ON THE ASYMPTOTIC BEHAVIOR OF THE CONTAMINATED SAMPLE MEAN

BEN BERCKMOES AND GEERT MOLENBERGHS

Abstract. An observation of a cumulative distribution function $F$ with finite variance is said to be contaminated according to the inflated variance model if it has a large probability of coming from the original target distribution $F$, but a small probability of coming from a contaminating distribution that has the same mean and shape as $F$, though a larger variance. In this paper we investigate under which conditions the sample mean, based on a finite number of independent observations of $F$ which are contaminated according to the inflated variance model, is a valid estimator for the mean of $F$. In particular, we examine up to what extent this estimator is weakly consistent for the mean of $F$ and asymptotically normal. As the classical central limit theory will in many situations turn out to be inaccurate to cope with the asymptotic normality in this setting, we will fall back on the more general quantitative central limit theory as developed by Berckmoes, Lowen and Van Casteren. Our theoretical results are illustrated by a specific example and a simulation study.

1. Introduction

Suppose that we are given a finite number of independent observations $X_1, \ldots, X_n$ of a cumulative distribution function $F$ on the real line with mean $\mu$ and finite variance. It belongs to the folklore of classical probability theory and mathematical statistics that the sample mean $\overline{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ is an accurate estimator for $\mu$ in the sense that it is consistent for $\mu$, i.e. $\overline{X}_n \overset{a.s.}{\to} \mu$, and asymptotically normal, i.e. $\frac{1}{\sqrt{n}} (\overline{X}_n - \mu) \overset{d}{\to} N(0, 1)$.

Now assume that there is an underlying mechanism causing each observation to be contaminated according to the inflated variance model (see [8], p.108). That is, instead of each $X_k$ having distribution $F$, there is a large probability that $X_k$ comes from the distribution $F$, but a small probability that it comes from a contaminating distribution $\tilde{F}_k = F\left(\cdot - \frac{\mu}{\sigma_k}\right)$ which has the same shape and mean as $F$, though a larger variance $\sigma_k^2 \geq \text{Var}[F]$.

In this paper we investigate under what conditions the sample mean in this contaminated setting remains (weakly) consistent for $\mu$ and asymptotically normal.

It will turn out that the weak consistency can be easily established under a fairly weak condition using Chebyshev’s inequality (Theorem 4.1) and that the asymptotic normality can be established using the classical Lindeberg-Feller central limit theory if the sequence of contaminating variances $(\sigma_k)_k$ can be controlled sufficiently (Theorem 4.4).

Ben Berckmoes is post doctoral fellow at the Fund for Scientific Research of Flanders (FWO).
However, if we lack sufficient control over the sequence \((\sigma_k)_k\), we end up with an interesting class of settings in which the classical central limit theory is inaccurate to describe the asymptotic behavior of the sample mean. This is due to the fact that in this case the question of whether the sample mean is asymptotically normal should be answered more subtly than with a simple yes or no. Instead we will use the quantitative generalization of the classical Lindeberg-Feller central limit theory as developed by Berckmoes, Lowen and Van Casteren in [4] to produce a number between 0 and 1 which serves as an upper bound for a canonical index measuring how far the sample mean deviates from being asymptotically normal (Theorem 4.9). We thus get a gradation in which we are closer to asymptotic normality in some cases, and further away from it in others. This will be made visible by QQ-plots coming from a simulation study.

The paper is structured as follows.

In section 2 the formal framework in which we will work and the notation we will use is sketched.

The key results of the quantitative central limit theory developed in [4] are explained in section 3.

Section 4 is the core of this paper. Here we give the theoretical results concerning the asymptotic properties of the sample mean in the contaminated setting.

Finally, in section 5 we give a specific example and an illustration via a simulation study.

2. Formal framework

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

\[
\int_{-\infty}^{\infty} x dF(x) = 0
\]

and

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

Fix \( \mu \in \mathbb{R} \) and let \( X_1, X_2, \ldots, X_k, \ldots \) be independent observations of \( F(\cdot - \mu) \) which are contaminated according to the inflated variance model (see [3], p.108), that is

\[
X_k \sim (1 - p_k)F(\cdot - \mu) + p_k F\left( \frac{\cdot - \mu}{\sigma_k} \right)
\]

where \( p_k \in [0, 1] \) and \( \sigma_k \in [1, \infty] \). Observe that

\[ E[X_k] = \mu \]

and

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

Now define the sample mean in the usual way as

\[ \bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k. \]

Notice that

\[ E[\bar{X}_n] = \mu \]
and
\[ \text{Var}[X_n] = \left( \frac{s_n}{n} \right)^2 \]
where
\[ s_n^2 = \sum_{k=1}^{n} [(1 - p_k) + p_k \sigma_k^2]. \]
Also,
\[ s_n^2 \geq n \]
(1)
because \( \sigma_k^2 \geq 1 \) for all \( k \). In this paper we investigate up to what extent the estimator \( \overline{X}_n \) is weakly consistent for \( \mu \) in the sense that
\[ \overline{X}_n \xrightarrow{p} \mu \]
(2)
and asymptotically normal in the sense that
\[ \frac{n}{s_n} (\overline{X}_n - \mu) \xrightarrow{d} N(0, 1). \]
(3)
Notice that in the uncontaminated case where \( \sigma_k = 1 \) for all \( k \), the Weak Law of Large Numbers implies the truth of (2) and the Central Limit Theorem entails the validity of (3). For our study of the asymptotic normality we fall back on the quantitative central limit theory as developed by Berckmoes, Lowen and Van Casteren in [4]. We briefly recall the basics of this theory in the next section.

3. QUANTITATIVE CENTRAL LIMIT THEORY

By a standard triangular array (STA) we mean a triangular array of real square integrable random variables
\[ \xi_{1,1}, \xi_{2,1}, \Xi_2, \xi_{3,1}, \xi_{3,2}, \xi_{3,3}, \ldots \]
with the following properties:
1. \( \forall n : \xi_{n,1}, \ldots, \xi_{n,n} \) are independent,
2. \( \forall n, k : E [\xi_{n,k}] = 0 \),
3. \( \forall n : \sum_{k=1}^{n} E [\xi_{n,k}^2] = 1. \)

We say that an STA \( \{\xi_{n,k}\} \) satisfies Lindeberg’s condition iff
\[ \forall \epsilon > 0 : \lim_{n \to \infty} \sum_{k=1}^{n} E [\xi_{n,k}^2; |\xi_{n,k}| \geq \epsilon] = 0 \]
and that it satisfies Feller’s condition iff
\[ \lim_{n \to \infty} \max_{k=1}^{n} E [\xi_{n,k}^2] = 0. \]
It is well known (and readily verified) that Lindeberg’s condition is strictly stronger than Feller’s condition.

The importance of the notions explained above is reflected by the following key result in the classical central limit theory ([5]).
Theorem 3.1 (Central Limit Theorem (CLT)). Consider, for $\xi \sim N(0,1)$ and $\{\xi_{n,k}\}$ an STA, the following assertions:

1. The weak convergence relation $\sum_{k=1}^{n} \xi_{n,k} \xrightarrow{w} \xi$ holds.
2. The STA $\{\xi_{n,k}\}$ satisfies Lindeberg’s condition.

Then assertion (2) implies assertion (1) and both assertions are equivalent if $\{\xi_{n,k}\}$ satisfies Feller’s condition.

Now suppose that we are given an STA $\{\xi_{n,k}\}$ which satisfies Feller’s condition, but fails to satisfy Lindeberg’s condition. Then we infer from the classical central limit theory (Theorem 3.1) that the row-wise sums of $\{\xi_{n,k}\}$ fail to be asymptotically normal. However, inspired by approach theory, a topological theory pioneered by Lowen [6],[7] (the details of which are not needed for a proper understanding of this paper), we could ask the following question: How far does $\sum_{k=1}^{n} \xi_{n,k}$ deviate from $\xi$ if $n$ gets large?

In order to formalize this question, recall that the Kolmogorov distance between random variables $\eta$ and $\eta'$ is given by

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

It is well known that for a continuously distributed random variable $\eta$ and an arbitrary sequence of random variables $(\eta_n)_n$ the following are equivalent:

1. $\eta_n \xrightarrow{w} \eta$,
2. $K(\eta, \eta_n) \to 0$.

Thus, even for an STA $\{\xi_{n,k}\}$ for which the sequence $(\sum_{k=1}^{n} \xi_{n,k})_n$ fails to converge weakly to $\xi$, it still makes sense to consider the number

$$\limsup_{n \to \infty} K\left(\xi, \sum_{k=1}^{n} \xi_{n,k}\right)$$

which takes values between 0 and 1 and measures in a precise sense how far the sequence of row-wise sums deviates from being asymptotically normal. In the language of approach theory, expression (4) is referred to as a limit operator or an index of convergence ([6],[2],[3]). Notice that $\limsup_{n \to \infty} K\left(\xi, \sum_{k=1}^{n} \xi_{n,k}\right) = 0$ if and only if $\sum_{k=1}^{n} \xi_{n,k} \xrightarrow{w} \xi$.

For an arbitrary STA $\{\xi_{n,k}\}$ it also makes sense to introduce the number

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

which lies between 0 and 1 and is a canonical index which measures how far $\{\xi_{n,k}\}$ deviates from satisfying Lindeberg’s condition. Expression (5) is called the Lindeberg index. Observe that $\text{Lin}(\{\xi_{n,k}\}) = 0$ if and only if $\{\xi_{n,k}\}$ satisfies Lindeberg’s condition.

The following result, which links the expressions (4) and (5) for an arbitrary STA satisfying Feller’s condition, lies at the heart of the quantitative central limit theory developed in [4]. The proof relies on Stein’s method ([1]).
Theorem 3.2 (Quantitative Central Limit Theorem (QCLT)). Consider \( \xi \sim N(0, 1) \) and \( \{\xi_{n,k}\} \) an STA which satisfies Feller’s condition. Then

\[
\limsup_{n \to \infty} K \left( \xi, \sum_{k=1}^{n} \xi_{n,k} \right) \leq \text{Lin} \left( \{\xi_{n,k}\} \right).
\]

Notice that the QCLT is a generalization of the CLT, which has the advantage that it can cope with STA’s that fail to satisfy Lindeberg’s condition. The QCLT heuristically states that if an STA is close to satisfying Lindeberg’s condition, then its row-wise sums are close to being asymptotically normally distributed.

We make use of the quantitative central limit theory in the next section, where we study the asymptotic behavior of the contaminated sample mean as introduced in the previous section.

4. Consistency and asymptotic normality

We keep the terminology and the notation of the previous sections. The proofs of the results obtained here are deferred to Appendix A.

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

**Theorem 4.1.** Suppose that

\[
\lim_{n \to \infty} \frac{1}{n^2} \sum_{k=1}^{n} p_k \sigma_k^2 = 0. \tag{6}
\]

Then

\( \overline{X}_n \xrightarrow{p} \mu. \)

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 4.2.** Suppose that

\[
\lim_{n \to \infty} \frac{1}{s_n^2} \max_{k=1}^{n} \sigma_k^2 \to 0. \tag{7}
\]

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

\[
\text{Lin} \left( \left\{ \frac{1}{s_n} (X_k - \mu) \right\} \right) = 0.
\]

**Remark 4.3.** Observe that (7) implies (6).

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

**Theorem 4.4.** Let \( \xi \sim N(0, 1) \) and suppose that

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

Then

\[
\frac{n}{s_n} (\overline{X}_n - \mu) \xrightarrow{w} \xi.
\]
If the sequence \( (\sigma_k)_k \) cannot be controlled by condition (7), then it turns out to be more appropriate to make use of the quantitative central limit theory. As Feller’s condition plays an important role in this theory, we start with the following characterization.

**Theorem 4.5.** The STA \( \left\{ \frac{1}{s_n} (X_k - \mu) \right\} \) satisfies Feller’s condition if and only if
\[
\lim_{n \to \infty} \frac{1}{s_n^2} \max_{1 \leq k \leq n} p_k \sigma_k^2 = 0. \tag{8}
\]

**Remark 4.6.** Observe that (7) implies (8) and that (8) implies (6).

Theorems 4.7 and 4.8 reveal that even in the absence of condition (7), the Lindeberg index of the STA \( \left\{ \frac{1}{s_n} (X_k - \mu) \right\} \) can still be bounded from above. Moreover, it can be explicitly computed under a fairly easy set of conditions.

**Theorem 4.7.** The inequality
\[
\text{Lin} \left( \left\{ \frac{1}{s_n} (X_k - \mu) \right\} \right) \leq \limsup_{n \to \infty} \frac{1}{s_n^2} \sum_{k=1}^{n} p_k \sigma_k^2 \tag{9}
\]
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) \(\frac{1}{n} \sum_{k=1}^{n} p_k \sigma_k^2\) is bounded,

then the inequality in (9) becomes an equality.

**Theorem 4.8.** Suppose that

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

(2) \(\frac{1}{n} \sum_{k=1}^{n} p_k\) is convergent to 0.

Then the inequality
\[
\text{Lin} \left( \left\{ \frac{1}{s_n} (X_k - \mu) \right\} \right) \leq \frac{L}{1 + L} \tag{10}
\]
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 (10) becomes an equality.

Now the quantitative central limit theory gives the following result. As in the previous section, \(K\) stands for the Kolmogorov distance.

**Theorem 4.9.** Let \(\xi \sim \mathcal{N}(0,1)\) and suppose that
\[
\left(1 \frac{1}{n} \sum_{k=1}^{n} p_k \sigma_k^2\right) \text{ is convergent to } L \in \mathbb{R}^+,
\]

\[
\left(2 \frac{1}{n} \sum_{k=1}^{n} p_k\right) \text{ is convergent to } 0,
\]

\[
\left(3 \max_{k=1}^{n} p_k \sigma_k^2\right) \text{ is convergent to } 0.
\]

Then
\[
\limsup_{n \to \infty} K\left(\xi, \frac{n}{s_n} (X_n - \mu)\right) \leq \frac{L}{1 + L}.
\]

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 3.1) applied to the set of conditions imposed in Theorem 4.9 leads to the conclusion that the estimator \(X_n\) fails to be asymptotically normal in the sense that the sequence \(\left(\frac{\mu}{s_n} (X_n - \mu)\right)\) does not converge weakly to \(\xi\). However, inequality (11), derived from the more general quantitative central limit theory, shows that \(X_n\) 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.

5. Example and simulation study

We keep the terminology and the notation of the previous sections.

In the following theorem we apply the results obtained in the previous section to a specific choice for \(p_k\) and \(\sigma_k^2\). Recall that we say that \(X_n\) is weakly consistent (WC) for \(\mu\) if (2) holds and asymptotically normal (AN) if (3) holds.

\textbf{Theorem 5.1.} Let
\[
p_k = pk^{-a} \text{ with } p \in [0, 1] \text{ and } a \in [0, \infty[.
\]
and
\[
\sigma_k^2 = s^2 k^b \text{ with } s \in [1, \infty[ \text{ and } b \in [0, \infty[.
\]
Then the following assertions are true.

(1) If \(b < 1\), then \(X_n\) is WC for \(\mu\) and AN.

(2) If \(b \geq 1\) and \(a > b\), then \(X_n\) is WC for \(\mu\) and AN.

(3) If \(b \geq 1\) and \(a = b\), then \(X_n\) is WC for \(\mu\), but fails to be AN.

However,
\[
\limsup_{n \to \infty} K\left(\xi, \frac{n}{s_n} (X_n - \mu)\right) \leq \frac{ps^2}{1 + ps^2}.
\]

\textbf{Theorem 5.1} 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 5.1 we have conducted a simulation study with the following setup. For specific instances of \(p, s, a, b\) we have created an empirical cdf \(E\) for \(\frac{X_n - \mathbb{E}[X_n]}{\sqrt{\text{Var}[X_n]}}\) 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 \((\Phi^{-1}(t), E^{-1}(t))\), where \(t\) runs over a specific grid from 0 to 1. If a bullet \((\Phi^{-1}(t), E^{-1}(t))\) is close to the line \(y = x\), then \(\Phi^{-1}(t) \approx E^{-1}(t)\), whence \(E(\Phi^{-1}(t)) \approx t = \Phi(\Phi^{-1}(t))\). 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 QQ-plots can be found in Appendix B.

The following conclusions can be drawn from this study.

If \(b < 1\), then the first assertion in Theorem 5.1 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 5.1 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 5.1 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, 4, 5 and 6.

6. Concluding remarks and further questions

In this paper we have investigated the weak consistency and the asymptotic normality of the sample mean based on a sample of independent observations which are contaminated according to the inflated variance model. We have shown that weak consistency is easily established under a fairly weak condition and that asymptotic normality can be established if we have sufficient control over the sequence of contaminating variances.

When the contaminating variances increase to infinity very rapidly, the situation concerning asymptotic normality completely changes. Instead of answering the question of whether the sample mean is asymptotically normal in this case with a simple yes or no, we have shown that, depending on the specific situation, there is a concrete upper bound for a canonical number measuring how much the estimator deviates from asymptotic normality. From a practical point of view, this means that in the case where this upper bound is small, we might safely assume asymptotic normality, even when it is lacking in the classical sense.

A simulation study has confirmed our theoretical conclusions.

Finally, we wish to mention that the following interesting related questions remain unsolved.
Question 1. Theorem 5.1 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 4.2, we easily see that

$$\text{Lin}\left(\left\{ \frac{1}{s_n}X_k \right\} \right) = \sup_{\epsilon > 0} \lim_{n \to \infty} \sup \frac{1}{s_n^a} \sum_{k=1}^{n} E \left[ X^2; |X| \geq \epsilon s_n \right]$$

$X$ being a random variable with $F$ as cumulative distribution function and

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

and

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

It would be of interest to examine the existence of a more explicit formula for the Lindeberg index in this case. Simulation points out that the Lindeberg index is always 1 in this case, but we have not been able to prove this formally. Also, the weak consistency should be investigated.

Question 2. Theoretically, inequality (11) 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 $\limsup_{n \to \infty} K \left( \frac{1}{s_n} \left( \bar{X}_n - \mu \right) \right)$ 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 [4], but they are so unsharp that they do not have the power to predict what we have seen in our simulation study.
Appendix A: Proofs

Proof of Theorem 4.1. Assume without loss of generality that $\mu = 0$. For $\epsilon > 0$, Chebyshev’s inequality gives
\[
P \left[ |X_n| \geq \epsilon \right] \leq \frac{1}{\epsilon^2} \text{Var}[X_n]
\]
which easily implies that
\[
\limsup_{n \to \infty} P \left[ |X_n| \geq \epsilon \right] \leq \frac{1}{\epsilon^2} \limsup_{n \to \infty} \frac{1}{n^2} \sum_{k=1}^{n} (1 - p_k) + \frac{1}{\sqrt{n}} \sum_{k=1}^{n} p_k \sigma_k^2.
\]
This finishes the proof.

Proof of Theorem 4.2. Assume without loss of generality that $\mu = 0$ and let $X$ be a random variable with $F$ as cumulative distribution function. Then, for $\epsilon > 0$,
\[
\frac{1}{s_n^2} \sum_{k=1}^{n} (1 - p_k) E \left[ X_k^2 ; |X| \geq \epsilon s_n \right] + \frac{1}{s_n^2} \sum_{k=1}^{n} p_k \sigma_k^2 E \left[ X^2 ; |X| \geq \epsilon s_n \right]
\]
which is
\[
\leq \frac{1}{s_n^2} \sum_{k=1}^{n} (1 - p_k) E \left[ X_k^2 ; |X| \geq \epsilon s_n \right] + \frac{1}{s_n^2} \sum_{k=1}^{n} p_k \sigma_k^2 E \left[ X^2 ; |X| \geq \epsilon \sqrt{\frac{s_n^2}{\max_k \sigma_k^2}} \right]
\]
\[
\leq E[X^2 ; |X| \geq \epsilon s_n] + E \left[ X^2 ; |X| \geq \epsilon \sqrt{\frac{s_n^2}{\max_k \sigma_k^2}} \right].
\]
The latter quantity converges to 0 as $n$ tends to $\infty$ by (1) and (7). This finishes the proof.

Proof of Theorem 4.4. Notice that the $n$-th rowwise sum of the STA $\left\{ \frac{1}{s_n} (X_k - \mu) \right\}$ coincides with $\frac{n}{s_n} (\overline{X}_n - \mu)$. Now apply Theorem 4.2 and Theorem 3.1.

Proof of Theorem 4.5. Assume without loss of generality that $\mu = 0$. Now
\[
\max_k E \left[ \frac{1}{s_n^2} X_k^2 \right] = \frac{1}{s_n^2} \max_k (1 - p_k) + \frac{1}{s_n^2} \max_k p_k \sigma_k^2
\]
whence, by (1),
\[
\limsup_{n \to \infty} \max_k E \left[ \frac{1}{s_n^2} X_k^2 \right] = \limsup_{n \to \infty} \frac{1}{s_n^2} \max_k p_k \sigma_k^2.
\]
This finishes the proof.
The proof of Theorem 4.7 is based on the following lemma.

**Lemma A.1.** 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 \( F \) as cumulated distribution function. Then

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

where \( \lceil \cdot \rceil \) is the ceiling function.

**Proof of Lemma A.1.** Assume w.l.o.g. that \( \mu = 0 \) and choose \( K \in \mathbb{R}_{+} \) such that for all \( n \)

\[
\frac{1}{n} \sum_{k=1}^{n} p_k \sigma_k^2 \leq K. \tag{12}
\]

Next, fix \( \gamma > 0 \) small. Then, for \( n \) large, by (1) and (12),

\[
\frac{1}{s_n^2} \sum_{k=1}^{\lceil \gamma n \rceil - 1} p_k \sigma_k^2 \mathbb{E} \left[ X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k} \right]
\]

\[
\leq \gamma \frac{1}{\gamma n} \sum_{k=1}^{\lceil \gamma n \rceil - 1} p_k \sigma_k^2
\]

\[
\leq \gamma \frac{1}{\gamma n} \sum_{k=1}^{\lceil \gamma n \rceil - 1} p_k \sigma_k^2
\]

\[
\leq K \gamma
\]

whence

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

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

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

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

Thus we have shown that

\[
\limsup_{n \to \infty} \frac{1}{s_n^2} \sum_{k=1}^{n} p_k \sigma_k^2 \mathbb{E} \left[ X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k} \right] = \sup_{\gamma > 0} \limsup_{n \to \infty} \frac{1}{s_n^2} \sum_{k=\lceil \gamma n \rceil}^{n} p_k \sigma_k^2 \mathbb{E} \left[ X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k} \right]. \tag{13}
\]
Now, arguing analogously as in the proof of Theorem 4.2 and using (13), we get
\[
\text{Lin}\left(\left\{ \frac{1}{s_n}X_k \right\} \right) = \sup_{\epsilon > 0} \limsup_{n \to \infty} \frac{1}{s_n^2} \sum_{k=1}^{n} E \left[ X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k} \right]
\]
\[
= \sup_{\epsilon > 0} \limsup_{n \to \infty} \frac{1}{s_n^2} \sum_{k=\lceil \gamma n \rceil}^{n} p_k \sigma_k^2 E \left[ X^2; |X| \geq \frac{\epsilon s_n}{\sigma_k} \right]
\]
and we are done.

Proof of Theorem 4.7. Inequality (9) is easily established by the fact that
\[
E \left[ X^2 \right] = 1.
\]
Now suppose that the three additional conditions in Theorem 4.7 are fulfilled. The fact that
\[
\liminf_{n \to \infty} \frac{-\sigma_n^2}{n} > 0
\]
allows us to choose \( \delta > 0 \) and \( n_0 \) such that for all \( n \geq n_0 \)
\[
\sigma_n^2 \geq \delta n. \tag{14}
\]
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 \( n \)
\[
\frac{1}{n} \sum_{k=1}^{n} p_k \sigma_k^2 \leq K. \tag{15}
\]
Now fix \( \gamma > 0 \) small. Then for \( n \) so large that
\[
\lceil \gamma n \rceil \geq n_0 \tag{16}
\]
and for \( k \) such that
\[
\lceil \gamma n \rceil \leq k \leq n \tag{17}
\]
we have, by (16), (17), (15) and (14),
\[
\left( \frac{s_n}{\sigma_k} \right)^2 = \frac{\sum_{k=1}^{n} (1 - p_k) + \sum_{k=1}^{n} p_k \sigma_k^2}{\sigma_k^2} \leq \frac{\sum_{k=1}^{n} (1 - p_k) + \sum_{k=1}^{n} p_k \sigma_k^2}{\delta k} \leq \frac{\sum_{k=1}^{n} (1 - p_k) + \sum_{k=1}^{n} p_k \sigma_k^2}{\delta \lceil \gamma n \rceil} \leq \frac{1}{\delta \gamma} \left( \frac{1}{n} \sum_{k=1}^{n} (1 - p_k) + \frac{1}{n} \sum_{k=1}^{n} p_k \sigma_k^2 \right) \leq \frac{1 + K}{\delta \gamma}
\]
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 \( F \) as cumulative distribution function. In particular,
\[
\sup_{\epsilon > 0} \limsup_{n \to \infty} \frac{1}{s^2_n} \sum_{k=[\gamma n]}^{n} p_k \sigma_k^2 \mathbb{E} \left[ X^2 ; |X| \geq \frac{\epsilon s_n}{\sigma_k} \right] \geq \sup_{\epsilon > 0} \limsup_{n \to \infty} \frac{1}{s^2_n} \sum_{k=[\gamma n]}^{n} p_k \sigma_k^2 \left( 1 + \frac{K}{\delta \gamma} \right)
\]
where the last equality follows from the fact that \( \mathbb{E} \left[ X^2 \right] = 1 \). Combining Lemma A.1 and the inequality shown by (18) gives
\[
\text{Lin} \left( \left\{ \frac{1}{s_n} (X_k - \mu) \right\} \right) \leq \limsup_{n \to \infty} \frac{1}{s^2_n} \sum_{k=[\gamma n]}^{n} p_k \sigma_k^2 \left( 1 + \frac{K}{\delta \gamma} \right)
\]
the last equality being seen by mimicking the proof of Lemma A.1. This finishes the proof.

**Proof of Theorem 4.8.** Theorem 4.7 gives
\[
\text{Lin} \left( \left\{ \frac{1}{s_n} (X_k - \mu) \right\} \right) \leq \limsup_{n \to \infty} \frac{1}{s^2_n} \sum_{k=1}^{n} p_k \sigma_k^2
\]
the last equality following from conditions (1) and (2) in Theorem 4.8. This establishes (10). If conditions (3) and (4) in Theorem 4.8 are also satisfied,
then Theorem 4.7 shows that the first inequality in the above calculation becomes an equality and we are done.

Proof of Theorem 4.9. Theorem 4.5 is applicable in order to conclude that the STA \( \{ \frac{1}{s_n} (X_k - \mu) \} \) satisfies Feller’s condition. Furthermore, Theorem 4.8 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 \( \frac{n}{s_n} (\bar{X}_n - \mu) \), it suffices to apply Theorem 3.2.

Proof of Theorem 5.1. First suppose that \( b < 1 \). Now, by (1),
\[
\frac{1}{s_n^2} \max_{k=1}^{n} \sigma_k^2 = \frac{n^b}{s_n^2} \leq n^{b-1}
\]
which clearly converges to 0 as \( n \) tends to \( \infty \). Thus condition (7) is satisfied which allows us to conclude from Theorem 4.4 that \( \bar{X}_n \) is AN. Also, Remark 4.3 shows that condition (6) holds whence we infer from Theorem 4.1 that \( \bar{X}_n \) is WC for \( \mu \). This establishes the first assertion.

Next, consider the case where \( b \geq 1 \) and \( a > b \). Then the sequence
\[
p_k \sigma_k^2 = p s_k^{2b-a}
\]
converges to 0 as \( k \) tends to \( \infty \) whence
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} p_k \sigma_k^2 = 0.
\]
Now it easily follows from Theorem 4.1 that \( \bar{X}_n \) is WC for \( \mu \) and from Theorem 4.9 that
\[
\limsup_{n \to \infty} K \left( \xi, \frac{n}{s_n} (\bar{X}_n - \mu) \right) = 0.
\]
Put otherwise, \( \bar{X}_n \) is AN and the second assertion is proved.

Finally, let \( b \geq 1 \) and \( a = b \). Then
\[
\frac{1}{n} \sum_{k=1}^{n} p_k \sigma_k^2 = p s_n^2.
\]
Now the proof of the third assertion goes along the same lines as the proof of the second one.
APPENDIX B: QQ-plots

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)
References

[1] 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.

[2] Berckmoes, B.; Lowen, R.; Van Casteren, J. Approach theory meets probability theory. Topology Appl. 158 (2011), no. 7, 836–852.

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

[4] Berckmoes, B.; Lowen, R.; Van Casteren, J. An isometric study of the Lindeberg-Feller CLT via Stein’s method J. Math. Anal. Appl. 405 (2013), no. 2, 484-498.

[5] Feller, W. An introduction to probability theory and its applications Vol. II. Second edition John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[6] Lowen, R. Approach Spaces: The Missing Link in the Topology-Uniformity-Metric Triad Oxford Mathematical Monographs, Oxford University Press, 1997.

[7] Lowen, R. Index Analysis: Approach Theory at Work, Springer scheduled to appear in 2015.

[8] 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.