A note on the normal approximation error for randomly weighted self-normalized sums

Siegfried Hörmann[1] and Yvik Swan[2]

Département de Mathématique, Université Libre de Bruxelles, Bd. Triomphe, CP210, 1050 Brussels, Belgium.[3]

Abstract

Let $\mathbf{X} = \{X_n\}_{n \geq 1}$ and $\mathbf{Y} = \{Y_n\}_{n \geq 1}$ be two independent random sequences. We obtain rates of convergence to the normal law of randomly weighted self-normalized sums

$$\psi_n(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^{n} X_i Y_i / V_n, \quad V_n = \sqrt{Y_1^2 + \cdots + Y_n^2}.$$

These rates are seen to hold for the convergence of a number of important statistics, such as for instance Student's t-statistic or the empirical correlation coefficient.

1   Introduction

Let $\mathbf{X} = \{X_n\}_{n \geq 1}$ and $\mathbf{Y} = \{Y_n\}_{n \geq 1}$ be two random sequences. In this paper we investigate the rate of convergence to the normal distribution of the randomly weighted self-normalized sums

$$\psi_n = \psi_n(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^{n} X_i Y_i / V_n, \quad V_n = \sqrt{Y_1^2 + \cdots + Y_n^2}. \tag{1}$$

The random variables ψ_n appear in some important statistics. For example, when testing the null that the mean of a population Y is equal to 0, one uses the Student t-statistic

$$T_n = \frac{\sqrt{n}\bar{Y}}{\left(\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2\right)^{1/2}}.$$

Denoting by $\psi_n = \psi_n(\mathbf{1}, \mathbf{Y}) = \sum_{i=1}^{n} Y_i / V_n$ the usual self-normalized partial sums, one can easily see that

$$T_n = \psi_n [(n-1)/(n-\psi_n^2)]^{1/2},$$

so that ψ_n and T_n are equivalent (in terms of a 1:1 correspondence). See, e.g., Efron [10], Logan *et al.* [15] or Giné *et al.* [12] for a discussion.

More generally, we could phrase the above testing problem as $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$ in the linear model $Z_i = \beta X_i + Y_i$. (The setup $X_i = 1$ for all $1 \leq i \leq n$ is

---

[1]Research supported by the Banque Nationale de Belgique and the Communauté française de Belgique - Actions de Recherche Concertées.

[2]Research supported by a Mandat de Chargé de Recherche from the Fonds National de la Recherche Scientifique, Communauté française de Belgique.

[3]E-mail: `shormann@ulb.ac.be` and `yvswan@ulb.ac.be`

contained as a special case.) Then $\psi_n(\mathbf{X}, \mathbf{Z})$, which reduces under H_0 to $\psi_n(\mathbf{X}, \mathbf{Y})$, will serve as a natural test statistic. As a matter of fact our research was originally motivated by this problem (see Hallin *et al.* [13]). We were interested in obtaining asymptotic normality of this test under as general as possible assumptions on the errors Y_i.

Another related example where ψ_n appears is the empirical correlation coefficient. If the sequences $\mathbf{X}$ and $\mathbf{Y}$ are centered then the empirical correlation is

$$\rho_n(\mathbf{X}, \mathbf{Y}) = \psi_n(\mathbf{X}, \mathbf{Y})/B_n = \frac{\sum_{k=1}^n X_k Y_k}{B_n V_n}, \quad B_n = \sqrt{X_1^2 + \cdots + X_n^2}.$$

We will see how, under moment conditions on $\mathbf{X}$, convergence rates for ψ_n can be transfered to ρ_n (see Lemma 2.2 below).

Besides their statistical applications, self-normalized sums have proven to be challenging mathematical objects with interesting properties. As a consequence they have attracted considerable attention in probability theory. For example Logan *et al.* [15] studied the limiting distributions of $\psi_n(\mathbf{1}, \mathbf{Y})$ when $\mathbf{Y}$ is a centered i.i.d. sequence with heavy tails, and conjectured that $\psi_n(\mathbf{1}, \mathbf{Y})$ is asymptotically normal if and only if $\mathbf{Y}$ is in the domain of attraction of the normal law. Giné *et al.* [12] proved that this conjecture holds true, while Chistyakov and Götze [9] settled the question of the convergence of Student's statistic by giving necessary and sufficient conditions for these sums to allow limiting distributions which are not concentrated on $\{\pm 1\}$. More recently Benktus *et al.* [3] studied the limiting distribution of the non-central t-statistic under different assumptions on $\mathbf{Y}$; they show, inter alia, how this limiting distribution depends critically on the existence of fourth moments for the Y_i. For a comprehensive study of these and related questions we refer the reader to the book Lai *et al.* [14].

In a slightly different setup, Breiman [6] provides necessary and sufficient conditions for the weak convergence of randomly weighted self-normalized sums of the form $\sum_i X_i Y_i / \sum_i Y_i$. Mason and Zinn [16] settle several questions left open by Breiman [6], and deduce the asymptotic distribution of $\psi_n(\mathbf{1}, \mathbf{Y})$ in the case of symmetry.

In this paper we will be interested in the rate of convergence to the normal distribution of $\psi_n(\mathbf{X}, \mathbf{Y})$ as well as of $\rho_n(\mathbf{X}, \mathbf{Y})$. The case when $\mathbf{X} = \mathbf{1}$ and $\{Y_i\}$ are independent with finite variance is already well established. Bentkus *et al.* [2] give sharp rates for convergence of Student's statistic, and thus equivalently for $\psi_n(\mathbf{1}, \mathbf{Y})$, in the non-i.i.d. case. Explicit constants in these bounds were derived by Shao [18]. See also Bentkus and Götze [4] for further references. There seem to be no similar investigations for $\psi_n(\mathbf{X}, \mathbf{Y})$. To the best of our knowledge, no similar results for the convergence rate of the correlation coefficient $\rho_n(\mathbf{X}, \mathbf{Y})$ exist.

Our approach is as follows. We first state in Lemma 2.1 a general result, which is simple to prove and which provides a bound that holds without *any* hypothesis on $\mathbf{Y}$, be it on its moments or dependence structure. The main target is then to work out the thus obtained rates explicitly by imposing different assumptions on the sequence $\{Y_i\}$. This is done through a number of subsequent results. An interesting feature in our approach is that (with one exception) we do not work with truncation arguments, even when assuming an infinite variance for the Y's.

2 Results

Recall that if P and Q are any probability measures on the real line, then the *Wasserstein distance* is given by

$$d_W(P,Q) = \sup_{h \in \mathcal{H}} \left| \int h dP - \int h dQ \right|,$$

where $\mathcal{H}$ is the class of Lipschitz 1 functions, i.e. $\mathcal{H} = \{h : \mathbb{R} \to \mathbb{R}; \|h'\| \leq 1\}$ with $\|f\| = \sup_{x \in \mathbb{R}} |f(x)|$. The *Kolmogorov distance* $d_K(P,Q)$ is defined similarly, with $\mathcal{H}$ replaced by the class of indicator functions $h_z(\cdot) = I\{\cdot \leq z\}$, $z \in \mathbb{R}$. If V and S are random variables on the space $(\Omega, \mathcal{A}, P)$ then $d_W(V, S)$ will be written for $d_W(P \circ V^{-1}, P \circ S^{-1})$, where $P \circ V^{-1}$ is the image measure of V under P. Similar is the definition for $d_K(V, S)$.

Throughout Z stands for a standard normal random variable and we are interested in

$$d_W\big(\psi_n(\mathbf{X}, \mathbf{Y}), Z\big) \quad \text{and} \quad d_W\big(\rho_n(\mathbf{X}, \mathbf{Y}), Z\big),$$

and

$$d_K\big(\psi_n(\mathbf{X}, \mathbf{Y}), Z\big) \quad \text{and} \quad d_K\big(\rho_n(\mathbf{X}, \mathbf{Y}), Z\big),$$

under the assumption that $\mathbf{X}$ and $\mathbf{Y}$ are independent.

The following simple Lemma gives the first step in our approach.

Lemma 2.1. *Let $\psi_n(\mathbf{X}, \mathbf{Y})$ be defined as in* (1), *where $\mathbf{X}$ and $\mathbf{Y}$ are two mutually independent sequences. Assume that $\{X_k\}$ is i.i.d. with $EX_1 = 0$, $EX_1^2 = 1$, $\xi_3 = E|X_1|^3 < \infty$. Then*

$$d_K\big(\psi_n(\mathbf{X}, \mathbf{Y}), Z\big) \leq 0.56\,\xi_3 \Delta, \tag{2}$$

where

$$\Delta = \sum_{k=1}^{n} E|\delta_{k,n}|^3 \quad \text{with} \quad \delta_{k,n} = Y_k/V_n. \tag{3}$$

Furthermore

$$d_W\big(\psi_n(\mathbf{X}, \mathbf{Y}), Z\big) \leq \xi_3 \Delta. \tag{4}$$

Proof of Lemma 2.1. We show first (2). Let $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$, $F_n(\mathbf{y}_n)$ be the joint law of $\mathbf{Y}_n$ and set $v_n^2 = \sum_{i=1}^{n} y_i^2$. Then using a version of the Berry-Esseen theorem for independent random variables, we obtain for any $z \in \mathbb{R}$

$$|P(Z \leq z) - P(\psi_n(\mathbf{X}, \mathbf{Y}) \leq z)| = \left| \int_{\mathbb{R}^n} P(Z \leq z) - P(\psi_n(\mathbf{X}, \mathbf{Y}) \leq z | \mathbf{Y}_n = \mathbf{y}_n) dF_n(\mathbf{y}_n) \right|$$

$$\leq \int_{\mathbb{R}^n} |P(Z \leq z) - P(\psi_n(\mathbf{X}, \mathbf{y}_n) \leq z)| \, dF_n(\mathbf{y}_n)$$

$$\leq CE|X_1|^3 \int_{\mathbb{R}^n} \sum_{i=1}^{n} (y_i/v_n)^3 dF_n(\mathbf{y}_n) = C\xi_3 E\Delta.$$

By a recent result of Shevtsova [19], $C \leq 0.56$.

The proof of (4) can be done in the exact same way, using Corollary 4.2 in [8]. $\qquad \square$

We remark that in Lemma 2.1 we do not put *any* restrictions on the sequence $\{Y_k\}$. This means that, in theory, we can obtain non-trivial bounds even if this sequence is not independent or identically distributed. Of course, the difficulty then resides in working

out Δ explicitly, which we do under different assumptions in Theorems 2.1, 2.2 and 2.3 below. We will see that Lemma 2.1 provides optimal bounds in several special cases.

Let us consider first the following special case, which gives an application to self-normalized sums $\psi_n(\mathbf{1}, \mathbf{Y})$ when the Y_i are not necessarily independent nor identically distributed. We assume instead that

$$(Y_1, \ldots, Y_n) \stackrel{d}{=} (\pm Y_1, \ldots, \pm Y_n) \tag{5}$$

for all choices of $+, -$. This form of symmetry, known as *sign-symmetry*, is more general than spherical symmetry (see e.g. Serfling [17]) and is to be likened with the concept of *orthant symmetry* discussed by Efron [10]. Sign-symmetry is obviously satisfied if the Y_i are symmetric and independent random variables. Under this condition the following result (which should be also compared to Mason and Zinn [16, Corollary 6]) holds.

Corollary 2.1. *Assume that* (5) *holds and set* $S_n = \sum_{k=1}^n Y_k$. *Then*

$$d_W\left(S_n/V_n, Z\right) \leq \Delta \quad and \quad d_K\left(S_n/V_n, Z\right) \leq 0.56\, \Delta,$$

with $\Delta = \sum_{k=1}^n E|\delta_{k,n}|^3$.

The proof follows simply by applying Lemma 2.1 to $\psi_n(\mathbf{X}, \mathbf{Y})$ with $\{X_k\}$ i.i.d. Rademacher variables, i.e. $X_k = \pm 1$ with probability $1/2$. Then due to the symmetric distribution of the Y_k, we have that S_n/V_n and $\psi_n(\mathbf{X}, \mathbf{Y})$ have the same distribution.

The next Lemma gives a simple criterion for switching from $\psi_n(\mathbf{X}, \mathbf{Y})$ to $\rho_n(\mathbf{X}, \mathbf{Y})$. While we impose 4 moments for X_1, we keep the assumptions on Y_1 general.

Lemma 2.2. *Let the assumptions of Lemma 2.1 hold and assume in addition that* $m_4 := EX_1^4 < \infty$. *Then, if* $n/\log n \geq 8m_4$,

$$d_W(\psi_n(\mathbf{X}, \mathbf{Y}), \sqrt{n}\rho_n(\mathbf{X}, \mathbf{Y})) \leq \sqrt{\frac{2m_4}{n}}.$$

Let us consider once more the testing problem $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$ in the linear model $Z_i = \beta X_i + Y_i$. The previous lemma in connection with Lemma 2.1 shows, if the regressors X_i are centered (a condition which is convenient but could be modified) and have 4 moments then we get under H_0 for very general errors Y_i the convergence of the correlation test statistic $\sqrt{n}\rho_n(\mathbf{X}, \mathbf{Z})$ to the normal, with an approximation error of order $O(n^{-1/2} + \Delta)$.

When $\{Y_k\}$ is a stationary sequence, then $\Delta = nE|\delta_{1,n}|^3$ and obtaining a rate of convergence to the normal distribution is entirely reduced to calculating the third absolute moment of $\delta_{1,n}$. We now concentrate on obtaining Δ under different moment and tail assumptions on the sequence $\{Y_k\}$ under the i.i.d. setup. We first work out Δ under the sole assumption $E|Y_1|^p < \infty$, $p \in (2, 3]$. In this case we obtain the "usual" convergence rates.

Theorem 2.1. *Let* $\{Y_i\}$ *be an i.i.d. sequence, let* $p \in (2, 3]$ *and assume* $EY_1^2 = 1$ *and* $E|Y_1|^p < \infty$. *Then*

$$\Delta = nE\left|\delta_{1,n}\right|^3 \leq nE\left|\delta_{1,n}\right|^p \sim E|Y_1|^p\, n^{1-p/2}. \tag{6}$$

(Note that the first inequality in (6) follows from $|\delta_{1,n}| \leq 1$.)

Remark 2.1. *A look at the proof of Theorem 2.1 suggests that similar results may be obtained under different dependence conditions, too. In fact, besides some purely analytic estimates, which hold for any sequence $\{Y_k\}$, we only make use of moment inequalities which exist in different generality for many weak dependence and mixing concepts, respectively.*

Next we consider the case when we have knowledge on the tail probabilities of the Y_k. Let $\Gamma(p)$ denote Euler's gamma function.

Theorem 2.2. *Let $\{Y_i\}$ be an i.i.d. sequence, let $1 \leq \alpha < 2$ and $P(Y_k^2 > x) \sim \ell(x)x^{-\alpha}$, where $\ell(x)$ is slowly varying at ∞. If $\sigma_Y^2 := EY_1^2 < \infty$, then we have for any $\gamma > \alpha$*

$$nE\,|\delta_{1,n}|^{2\gamma} \sim \frac{\Gamma(\gamma - \alpha)\Gamma(1+\alpha)}{\sigma_Y^2\Gamma(\gamma)}\, n^{1-\alpha}\ell(n).$$

Example 2.1. *Consider the case $P(|Y_1|^p > x) \sim \frac{1}{x\log^2 x}$, $p \in (2,3)$. Then $E|Y_1|^\beta = \infty$ for any $\beta > p$, while $E|Y_1|^p < \infty$. Applying the above result with $\gamma = 3/2$ we obtain*

$$\Delta = nE\,|\delta_{1,n}|^3 \sim \frac{\Gamma((3-p)/2)\Gamma(1+p/2)}{\sigma_Y^2\Gamma(3/2)}\, \frac{n^{1-p/2}}{\log^2 n}.$$

Hence the additional knowledge of the tail behavior yields a slightly better rate than the one obtained in (6).

We now turn to the case when we have infinite second moments.

Theorem 2.3. *Let $\{Y_i\}$ be an i.i.d. sequence, let $P(Y_1^2 > x) \sim \ell(x)x^{-1}$, with $\ell(x)$ slowly varying at ∞. If $E(Y_1^2) = \infty$, then, for any $\gamma > 1$, we have*

$$nE\,|\delta_{1,n}|^{2\gamma} \sim \frac{1}{\gamma-1}\frac{\ell(a_n)}{L(a_n)},$$

where $L(x) = \int^x \left(\ell(t)/t\right)dt$ and $\{a_n\}$ is a sequence satisfying $a_n \sim nL(a_n)$.

Example 2.2. *Assume that $P(Y_1^2 > x) \sim x^{-1}(\log x)^{-2}$. Then $EY_1^2 < \infty$ and by Theorem 2.2 we get that for all $n \geq 1$*

$$d_K\big(\psi_n(\mathbf{X},\mathbf{Y}),Z\big) \leq A(\log n)^{-2},$$

for some large enough constant A. If $P(Y_k^2 > x) \sim x^{-1}(\log x)^{-1}$ then $EY_1^2 = \infty$. But since $\tilde{\ell}(x) = \log\log x$ and $a_n \sim n\log\log n$ we still get by Theorem 2.3 applied with $\gamma = 3/2$

$$d_K\big(\psi_n(\mathbf{X},\mathbf{Y}),Z\big) \leq A\big((\log n)\log\log n\big)^{-1},$$

and thus an explicit convergence rate to the normal law.

We conclude with a result which shows that, even in the case when Y_1 is in the domain of attraction of an α-stable law with α strictly less but close to 2 we can get non-trivial bounds.

Theorem 2.4. *Let $\{Y_i\}$ be an i.i.d. sequence, assume that $P(|Y_1| > x) \sim \ell(x)x^{-\alpha}$ with $\alpha \in (0,2)$. Then for any $\gamma > 1$*

$$nE\,|\delta_{1,n}|^{2\gamma} \sim \frac{\Gamma(\gamma - \alpha/2)}{\Gamma(\gamma)\Gamma(1-\alpha/2)}. \tag{7}$$

5

Example 2.3. *We apply this result with $\gamma = 3/2$. Let $\alpha = 2 - \varepsilon$ for small $\varepsilon > 0$. Observing that under the above assumptions $\Gamma(3/2 - \alpha/2)/\Gamma(3/2) < 2$, and $1/\Gamma(\varepsilon) \sim \varepsilon$ for $\varepsilon \to 0$ we get by Lemma 2.1 that for large enough n*

$$d_K(\psi_n(\mathbf{X}, \mathbf{Y}), Z) \leq 0.56 \, \xi_3 \, \varepsilon.$$

When the distribution of the Y_k is symmetric, we can conclude that for sufficiently large sample size n we have $d_K(S_n/V_n, Z) \leq 0.56 \, \varepsilon$.

3 Proofs

In the sequel we need the following version of Hoeffding's inequality (see e.g. Shao [18, p. 145]).

Lemma 3.1. *Let $\{Z_i, 1 \leq i \leq n\}$ be independent non-negative random variables with $\mu = \sum_{i=1}^{n} EZ_i$ and $\sigma^2 = \sum_{i=1}^{n} EZ_i^2 < \infty$. Then for $0 < x < \mu$*

$$P\left(\sum_{i=1}^{n} Z_i \leq x\right) \leq \exp\left(-\frac{(\mu - x)^2}{2\sigma^2}\right).$$

Proof of Lemma 2.2. Let

$$\phi := \sqrt{n}\rho_n(\mathbf{X}, \mathbf{Y}).$$

Note that $\phi = \sqrt{\frac{n}{B_n^2}}\psi$. Then, for all $h \in \mathcal{H}$, by the mean value theorem we get (recall that $\|h'\| \leq 1$)

$$|Eh(\phi) - Eh(\psi)| \leq E\left[\left|\sqrt{\frac{n}{B_n^2}} - 1\right| |\psi|\right].$$

Let $\varepsilon \in (0, 1)$. Then the last term is bounded by $\frac{1}{2}A_n^{(1)} + A_n^{(2)}$, where

$$A_n^{(1)} := E\left[\left|\frac{n}{B_n^2} - 1\right| |\psi| I\{B_n^2 > (1 - \varepsilon)n\}\right]$$

and

$$A_n^{(2)} := E\left[\left|\sqrt{\frac{n}{B_n^2}} - 1\right| |\psi| I\{B_n^2 \leq (1 - \varepsilon)n\}\right].$$

Define $\kappa_4 = E(X_1^2 - 1)^2$. Then

$$A_n^{(1)} \leq \frac{1}{1 - \varepsilon}E\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 1)\right| |\psi|\right]$$

$$\leq \frac{1}{1 - \varepsilon}\left(E\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 1)\right|^2\right]\right)^{1/2} \times \left(E|\psi|^2\right)^{1/2}$$

$$= \frac{1}{1 - \varepsilon}\sqrt{\frac{\kappa_4}{n}}.$$

For estimating $A_n^{(2)}$ we use

$$|\psi| = \left|\sum_{i=1}^{n} X_i\frac{Y_i}{V_n}\right| \leq \left(\sum_{i=1}^{n} X_i^2\right)^{1/2}\left(\sum_{i=1}^{n} \frac{Y_i^2}{V_n^2}\right)^{1/2} = B_n.$$

Thus

$$A_n^{(2)} \leq E|\sqrt{n} - B_n|I\{B_n^2 \leq (1-\varepsilon)n\} \leq \sqrt{n}P(B_n^2 \leq (1-\varepsilon)n).$$

By Lemma 3.1 we get that $P(B_n^2 \leq (1-\varepsilon)n) \leq \exp(-\varepsilon^2 n/(2m_4))$. Collecting our estimates we have

$$|Eh(\phi) - Eh(\psi)| \leq \frac{1}{2}\frac{1}{1-\varepsilon}\sqrt{\frac{\kappa_4}{n}} + \sqrt{n}\exp(-\varepsilon^2 n/(2m_4)).$$

For large enough n we have $\varepsilon^2 := (2m_4 \log n)/n \leq 1/4$, and since $m_4 = \kappa_4 + 1$ we conclude

$$|Eh(\phi) - Eh(\psi)| \leq \sqrt{\frac{\kappa_4}{n}} + \frac{1}{\sqrt{n}} \leq \sqrt{\frac{2m_4}{n}}.$$

$\square$

Proof of Theorem 2.1. Let $\tilde{Y}_{i,n} = Y_i \wedge n^{1/p}$. Then

$$E\tilde{Y}_{i,n}^2 = 1 - \varepsilon_n \quad \text{with } \varepsilon_n \to 0. \tag{8}$$

Further

$$E\tilde{Y}_{i,n}^4 = \int_0^{n^{1/p}} x^4 dP(Y_1 \leq x) \leq n^{2/p}\int_0^\infty x^2 dP(Y_1 \leq x) = n^{2/p}. \tag{9}$$

Now fix an arbitrarily small $\varepsilon > 0$ and let n be large enough in order to have $\varepsilon_n \leq \varepsilon/2$. Using Lemma 3.1 with (8) and (9) it follows that

$$P\left(\tilde{Y}_{2,n}^2 + \cdots + \tilde{Y}_{n,n}^2 \leq (1-\varepsilon)(n-1)\right) \leq \exp\left(-\frac{\varepsilon^2}{8}(n-1)^{1-2/p}\right). \tag{10}$$

Next we observe that

$$\left(\frac{Y_1^2}{Y_1^2 + \cdots + Y_n^2}\right)^{p/2} \leq \min\left\{1, \frac{|Y_1|^p}{(\tilde{Y}_{2,n}^2 + \cdots + \tilde{Y}_{n,n}^2)^{p/2}}\right\}$$

$$\leq I\{\tilde{Y}_{2,n}^2 + \cdots + \tilde{Y}_{n,n}^2 \leq (1-\varepsilon)n\}$$

$$+ |Y_1|^p\left(\frac{1}{n(1-\varepsilon)}\right)^{p/2}I\{\tilde{Y}_{2,n}^2 + \cdots + \tilde{Y}_{n,n}^2 > (1-\varepsilon)n\}$$

This and (10) give

$$nE|\delta_{1,n}|^p = nE\left(\frac{Y_1^2}{\sum_{k=1}^n Y_k^2}\right)^{p/2}$$

$$\leq nP(\tilde{Y}_{2,n}^2 + \cdots + \tilde{Y}_{n,n}^2 \leq (1-\varepsilon)n) + E|Y_1|^p\left(\frac{1}{1-\varepsilon}\right)^{p/2}n^{1-p/2}$$

$$\sim E|Y_1|^p\left(\frac{1}{1-\varepsilon}\right)^{p/2}n^{1-p/2} \quad (n \to \infty). \tag{11}$$

On the other hand we have

$$nE\left(\frac{Y_1^2}{\sum_{k=1}^n Y_k^2}\right)^{p/2}$$

$$\geq E|Y_1|^p I\{|Y_1|^p < n\} E\left(\frac{1}{n^{-1}\sum_{k=2}^n Y_k^2 + n^{(2-p)/p}}\right)^{p/2} n^{1-p/2}$$

$$\geq E|Y_1|^p I\{|Y_1|^p < n\}\left(\frac{1}{1 + n^{(2-p)/p}}\right)^{p/2} P\left(\frac{1}{n}\sum_{k=2}^n Y_k^2 \leq 1\right) n^{1-p/2}$$

$$\sim E|Y_1|^p n^{1-p/2} \quad (n \to \infty), \tag{12}$$

where in the last step we used the law of large numbers to obtain $P(\frac{1}{n}\sum_{k=2}^n Y_k^2 \leq 1) \to 1$. Now (11) holds for arbitrarily small $\varepsilon > 0$. Together with (12) the proof follows. $\qquad\square$

Proof of Theorem 2.3. We borrow an idea of Albrecher and Teugels [1]. The crucial trick is to write

$$\frac{1}{x^\gamma} = \frac{1}{\Gamma(\gamma)}\int_0^\infty s^{\gamma-1}e^{-sx}ds, \quad \gamma > 0.$$

Then, since

$$E|\delta_{1,n}^2|^\gamma = E\left|\frac{Y_1^2}{\sum_{k=1}^n Y_k^2}\right|^\gamma,$$

we obtain

$$E|\delta_{1,n}^2|^\gamma = \frac{1}{\Gamma(\gamma)}\int_0^\infty s^{\gamma-1}(\varphi_1(s))^{n-1}\varphi_2(s)ds,$$

with $\varphi_1(s) := E(e^{-sY_1^2})$ and $\varphi_2(s) := E\left([Y_1^2]^\gamma e^{-sY_1^2}\right)$. Choosing $a_n \to \infty$ such that $na_n^{-1}L(a_n) \to 1$ one easily shows (see [1]) that for any $s > 0$

$$\lim_{n\to\infty}\varphi_1^{n-1}\left(\frac{s}{a_n}\right) = e^{-s}. \tag{13}$$

In order to determine $\varphi_2(s)$ we introduce

$$\Gamma_{\gamma,s}(x) = \int_0^x t^\gamma e^{-st}dt, \quad \gamma > 1, \quad s, x > 0.$$

Note that $\lim_{x\to\infty}\Gamma_{p,s}(x) = s^{-(\gamma+1)}\Gamma(\gamma+1)$. Further we let F be the distribution function of Y_1^2. Using integration by parts, we get

$$\int_0^\infty F(x)d\Gamma_{\gamma,s}(x) = s^{-(\gamma+1)}\Gamma(\gamma+1) - \int_0^\infty \Gamma_{\gamma,s}(x)dF(x).$$

A simple consequence is that

$$\int_0^\infty \Gamma_{\gamma,s}(x)dF(x) = \int_0^\infty (1 - F(x))d\Gamma_{\gamma,s}(x) = \int_0^\infty x^\gamma e^{-sx}(1 - F(x))dx.$$

Since $\gamma\Gamma_{\gamma-1,s}(x) - s\Gamma_{\gamma,s}(x) = x^\gamma e^{-sx}$ we conclude that

$$\varphi_2(s) = \int_0^\infty x^\gamma e^{-sx}dF(x)$$

$$= \gamma\int_0^\infty \Gamma_{\gamma-1,s}(x)dF(x) - s\int_0^\infty \Gamma_{\gamma,s}(x)dF(x)$$

$$= \gamma\int_0^\infty x^{\gamma-1}(1 - F(x))e^{-sx}dx - s\int_0^\infty x^\gamma(1 - F(x))e^{-sx}dx.$$

8

By our assumption $1 - F(x) \sim x^{-1}\ell(x)$ and thus by Karamata's Tauber theorem (see e.g. Bingham *et al.* [5, Theorem 1.7.6]) we have for any $\rho > -1$

$$\int_0^\infty x^\rho (1 - F(x)) e^{-sx} dx \sim \Gamma(\rho) s^{-\rho} \ell(1/s) \quad \text{as} \quad s \to 0.$$

Combining with our just derived formula for $\varphi_2(s)$ we have for $\gamma > 1$

$$\varphi_2(s) \sim \frac{\ell(1/s)}{s^{\gamma-1}} \Gamma(\gamma - 1) \quad \text{as } s \to 0. \tag{14}$$

Finally consider the quantity

$$nE\big|\delta_{1,n}^2\big|^\gamma = \frac{n}{\Gamma(\gamma)} \int_0^\infty t^{\gamma-1} \varphi_1^{n-1}(t) \varphi_2(t) dt.$$

It is easy to show that $n \int_\epsilon^\infty t^{\gamma-1} \varphi_1^{n-1}(t) \varphi_2(t) dt \to 0$ for all $\epsilon > 0$. Hence we can restrict the integration to the compact interval $[0, \epsilon]$, on which Lemma 5.1 of Fuchs *et al.* [11] can be used to uniformly bound the integrand above by an integrable function. Using the already defined a_n, we therefore get from (13), (14) and dominated convergence

$$
\begin{aligned}
nE\big|\delta_{1,n}^2\big|^\gamma &\sim \frac{n}{\Gamma(\gamma)} \int_0^\epsilon t^{\gamma-1} \varphi_1^{n-1}(t) \varphi_2(t) dt \\
&\sim \frac{n}{\Gamma(\gamma)} \left(\frac{1}{a_n}\right)^\gamma \int_0^\infty t^{\gamma-1} \varphi_1^{n-1}(t/a_n) \varphi_2(t/a_n) dt \\
&\sim \frac{\Gamma(\gamma - 1)}{\Gamma(\gamma)} \frac{n\ell(a_n)}{a_n}, \quad \text{as } n \to \infty.
\end{aligned}
$$

The relation $\frac{n\ell(a_n)}{a_n} \sim \frac{\ell(a_n)}{L(a_n)}$ concludes the proof. $\qquad\square$

The proof of Theorem 2.2 is similar to the proof of Theorem 2.3 and will therefore be omitted.

Acknowledgments.

References

[1] Albrecher, H., and Teugels, J. (2006). Asymptotic analysis of a measure of variation. *Theor. Probab. Math. Stat.*, **74,** 1–9.

[2] Bentkus, V., Bloznelis, M., and Götze, F. (1996). A Berry-Esséen bound for Student's statistic in the non-i.i.d. case. *J. Theor. Probab.*, **9,** 765–796.

[3] Bentkus, V., Bing-Yi, J., M., Shao, Q.M. and Wang, Z. (2007). Limiting distributions of the non-central t-statistic and their applications to the power of t-tests under non-normality. *Bernoulli*, **13,** 346–364.

[4] Bentkus, V., and Götze, F. (1996). The Berry-Esséen bound for Student's statistic. *Ann. Probab.*, **24,** 491–503.

[5] Bingham, N.H., Goldie, C.M., and Teugels, J.L. (1987). *Regular Variation.* Cambridge University Press.

[6] Breiman, L. (1965). On some limit theorems similar to the arc-sin law. *Teor. Veroyatnost. i Primenen.*, **10,** 351–359.

[7] Chen, L.H.Y, and Shao, Q-M. (2004). Stein's method for normal approximation. In *An Introduction to Stein's Method* (A.D. Barbour and L.H.Y. Chen eds). Lecture Notes Series, Institute for Mathematical Sciences, NUS, Vol. 4, p. 1-59.

[8] Chen, L.H.Y, Goldstein, L. and Shao, Q-M. (2011). Normal Approximation by Stein's Method. *Springer Series in Probability and its Application.*

[9] Chistyakov, G. P. and Götze, F. (2004). Limit distributions of studentized means. *Ann. Probab.*, **32,** No. 1 A, 28–77.

[10] Efron, B. (1969). Student's *t*-test under symmetry conditions. *JASA*, **64,** 1278–1302.

[11] Fuchs, A., Joffe, A. and Teugels, J. (2001). Expectation of the Ratio of the Sum of Squares to the Square of the Sum: Exact and Asymptotic Results. *Teor. Veroyatnost. i Primenen*, **46,** 297–310.

[12] Giné, E., Götze, F., and Mason, D.M. (1997). When is the student *t*-statistic asymptotically normal? *Ann. Probab.*, **25,** 1514–1531.

[13] Hallin, M., Swan, Y., Verdebout, T. and Veredas, D. (2010). Rank-based Inference in Linear Models with Stable Errors. *JNPS*, forthcoming.

[14] Lai, T.L., de la Pena, V. and Shao, Q. M. (2009). *Self-normalized Processes: Theory and Statistical Applications.* Springer Series in Probability and its Applications, Springer-Verlag, New York.

[15] Logan, B.F., Mallows, C.L., Rice, S.O., and Shepp, L.A. (1973). Limit distributions of self-normalized sums. *Ann. Probab.*, **5,** 788–809.

[16] Mason, D., and Zinn, J. (2005). When does a self-normalized weighted sum converge in distribution? *Elec. Comm. in Probab.*, **10,** 70–81.

[17] Serfling, R. (2006). Multivariate symmetry and asymmetry. In: *Encyclopedia of Statistical Sciences, 2nd Ed. (Kotz, Balakrishnan, Read and Vidakovic, Eds.).* Wiley, 5338–5345.

[18] Shao, Q-M. (2005). An explicit Berry-Esseen bound for the Student t-statistic via Stein's method. In: *Stein's Method and Applications (A.D. Barbour and L.H.Y. Chen eds).* Lecture Notes Series, Institute for Mathematical Sciences, NUS, Vol. 5, 143–155.

[19] Shevtsova, I.G. (2010). An improvement of convergence rate estimates in the Lyapunov theorem. *Doklady Mathematics*, **82,** 862–864.