdot\rceil$ is the ceiling function.

Proof. Assume w.l.o.g. that $\mu = 0$ and choose $K \in \mathbb{R}_0^+$ such that for all n

$$\frac{1}{n}\sum_{k=1}^{n} p_k\sigma_k^2 \leq K. \tag{29}$$

Next, fix $\gamma > 0$ small. Then, for n large, by (23) and (29),

$$\begin{aligned} &\frac{1}{s_n^2}\sum_{k=1}^{\lceil\gamma n\rceil-1} p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right] \\ \leq\ & \gamma\frac{1}{\gamma n}\sum_{k=1}^{\lceil\gamma n\rceil-1} p_k\sigma_k^2 \\ \leq\ & \gamma\frac{1}{\lceil\gamma n\rceil-1}\sum_{k=1}^{\lceil\gamma n\rceil-1} p_k\sigma_k^2 \\ \leq\ & K\gamma, \end{aligned}$$

whence

$$\begin{aligned} &\limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^{n} p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right] \\ \leq\ & \limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^{\lceil\gamma n\rceil-1} p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right] \\ &+\limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=\lceil\gamma n\rceil}^{n} p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right] \\ \leq\ & K\gamma+\limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=\lceil\gamma n\rceil}^{n} p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right]. \end{aligned}$$

Thus we have shown that

$$\begin{aligned} &\limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^{n} p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right] \\ =\ & \sup_{\gamma>0}\limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=\lceil\gamma n\rceil}^{n} p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right]. \end{aligned} \tag{30}$$

Now, arguing analogously as in the proof of Theorem 3.2 and using (30), we get

$$\begin{aligned} &\mathrm{Lin}\left(\left\{\frac{1}{s_n}X_k\right\}\right) \\ =\ & \sup_{\epsilon>0}\limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^{n}\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right] \end{aligned}$$

$$= \sup_{\epsilon>0} \sup_{\gamma>0} \limsup_{n\to\infty} \frac{1}{s_n^2} \sum_{k=\lceil \gamma n \rceil}^{n} p_k \sigma_k^2 \mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right]$$

$$= \sup_{\gamma>0} \sup_{\epsilon>0} \limsup_{n\to\infty} \frac{1}{s_n^2} \sum_{k=\lceil \gamma n \rceil}^{n} p_k \sigma_k^2 \mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right],$$

completing the proof. □

Theorem 3.8. *The inequality*

$$\operatorname{Lin}\left(\left\{\frac{1}{s_n}(X_k - \mu)\right\}\right) \leq \limsup_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} p_k \sigma_k^2 \tag{31}$$

always holds. If, in addition,

(1) $(\sigma_n^2)_n$ *is monotonically increasing,*

(2) $\liminf_{n\to\infty} \frac{1}{n}\sigma_n^2 > 0$,

(3) $\left(\frac{1}{n}\sum_{k=1}^{n} p_k \sigma_k^2\right)_n$ *is bounded,*

then the inequality in (31) becomes an equality.

Proof. Inequality (31) is easily established by the fact that $\mathbb{E}[X^2] = 1$. Now suppose that the three additional conditions in Theorem 3.8 are fulfilled. The fact that

$$\liminf_{n\to\infty} \frac{1}{n}\sigma_n^2 > 0$$

allows us to choose $\delta > 0$ and n_0 such that for all $n \geq n_0$

$$\sigma_n^2 \geq \delta n. \tag{32}$$

Furthermore, the boundedness of $\left(\frac{1}{n}\sum_{k=1}^{n} p_k \sigma_k^2\right)_n$ allows us to pick $K \in \mathbb{R}_0^+$ such that for all n

$$\frac{1}{n}\sum_{k=1}^{n} p_k \sigma_k^2 \leq K. \tag{33}$$

Now fix $\gamma > 0$ small. Then, for n so large that

$$\lceil \gamma n \rceil \geq n_0 \tag{34}$$

and for k such that

$$\lceil \gamma n \rceil \leq k \leq n, \tag{35}$$

we have, by (34), (35), (33), and (32),

$$\begin{aligned}
\left(\frac{s_n}{\sigma_k}\right)^2 &= \frac{\sum_{k=1}^{n}(1-p_k) + \sum_{k=1}^{n} p_k \sigma_k^2}{\sigma_k^2} \\
&\leq \frac{\sum_{k=1}^{n}(1-p_k) + \sum_{k=1}^{n} p_k \sigma_k^2}{\delta k} \\
&\leq \frac{\sum_{k=1}^{n}(1-p_k) + \sum_{k=1}^{n} p_k \sigma_k^2}{\delta \lceil \gamma n \rceil}
\end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{\delta\gamma}\left(\frac{1}{n}\sum_{k=1}^{n}(1-p_k)+\frac{1}{n}\sum_{k=1}^{n}p_k\sigma_k^2\right) \\ &\leq \frac{1+K}{\delta\gamma}, \end{aligned}$$

whence

$$\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right] \geq \mathbb{E}\left[X^2; |X| \geq \epsilon\sqrt{\frac{1+K}{\delta\gamma}}\right],$$

with X a random variable with cumulative distribution function F. In particular,

$$\sup_{\epsilon>0}\limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=\lceil\gamma n\rceil}^{n}p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right] \tag{36}$$

$$\begin{aligned} &\geq \sup_{\epsilon>0}\limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=\lceil\gamma n\rceil}^{n}p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \epsilon\sqrt{\frac{1+K}{\delta\gamma}}\right] \\ &= \sup_{\epsilon>0}\mathbb{E}\left[X^2; |X| \geq \epsilon\sqrt{\frac{1+K}{\delta\gamma}}\right]\limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=\lceil\gamma n\rceil}^{n}p_k\sigma_k^2 \\ &= \limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=\lceil\gamma n\rceil}^{n}p_k\sigma_k^2, \end{aligned}$$

where the last equality follows from the fact that $\mathbb{E}\left[X^2\right] = 1$. Combining Lemma 3.7 and the inequality shown by (36) gives

$$\begin{aligned} &\mathrm{Lin}\left(\left\{\frac{1}{s_n}(X_k-\mu)\right\}\right) \\ &= \sup_{\gamma>0}\sup_{\epsilon>0}\limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=\lceil\gamma n\rceil}^{n}p_k\sigma_k^2\mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right] \\ &\geq \sup_{\gamma>0}\limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=\lceil\gamma n\rceil}^{n}p_k\sigma_k^2 \\ &= \limsup_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^{n}p_k\sigma_k^2, \end{aligned}$$

the last equality following by mimicking the proof of Lemma 3.7. This finishes the proof. □

Theorem 3.9. *Suppose that*

(1) $\left(\frac{1}{n}\sum_{k=1}^{n}p_k\sigma_k^2\right)_n$ *is convergent to* $L \in \mathbb{R}^+$,

(2) $\left(\frac{1}{n}\sum_{k=1}^{n} p_k\right)_n$ *is convergent to 0.*

Then the inequality

$$\operatorname{Lin}\left(\left\{\frac{1}{s_n}(X_k-\mu)\right\}\right) \le \frac{L}{1+L} \tag{37}$$

holds. If, in addition,

(3) $(\sigma_n^2)_n$ *is monotonically increasing,*

(4) $\liminf_{n\to\infty} \frac{1}{n}\sigma_n^2 > 0,$

then the inequality in (37) becomes an equality.

Proof. Theorem 3.8 gives

$$\begin{aligned}
\operatorname{Lin}\left(\left\{\frac{1}{s_n}\left(X_k-\mu\right)\right\}\right) &\le \limsup_{n\to\infty} \frac{1}{s_n^2}\sum_{k=1}^{n} p_k\sigma_k^2 \\
&= \limsup_{n\to\infty} \frac{\sum_{k=1}^{n} p_k\sigma_k^2}{\sum_{k=1}^{n}(1-p_k)+\sum_{k=1}^{n} p_k\sigma_k^2} \\
&= \limsup_{n\to\infty} \frac{\frac{1}{n}\sum_{k=1}^{n} p_k\sigma_k^2}{1-\frac{1}{n}\sum_{k=1}^{n} p_k+\frac{1}{n}\sum_{k=1}^{n} p_k\sigma_k^2} \\
&= \frac{L}{1+L},
\end{aligned}$$

the last equality following from conditions (1) and (2) in Theorem 3.9. This establishes (37). If conditions (3) and (4) in Theorem 3.9 are also satisfied, then Theorem 3.8 shows that the first inequality in the above calculation becomes an equality and we are done. □

Now Theorem 2.10 gives the following result.

Theorem 3.10. *Let ξ be as in Section 1 and suppose that*

(1) $\left(\frac{1}{n}\sum_{k=1}^{n} p_k\sigma_k^2\right)_n$ *is convergent to $L \in \mathbb{R}^+$,*

(2) $\left(\frac{1}{n}\sum_{k=1}^{n} p_k\right)_n$ *is convergent to* 0*,*

(3) $\left(\max_{k=1}^{n} p_k\sigma_k^2\right)_n$ *is convergent to* 0*.*

Then, for each $\delta \in \{K, W, P\}$,

$$\lambda_\delta\left(\frac{n}{s_n}\left(\overline{X}_n-\mu\right)\to\xi\right) \le C_\delta \frac{L}{1+L}, \tag{38}$$

with $C_K = 1$, $C_W = 8$, and $C_P = 4$.

Proof. Theorem 3.5 is applicable to conclude that the STA $\left\{\frac{1}{s_n}(X_k-\mu)\right\}$ satisfies Feller's condition. Furthermore, Theorem 3.9 reveals that the Lindeberg index of this STA is bounded from above by $\frac{L}{1+L}$. Finally, the n-th row-wise sum of this STA coinciding with $\frac{n}{s_n}\left(\overline{X}_n-\mu\right)$, it suffices to apply Theorem 2.10. □

We wish to make the following final reflection. If $(\sigma_n)_n$ increases monotonically and $\liminf_{n\to\infty}\frac{1}{n}\sigma_n>0$, then classical central limit theory (Theorem 1.1) applied to the set of conditions imposed in Theorem 3.10 leads to the conclusion that the estimator $\overline{X}_n$ fails to be asymptotically normal in the sense that the sequence $\left(\frac{n}{s_n}\left(\overline{X}_n-\mu\right)\right)_n$ does not converge weakly to ξ. However, inequality (38), derived from the more general approximate central limit theory (Theorem 2.10), shows that $\overline{X}_n$ is still close to being asymptotically normal when L is small.

We empirically demonstrate these ideas in the next section through an example and a simulation study.

3.3. **Example and simulation study.** We keep the terminology and the notation from above.

In the following theorem we apply the results obtained in the previous section to a specific choice for p_k and σ_k^2. Recall that we say that $\overline{X}_n$ is weakly consistent (WC) for μ if (24) holds and asymptotically normal (AN) if (25) holds.

Theorem 3.11. *Let*

$$p_k = pk^{-a} \text{ with } p\in[0,1] \text{ and } a\in[0,\infty[$$

and

$$\sigma_k^2 = s^2k^b \text{ with } s\in[1,\infty[\text{ and } b\in[0,\infty[\,.$$

Then the following assertions are true.

(1) *If $b<1$, then $\overline{X}_n$ is WC for μ and AN.*
(2) *If $b\geq 1$ and $a>b$, then $\overline{X}_n$ is WC for μ and AN.*
(3) *If $b\geq 1$ and $a=b$, then $\overline{X}_n$ is WC for μ, but fails to be AN. However, for each $\delta\in\{K,W,P\}$,*

$$\lambda_\delta\left(\frac{n}{s_n}\left(\overline{X}_n-\mu\right)\to\xi\right)\leq C_\delta\frac{ps^2}{1+ps^2}, \tag{39}$$

with $C_K=1$, $C_W=8$, and $C_P=4$.

Proof. Firstly, suppose that $b<1$. Now, by (23),

$$\frac{1}{s_n^2}\max_{k=1}^{n}\sigma_k^2=\frac{n^b}{s_n^2}\leq n^{b-1}$$

which clearly converges to 0 as n tends to ∞. Thus condition (27) is satisfied, which allows us to conclude from Theorem 3.4 that $\overline{X}_n$ is AN. Also, Remark 3.3 shows that condition (26) holds, whence we infer from Theorem 3.1 that $\overline{X}_n$ is WC for μ. This establishes the first assertion.

Next, consider the case where $b \geq 1$ and $a > b$. Then the sequence

$$p_k \sigma_k^2 = ps^2 k^{b-a}$$

converges to 0 as k tends to ∞, whence

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_k \sigma_k^2 = 0.$$

Now it easily follows from Theorem 3.1 that $\overline{X}_n$ is WC for μ and from Theorem 3.10 that

$$\lambda_\delta \left(\frac{n}{s_n} \left(\overline{X}_n - \mu \right) \to \xi \right) = 0$$

for any $\delta \in \{K, W, P\}$. Put otherwise, $\overline{X}_n$ is AN and the second assertion is proved.

Finally, let $b \geq 1$ and $a = b$. Then

$$\frac{1}{n} \sum_{k=1}^{n} p_k \sigma_k^2 = ps^2.$$

Now the proof of the third assertion goes along the same lines as the proof of the second one. □

Theorem 3.11 shows for a specific example in which cases the sample mean is an accurate estimator with desirable asymptotic properties. Especially the third case is interesting, because although asymptotic normality is lacking in the classical sense, it gives a concrete numerical upper bound for how far the sample mean can maximally deviate from being asymptotically normal. This allows us to conclude that when this upper bound is small, it is still safe to assume asymptotic normality. This might be interesting from a practical point of view.

In order to illustrate Theorem 3.11, we have conducted a simulation study with the following setup. For specific instances of p, s, a, b we have created an empirical cdf $\mathcal{E}$ for $\frac{\overline{X}_n - \mathbb{E}[\overline{X}_n]}{\sqrt{\mathrm{Var}[\overline{X}_n]}}$ with sample size $n = 1000$, where we have assumed that $F = \Phi$, the cdf of a standard normal distribution. In each case the empirical cdf was based on 5000 simulations. We have tested for asymptotic normality by creating a QQ-plot the graph of which contains bullets with coordinates $(\Phi^{-1}(t), \mathcal{E}^{-1}(t))$, where t runs over a specific grid from 0 to 1. If a bullet $(\Phi^{-1}(t), \mathcal{E}^{-1}(t))$ is close to the line $y = x$, then $\Phi^{-1}(t) \approx \mathcal{E}^{-1}(t)$, whence $\mathcal{E}(\Phi^{-1}(t)) \approx t = \Phi(\Phi^{-1}(t))$. Thus on each QQ-plot we have also added the graph of the line $y = x$. To each figure we have added the value of the Lindeberg index governing the asymptotic normality of the sample mean. Recall that the Lindeberg index takes values between 0 and 1.

The following conclusions can be drawn from this study.

Figure 1 (p = 0.5, s = 3, a = 0.2, b = 0.5, Lin = 0)

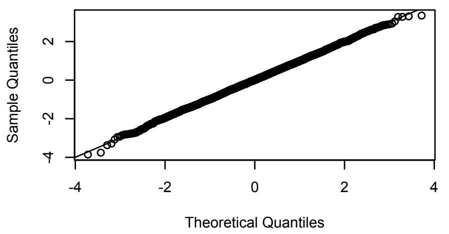


Figure 2 (p = 0.5, s = 3, a = 3, b = 1, Lin = 0)

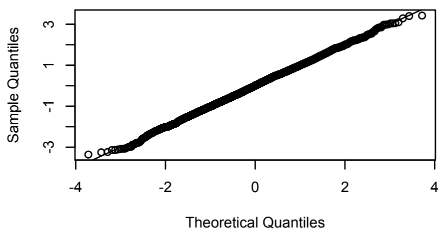


Figure 3 (p = 0.01, s = 1.5, a = 1, b = 1, Lin = 0.02)

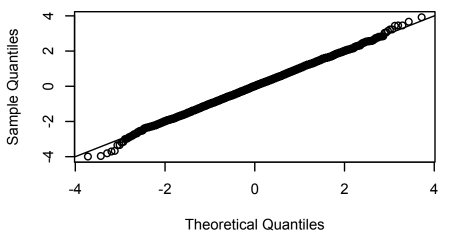


Figure 4 (p = 0.1, s = 1.5, a = 1, b = 1, Lin = 0.18)

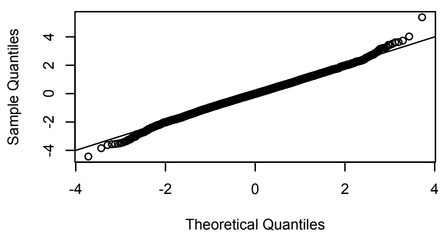


Figure 5 (p = 0.2, s = 2, a = 1, b = 1, Lin = 0.44)

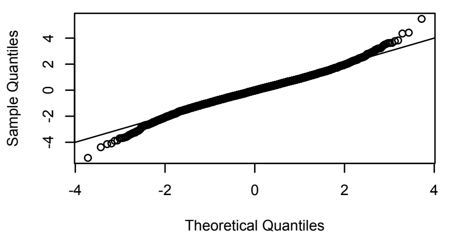
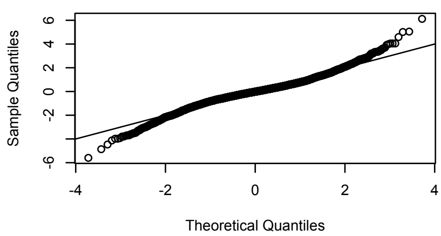


Figure 6 (p = 0.5, s = 4, a = 1, b = 1, Lin = 0.89)



If $b < 1$, then the first assertion in Theorem 3.11 states that - even if p and s are large and a is below b - the sample mean is asymptotically normal because the Lindeberg index is 0. This is confirmed by Figure 1.

If $b \geq 1$ and $a > b$, then the second assertion in Theorem 3.11 states that - even if p and s are large - the sample mean is asymptotically normal because the Lindeberg index is 0. This is confirmed by Figure 2.

If $b \geq 1$ and $a = b$, then the third assertion in Theorem 3.11 provides an upper bound for a canonical measure of the asymptotic normality of the sample mean because the Lindeberg index is $\frac{ps^2}{1+ps^2}$. The larger the Lindeberg index, the more deviation from asymptotic normality is expected. This is confirmed by Figures 3 – 6.

4. Open questions

We formulate some open questions which could be a source for future research.

Question 1. It would be nice to look for sharper constants in Theorem 2.10.

Question 2. The possibility of (partially) extending Theorem 2.10 to a multivariate setting, or even to a Hilbert- or Banach-valued context, is an interesting topic for further investigation.

Question 3. Theorem 3.11 does not handle the case where $b \geq 1$ and $a < b$. Assume without loss of generality that $\mu = 0$. Then, arguing analogously as in the proof of Theorem 3.2, we easily see that

$$\operatorname{Lin}\left(\left\{\frac{1}{s_n}X_k\right\}\right) = \sup_{\epsilon>0} \limsup_{n\to\infty} \frac{1}{s_n^2}\sum_{k=1}^{n} \mathbb{E}\left[X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k}\right],$$

X being a random variable with cumulative distribution function F and

$$\sigma_k^2 = s^2 k^b$$

and

$$s_n^2 = n - p\sum_{k=1}^{n} k^{-a} + ps^2 \sum_{k=1}^{n} k^{b-a}.$$

It would be of interest to examine the existence of a more explicit formula for the Lindeberg index in this case. Also, the weak consistency should be investigated.

Question 4. Strictly speaking, inequality (38) only shows that the Lindeberg index is an upper bound for a natural index measuring the asymptotic normality of the sample mean. This allows us to draw the conclusion that the sample mean is close to being asymptotically normal when the Lindeberg index is small, but we cannot say anything about what happens when the Lindeberg index is large. However, our simulation study empirically reveals that when the Lindeberg index gets larger, the sample mean tends to deviate more from asymptotic normality. It would be of interest to establish a useful lower bound for $\lambda_\delta\left(\frac{n}{s_n}\left(\overline{X}_n - \mu\right) \to \xi\right)$ in terms of the Lindeberg index, which serves as a theoretical underpinning of this observation. General lower bounds of this type have been obtained in [BLV13] for the Kolmogorov distance, but they are so unsharp that they do not have the power to predict what we have seen in our simulation study.

Appendix A: Proof of Theorem 2.1

We follow [BLV13], Section 2. We keep a continuously differentiable $h : \mathbb{R} \to [0, 1]$, with bounded derivative, fixed, and let f_h be its Stein transform defined by (6). Also, we put

$$\sigma_{n,k}^2 = \mathbb{E}[\xi_{n,k}^2].$$

The following lemma is easily verified. It can be found in e.g. [BC05] (p.10-11).

Lemma 1. *f_h is twice continuously differentiable, has bounded first and second derivatives, and*

$$\mathbb{E}\left[h(\xi)\right] - h(x) = x f_h(x) - f_h'(x). \tag{40}$$

The following lemma can be found in [BLV13] (Lemma 2.4). We give the proof for completeness.

Lemma 2. *Put*

$$\delta_{n,k} = f_h\left(\sum_{i\neq k}\xi_{n,i} + \xi_{n,k}\right) - f_h\left(\sum_{i\neq k}\xi_{n,i}\right) - \xi_{n,k} f_h'\left(\sum_{i\neq k}\xi_{n,i}\right)$$

and

$$\epsilon_{n,k} = f_h'\left(\sum_{i\neq k}\xi_{n,i} + \xi_{n,k}\right) - f_h'\left(\sum_{i\neq k}\xi_{n,i}\right) - \xi_{n,k} f_h''\left(\sum_{i\neq k}\xi_{n,i}\right).$$

Then

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{k=1}^{n}\xi_{n,k}\right) f_h\left(\sum_{k=1}^{n}\xi_{n,k}\right) - f_h'\left(\sum_{k=1}^{n}\xi_{n,k}\right)\right] \\
&= \sum_{k=1}^{n}\mathbb{E}\left[\xi_{n,k}\delta_{n,k}\right] - \sum_{k=1}^{n}\sigma_{n,k}^2\mathbb{E}\left[\epsilon_{n,k}\right].
\end{aligned} \tag{41}$$

Proof. Recalling that $\xi_{n,k}$ and $\sum_{i\neq k}\xi_{n,i}$ are independent, $\mathbb{E}\left[\xi_{n,k}\right] = 0$, and $\sum_{k=1}^{n}\sigma_{n,k}^2 = 1$, we get

$$\begin{aligned}
&\sum_{k=1}^{n}\mathbb{E}\left[\xi_{n,k}\delta_{n,k}\right] - \sum_{k=1}^{n}\sigma_{n,k}^2\mathbb{E}\left[\epsilon_{n,k}\right] \\
&= \sum_{k=1}^{n}\mathbb{E}\left[\xi_{n,k} f_h\left(\sum_{k=1}^{n}\xi_{n,k}\right)\right] - \mathbb{E}\left[\xi_{n,k} f_h\left(\sum_{i\neq k}\xi_{n,i}\right)\right] \\
&\quad -\sum_{k=1}^{n}\mathbb{E}\left[\xi_{n,k}^2 f_h'\left(\sum_{i\neq k}\xi_{n,i}\right)\right] - \sum_{k=1}^{n}\sigma_{n,k}^2\mathbb{E}\left[f_h'\left(\sum_{k=1}^{n}\xi_{n,k}\right)\right] \\
&\quad +\sum_{k=1}^{n}\mathbb{E}\left[\xi_{n,k}^2\right]\mathbb{E}\left[f_h'\left(\sum_{i\neq k}\xi_{n,i}\right)\right] + \sum_{k=1}^{n}\sigma_{n,k}^2\mathbb{E}\left[\xi_{n,k} f_h''\left(\sum_{i\neq k}\xi_{n,i}\right)\right].
\end{aligned}$$

The last expression further reduces to

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{k=1}^{n}\xi_{n,k}\right) f_h\left(\sum_{k=1}^{n}\xi_{n,k}\right)\right] - \mathbb{E}\left[\xi_{n,k}\right]\mathbb{E}\left[f_h\left(\sum_{i\neq k}\xi_{n,i}\right)\right] \\
&\quad -\sum_{k=1}^{n}\mathbb{E}\left[\xi_{n,k}^2 f_h'\left(\sum_{i\neq k}\xi_{n,i}\right)\right] - \mathbb{E}\left[f_h'\left(\sum_{k=1}^{n}\xi_{n,k}\right)\right]
\end{aligned}$$

$$+\sum_{k=1}^{n}\mathbb{E}\left[\xi_{n,k}^{2}f_h'\left(\sum_{i\neq k}\xi_{n,i}\right)\right]+\sum_{k=1}^{n}\sigma_{n,k}^{2}\mathbb{E}\left[\xi_{n,k}\right]\mathbb{E}\left[f_h''\left(\sum_{i\neq k}\xi_{n,i}\right)\right],$$

which is easily seen to equal

$$\mathbb{E}\left[\left(\sum_{k=1}^{n}\xi_{n,k}\right)f_h\left(\sum_{k=1}^{n}\xi_{n,k}\right)-f_h'\left(\sum_{k=1}^{n}\xi_{n,k}\right)\right].$$

This finishes the proof. □

The following lemma is an application of Taylor's theorem.

Lemma 3. *For any $a, x \in \mathbb{R}$,*

$$\begin{aligned}&|f_h(a+x)-f_h(a)-f_h'(a)x|\\&\quad\leq\min\left\{\left(\sup_{x_1,x_2\in\mathbb{R}}|f_h'(x_1)-f'(x_2)|\right)|x|\,,\frac{1}{2}\|f_h''\|_\infty x^2\right\}.\end{aligned}\qquad(42)$$

We are now in a position to present a proof of Theorem 2.1.

Proof of Theorem 2.1. For n and $\epsilon>0$, we have, by (40), (41), and (42),

$$\begin{aligned}&\left|\mathbb{E}\left[h\left(\xi\right)-h\left(\sum_{k=1}^{n}\xi_{n,k}\right)\right]\right|\\
&=\left|\mathbb{E}\left[\left(\sum_{k=1}^{n}\xi_{n,k}\right)f_h\left(\sum_{k=1}^{n}\xi_{n,k}\right)-f_h'\left(\sum_{k=1}^{n}\xi_{n,k}\right)\right]\right|\\
&\leq\sum_{k=1}^{n}\mathbb{E}\left[|\xi_{n,k}\delta_{n,k}|\right]+\sum_{k=1}^{n}\sigma_{n,k}^{2}\mathbb{E}\left[|\epsilon_{n,k}|\right]\\
&\leq\frac{1}{2}\|f_h''\|_\infty\sum_{k=1}^{n}\mathbb{E}\left[|\xi_{n,k}|^3;|\xi_{n,k}|<\epsilon\right]\\
&\quad+\left(\sup_{x_1,x_2\in\mathbb{R}}|f_h'(x_1)-f_h'(x_2)|\right)\sum_{k=1}^{n}\mathbb{E}\left[|\xi_{n,k}|^2;|\xi_{n,k}|\geq\epsilon\right]\\
&\quad+\left(\sup_{x_1,x_2\in\mathbb{R}}|f_h''(x_1)-f_h''(x_2)|\right)\sum_{k=1}^{n}\sigma_{n,k}^{2}\mathbb{E}\left[|\xi_{n,k}|\right],\end{aligned}$$

which proves the desired result since $\sum_{k=1}^{n}\sigma_{n,k}^{2}=1$.

References

[BC05] Barbour, A.D.; Chen, L.H.Y. *An introduction to Stein's method* Singapore University Press, Singapore; World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.

[BH84] Barbour, A. D.; Hall, P. *Stein's method and the Berry-Esseen theorem.* Austral. J. Statist. 26 (1984), no. 1, 8–15.

[BLV11] Berckmoes, B.; Lowen, R.; van Casteren, J. Distances on probability measures and random variables. J. Math. Anal. Appl. 374 (2011), no. 2, 412–428.

[BLV13] Berckmoes, B.; Lowen, R.; Van Casteren, J. *An isometric study of the Lindeberg-Feller central limit theorem via Stein's method.* J. Math. Anal. Appl. 405 (2013), no. 2, 484–498.

[BLV16] Berckmoes B.; Lowen R.; Van Casteren J. *Stein's Method and a Quantitative Lindeberg CLT for the Fourier Transforms of Random Vectors* J. Math. Anal. Appl. 433 (2016), no. 2, 1441–1458.

[B99] Billingsley, P. *Convergence of probability measures* Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. John Wiley and Sons, Inc., New York, 1999.

[B08] Bolley, F. *Separability and completeness for the Wasserstein distance.* Sminaire de probabilits XLI, 371–377, Lecture Notes in Math., 1934, Springer, Berlin, 2008.

[CGS11] Chen, L. H. Y.; Goldstein, L.; Shao, Q-M. *Normal approximation by Stein's method.* Probability and its Applications (New York). Springer, Heidelberg, 2011.

[CS01] Chen, L. H. Y.; Shao, Q-M. *A non-uniform Berry-Esseen bound via Stein's method.* Probab. Theory Related Fields 120 (2001), no. 2, 236–254.

[O66] Osipov, L. V. *A refinement of Lindberg's theorem.* (Russian) Teor. Verojatnost. i Primenen. 11 1966 339–342.

[F68] Feller, W. *On the Berry-Esseen theorem.* Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 10 1968 261–268.

[L75] Loh, W.Y. *On the normal approximation for sums of mixing random variables.* Master Thesis, Department of Mathematics, University of Singapore (1975)

[L15] Lowen, R. *Index analysis. Approach theory at work.* Springer Monographs in Mathematics. Springer, London, 2015.

[M09] Meckes, E. *On Stein's method for multivariate normal approximation.* High dimensional probability V: the Luminy volume, 153–178, Inst. Math. Stat. Collect., 5, Inst. Math. Statist., Beachwood, OH, 2009.

[NPR10] Nourdin, I.; Peccati, G.; Réveillac, A. *Multivariate normal approximation using Stein's method and Malliavin calculus.* Ann. Inst. Henri Poincaré Probab. Stat. 46 (2010), no. 1, 45–58.

[R91] Rachev, S. T. *Probability metrics and the stability of stochastic models.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1991.

[TSM85] Titterington, D. M.; Smith, A. F. M.; Makov, U. E. *Statistical analysis of finite mixture distributions. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics.* John Wiley & Sons, Ltd., Chichester, 1985.

[V03] Villani, C. *Topics in optimal transportation.* Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003.

[Z83] Zolotarev, V. M. *Probability metrics.* (Russian) Teor. Veroyatnost. i Primenen. 28 (1983), no. 2, 264–287.