Self-normalized limit theorems: A survey

Qi-Man Shao[*]

Department of Statistics, The Chinese University of Hong Kong
Shatin, NT, Hong Kong
e-mail: qmshao@cuhk.edu.hk

and

Qiying Wang[†]

School of Mathematics and Statistics, University of Sydney
NSW 2006, Australia
e-mail: qiying@maths.usyd.edu.au

Abstract: Let $X_1, X_2, \ldots$, be independent random variables with $EX_i = 0$ and write $S_n = \sum_{i=1}^{n} X_i$ and $V_n^2 = \sum_{i=1}^{n} X_i^2$. This paper provides an overview of current developments on the functional central limit theorems (invariance principles), absolute and relative errors in the central limit theorems, moderate and large deviation theorems and saddle-point approximations for the self-normalized sum S_n/V_n. Other self-normalized limit theorems are also briefly discussed.

MSC 2010 subject classifications: Primary 60F05, 60F17; secondary 62E20.

Keywords and phrases: Self-normalized sum, Student t statistic, central limit theorem, invariance principle, convergence rate, absolute error, relative error, Cramér moderate deviation, large deviation, saddle-point approximation, laws of the iterated logarithm, Darling-Erdös theorem, Hotelling's T^2 statistic.

Contents

1. Introduction

Let $X, X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean zero, and put $\bar{X} = S_n/n$,

$$S_n = \sum_{j=1}^{n} X_j, \quad V_n^2 = \sum_{j=1}^{n} X_j^2 \quad \text{and} \quad \sigma_n^2 = \frac{n-1}{n} \sum_{j=1}^{n} (X_j - \bar{X})^2.$$

There has been increasing interest in the investigation of limit behaviors for the so-called self-normalized sum S_n/V_n in the past decade. One of the reasons for this flourish comes from the fact:

$$P\big(t_n \geq x\big) = P\big\{S_n/V_n \geq x[n/(n+x^2-1)]^{1/2}\big\}, \quad x \geq 0, \tag{1.1}$$

where $t_n = S_n/\sigma_n$ is the classical Student t-statistic. This allows us to only consider the distribution properties of the less complex S_n/V_n to discuss the distribution properties of t_n, which is frequently used in practice to test hypotheses about the mean. More importantly, the limit theorems for S_n/V_n (and hence for t_n) usually require much less stringent moment conditions than the classical limit theorems do, and hence provide much wide practical applicability.

The past decade has witnessed significant development in the arena of weakening moment conditions for self-normalized limit theorems. Griffin and Kuelbs (1989) [31] obtained a self-normalized law of the iterated logarithm for all distributions in the domain of attraction of a normal or stable law. Shao (1997) [55] showed that no moment conditions are needed for a self-normalized large deviation result $P(S_n/V_n \geq x\sqrt{n})$, and that the tail probability of S_n/V_n is Gaussian-like when X_1 is in the domain of attraction of the normal law and sub-Gaussian-like when X is in the domain of attraction of a stable law, while Giné, Götze and Mason (1997) [29] proved that the tails of S_n/V_n are uniformly sub-Gaussian when the sequence is stochastically bounded. Shao (1999) [57] established a Cramér type moderate deviation result for self-normalized sums only under a finite third moment condition. Jing, Shao and Wang (2003) [37] proved a Cramér type moderate deviation result (for independent random variables) under a Lindeberg type condition. Jing, Shao and Zhou (2004) [38] obtained the saddlepoint approximation without any moment conditions. Other results include Wang and Jing (1999) [69] as well as Robinson and Wang (2005) [54] for an exponential non-uniform Berry-Esseen bound; Csörgő, Szyszkowicz and Wang (2003a, b) [18, 19] for Darling-Erdős theorems and Donsker's theorems; Wang (2005) [66] as well as Wang and Hall (2009) [68] for a refined moderate deviation; Hall and Wang (2004) [35] for exact convergence rates, and Chistyakov and Götze (2004b) [13] for all possible limiting distributions when X is in the domain of attraction of a stable law. We also refer to de la Peňa, Lai and Shao

(2009) [52] for a systematic presentation on self-normalized processes and their statistical applications.

The main aim of this paper is to provide an overview of new developments on the functional central limit theorems (invariance principles), absolute and relative errors in the central limit theorems, Cramér-Chernoff-type large deviations and saddle-point approxiamtions for the S_n/V_n. Partial materials have been collected in [56, 58, 60]. We represent these here for the sake of completeness. Explicitly, Section 2 will review weak convergence properties of S_n/V_n, including central limit theorems and invariance principles. The absolute and relative errors in the central limit theorems for the S_n/V_n will be given in Section 3 and 4 respectively. Section 5 reviews Cramér-Chernoff-type large deviations and saddle-point approximations. Finally in Section 6 we briefly review other self-normalized limit theorems, like the self-normalized law of the iterated logarithm, Darling-Erdös type theorem, limit theorem for studentized non-linear statistics, etc. Throughout the paper, we assume that $X, X_1, \ldots, X_n$ are i.i.d. random variables with $EX = 0$, except for those explicitly specified.

2. The central limit theorem and invariance principle

Efron (1969) [25] might be the first paper to investigate the limit behavior of the Student's t-statistic t_n or, equivalently, the S_n/V_n, in some nonstandard cases. The general research begins with Logan, Mallows, Rice and Shep (1973) [44] (LMRS for short) in which the authors showed, among many other results, that if X is in the domain of attraction of an α-stable law, $0 < \alpha \le 2$, centered if $\alpha > 1$ and symmetric if $\alpha = 1$, then S_n/V_n converges in distribution to a *limit*, which is sub-Gaussian, and if moreover X is symmetric, then the moments of S_n/V_n also converge to the corresponding moments of the limit. LMRS also conjectured that S_n/V_n is asymptotically normal if (and perhaps only if) X is in DAN (the domain of attraction of the normal law) and the only possible nontrivial limit distributions of S_n/V_n are those obtained when X follows a stable law.

Based on Raikov' theorem, as was noticed by Maller (1981) [45], among others, one can easily show the "if" part in the conjectures of LMRS. We refer to Csörgő and Horváth (1988) [15], Griffin and Mason (1991) [32] for more details in this regard. It is the "only if" part that has remained open until 1997 for the general case of not necessarily symmetric random variables, when Giné, Götze and Mason (1997) [29] proved that

Theorem 2.1. $S_n/V_n \to_D N(0,1)$ *if and only if X is in the domain of attraction of the normal law.*

Giné, Götze and Mason (1997) [29] also showed that, *if the self-normalized sums S_n/V_n, $n \in N$, are stochastically bounded, then they are uniformly sub-Gaussian in the sense that*

$$\sup_{n \in N} E e^{t S_n/V_n} \le 2 e^{ct^2} \text{ for all } t \in R \text{ and some } c < \infty.$$

This, in turn, implies a basic requirement in the proof of this result that the moments of S_n/V_n converge to those of a $N(0,1)$ r.v. whenever S_n/V_n is asymptotically standard normal.

The second conjecture of LMRS was confirmed by Chistyakov and Götze (2004b) [13] who proved the following theorem.

Theorem 2.2. *S_n/V_n converge weakly to a random variable Z such that*

(a) *$P(|Z| = 1) < 1$ if and only if*

   (1) *X is in the domain of attraction of a stable law with $\alpha \in (0,2]$;*

   (2) *$EX = 0$ if $1 < \alpha \leq 2$;*

   (3) *if $\alpha = 1$, then X is in the domain of attraction of Cauchy's law and Feller's condition holds: $\lim_{n\to\infty} nE\sin(X/a_n)$ exists and is finite, where*

$$a_n = \inf\left\{x > 0 : nx^{-2}\left[1 + E\{X^2 I(|X| < x)\}\right] \leq 1\right\}.$$

(b) *$P(|Z| = 1) = 1$ if and only if $P(|X| > x)$ is a slowly varying function at $+\infty$.*

The proofs of Chistyakov and Götze (2004b) [13] are very technical. It would be interesting to find an alternative approach. In the independent, but not identically distributed case, Mason (2005) [46] considered self-normalized triangular arrays. The result in [46] is stated as follows.

Theorem 2.3. *Let $X_{1,n}, \ldots, X_{n,n}, n \geq 1$, be a triangular array of independent infinitesimal random variables. Assume that*

$$\left(\sum_{i=1}^{n} X_{i,n}, \sum_{i=1}^{n} X_{i,n}^2\right) \to_D (U, V),$$

for a nondegenerate pair (U, V). Then $\sum_{i=1}^{n} X_{i,n} / \sqrt{\sum_{i=1}^{n} X_{i,n}^2} \to_D N(0,1)$ if and only if for some $\tau > 0$,

$$\sum_{i=1}^{n} X_{i,n} \to_D \tau N(0,1) \quad and \quad \sum_{i=1}^{n} X_{i,n}^2 \to_P \tau^2.$$

Mason (2005) [46] also claimed other general results. For instance, their Theorem 1 leads to an alternative proof of the Giné, Götze and Mason (1997) [29] result. In the independent symmetric case, the following result from Egorov (1996) [26] is also of interest.

Theorem 2.4. *Let $X_1, X_2, \ldots$, be independent symmetric random variables around mean zero. Then $S_n/V_n \to_D N(0,1)$ if and only if*

$$\max_{1 \leq j \leq n} |X_j|/V_n \to_P 0, \qquad as \ n \to \infty. \tag{2.1}$$

Note that (2.1) is equivalent to the condition that X is in the domain of attraction of the normal law if $\{X_j, j \geq 1\}$ is a sequence of i.i.d. random variables (cf. O'Brien (1980) [48]). Also, it is readily seen that the Lindeberg con-

dition implies (2.1). However, it is not clear at this moment whether or not Theorem 2.4 still holds for general independent random variables, i.e., without assuming $\{X_j, j \geq 1\}$ to be symmetric. In the i.i.d. case, Theorem 2.4 has been previously proved in Griffin and Mason (1991) [32].

The extension of the self-normalized central limit theorem to Donsker type functional central limit theorem was established in Csörgő, Szyszkowicz and Wang (CsSzW) (2003b) [19]. Define $S_{[nt]} = \sum_{i=1}^{[nt]} X_i$. The following theorem comes from their Theorem 1.

Theorem 2.5. *As $n \to \infty$, the following statements are equivalent:*

- *(a)* $EX = 0$ *and X is in the domain of attraction of the normal law;*
- *(b)* $S_{[nt_0]}/V_n \to_{\mathcal{D}} N(0, t_0)$, *for $t_0 \in (0, 1]$;*
- *(c)* $S_{[nt]}/V_n \to_{\mathcal{D}} W(t)$ *on $(D[0,1], \rho)$, where ρ is the sup-norm metric for functions in $D[0,1]$, and $\{W(t), 0 \leq t \leq 1\}$ is a standard Wiener process;*
- *(d)* *On an appropriate probability space for $X, X_1, X_2, \ldots$, we can construct a standard Wiener process $\{W(t), 0 \leq t < \infty\}$ such that*

$$\sup_{0 \leq t \leq 1} \left| S_{[nt]}/V_n - W(nt)/\sqrt{n} \right| = o_P(1).$$

Assuming appropriate conditions, there are two immediate analogs of Theorem 2.5 when $\{X_j, j \geq 1\}$ is a sequence of independent random variables with $EX_j = 0$ and finite variances EX_j^2. Write $B_n^2 = \sum_{j=1}^n EX_j^2$. If the Lindeberg condition holds, namely

$$\text{for all } \epsilon > 0, \qquad B_n^{-2} \sum_{j=1}^n E\{X_j^2 I(|X_j| > \epsilon B_n)\} \to 0, \qquad \text{as } n \to \infty,$$

then it is readily seen that $V_n^2/B_n^2 \to_P 1$. Hence it follows easily from classical results (cf., e.g., Prohorov (1956) [53]) that $S_{K_n(t)}/V_n \to_{\mathcal{D}} W(t)$ on $(D[0,1], \rho)$, where $K_n(t) = \sup\{m : B_m^2 \leq tB_n^2\}$. By using a similar method as in the Theorem 2.5, we can also redefine $\{X_j, j \geq 1\}$ on a richer probability space together with a sequence of independent normal random variables $\{Y_j, j \geq 1\}$ with mean zero and $Var(Y_j) = Var(X_j)$ such that

$$\sup_{0 \leq t \leq 1} \left| S_{[nt]}/V_n - \sum_{j=1}^{[nt]} Y_j/B_n \right| = o_P(1)$$

provided that the Lindeberg condition holds. Furthermore, CsSzW (2003b) [19] also proved the following result for self-normalized, self-randomized partial sums processes of independent random variables.

Theorem 2.6. *Let $X_1, X_2, \ldots$, be independent symmetric random variables around mean zero. Then (2.1) holds true if and only if*

$$S_{\tilde{K}_n(t)}/V_n \to_{\mathcal{D}} W(t) \qquad on \ (D[0,1], \rho), \tag{2.2}$$

where $\tilde{K}_n(t) = \sup\{m : V_m^2 \leq tV_n^2\}$.

Currently, several authors considered the extensions of Theorems 2.5 and 2.6 to weight approximation and dependent settings. See, for example, CsSzW (2004, 2008) [20, 21], Kulik (2006) [40], Balan and Kulik (2009) [5] and Choi and Moon (2010) [14].

Since we assume $EX_j = 0$, the investigation of limit behaviors for S_n/V_n in Theorems 2.1–2.6 is only related to centralized Student t-statistics. The limit behaviors of the non-central Student t-statistic was discussed in Bentkus, Jing, Shao and Zhou (2007) (BJSZ for short) [8]. Under the assumption of $EX^2 < \infty$, the limit behaviors of the non-central limit t-statistic are shown to be different for the following two cases: (a) $EX^4 < \infty$; (b) $EX^4 = \infty$ but $EX^2 < \infty$. Define $\nu = \mu/\sigma$, where $\mu = EX$ and $\sigma^2 = var(X)$. Write $X \in B(p, \mu, \sigma^2)$ if $X = \sigma Y + \mu$, where Y is a standardized Bernoulli random variable. The following theorem comes from BJSZ's Theorems 1 and 2.

Theorem 2.7. *(a) Assume $EX^4 < \infty$.*

(i) For $X \in B(p, \mu, \sigma^2)$ such that $\nu = 2\sqrt{pq}/(q - p)$ and $p \neq 1/2$, we have

$$a'_n(t_n - \nu\sqrt{n-1}) \to_D N^2(0, 1),$$

where $a'_n = 2\nu\sqrt{n}/(1 + \nu^2)$.

(ii) For any X other than the one given in (i), we have

$$\sigma_0^{-1}(t_n - \nu\sqrt{n}) \to_D N(0, 1),$$

where $\sigma_0^2 = 1 - \nu\alpha_3 + \nu^2(\alpha_4 - 1)/4$, and $\alpha_k = E(X - \mu)^k/\sigma^k$, $k = 3, 4$.

(b) Assume $\mu \neq 0$, $EX^2 < \infty$ and $EX^4 = \infty$. Further assume that X^2 is in the domain of attraction of the stable law with an index $\tau \in [1, 2]$, and that $(d_n - c_n)^2 = o(c_n)$ when $\tau = 1$, where $c_n > 0$ (slowly varying) and d_n (diverge to ∞) are constants related to the limit law:

$$c_n\, n^{-1/\tau}\Big(\sum_{i=1}^{n} Y_i^2 - n\Big) \to_D Z_\tau, \quad \tau > 1, \quad as\ n \to \infty,$$

$$c_n\, n^{-1}\sum_{i=1}^{n} Y_i^2 - d_n \to_D Z_1, \quad as\ n \to \infty,$$

where $Y_i = (X_i - \mu)/\sigma$. Define

$$a_n = 2c_n n^{1/2 - 1/\tau}/\nu, \quad b_n = \begin{cases} \nu\sqrt{n} & if\ 1 < \tau \leq 2, \\ \nu\sqrt{n}\big[3/2 - d_n/(2c_n)\big] & if\ \tau = 1. \end{cases}$$

We have

(i) if $\tau = 2$, then $a_n(t_n - b_n) \to_D N(0, 1)$,

(ii) if $\tau \in [1, 2)$, then $a_n(t_n - b_n) \to_D -Z_\tau$.

In BJSZ, the authors also considered the limit behavior of the non-centralized Student t when $EX^2 = \infty$ or $\mu = \mu_n = o(n)$.

3. Absolute errors in the central limit theorems

There are mainly two approaches for estimating the error of the normal approximation in Section 2. One is to study the absolute error in the self-normalized central limit theorem via a Berry-Esseen bound or an Edgeworth expansion. Put $b_n = \sup\{\, x \,:\, n\,x^{-2}\,E\{X^2\,I(|X| \le x)\} \ge 1\}$ and

$$
\begin{aligned}
\delta_n &= n\,P(|X| > b_n) + n\,b_n^{-1}\,|E\{X\,I(|X| \le b_n)\}| \\
&\quad + n\,b_n^{-3}\,E\{|X|^3\,I(|X| \le b_n)\}.
\end{aligned} \tag{3.1}
$$

As a major advance in this direction, Bentkus and Götze (1996) [7] refined the results in Slavova (1985) [63] as well as Hall (1988) [34] and showed that

Theorem 3.1. *If X is in the domain of attraction of the normal law, then*

$$
\sup_x \big|P\big(S_n/V_n \le x\big) - \Phi(x)\big| \le A\,\delta_n,
$$

where A is an absolute constant.

This result was extended to the independent, but not identically distributed case by Bentkus, Bloznelis and Götze (1996) [6] and Shao (2005) [59]. Differing from Bentkus and Götze (1996) [7] as well as Bentkus, Bloznelis and Götze (1996) [6], Shao (2005) [59] used Stein's method to provide an explicit constant in his theorem, which is stated as follows.

Theorem 3.2. *Let $X_1, X_2, \ldots X_n$ be a sequence of independent random variables with mean $EX_j = 0$, $j = 1, 2, \ldots, n$. Then,*

$$
\begin{aligned}
\sup_x \big|P\big(S_n/V_n \le x\big) - \Phi(x)\big| &\le \frac{10.2}{B_n^2} \sum_{i=1}^{n} E\big\{X_i^2 I(|X_i| \ge B_n/2)\big\} \\
&\quad + \frac{25}{B_n^3} \sum_{i=1}^{n} E\big\{|X_i|^3 I(|X_i| \le B_n/2)\big\},
\end{aligned}
$$

where $B_n^2 = \sum_{i=1}^{n} EX_i^2$.

For explicit constants in the i.i.d case and other results related to the uniform Berry-bound, we refer to Egorov (2002) [27] and Novak (2004) [47].

The Berry-Esseen bounds provide an upper bound for the rate of convergence in the central limit theorem. In order to characterize the rate of convergence in the central limit theorem for S_n/V_n, Hall and Wang (2004) [35] investigated the leading term arguments under the optimal conditions. Letting

$$
L_n(x) = n\,E\Big(\Phi\Big[x\,\big\{1 + (X/b_n)^2\big\}^{1/2} - (X/b_n)\Big] - \Phi(x)\Big),
$$

[35] established the following result, among others.

Theorem 3.3. *If X is in the domain of attraction of the normal law, then*

$$\sup_x \left| P(S_n/V_n \leq x) - \{\Phi(x) + L_n(x)\} \right| = o(\delta_{1n}) + O\left(n^{-1/2}\right), \qquad (3.2)$$

where

$$\begin{aligned}
\delta_{1n} \quad &= \quad n\, P(|X| > b_n) + n\, b_n^{-1} \left| E\{X\, I(|X| \leq b_n)\} \right| \\
&\quad + n\, b_n^{-3} \left| E\{X^3\, I(|X| \leq b_n)\} \right| + n\, b_n^{-4}\, E\{X^4\, I(|X| \leq b_n)\}.
\end{aligned}$$

If in addition Cramér's condition holds, i.e.

$$\limsup_{|t| \to \infty} \left| E\left(e^{itX}\right) \right| < 1,$$

then $O(n^{-1/2})$ on the right-hand side of (3.2) can be replaced by $O(n^{-1})$.

Theorem 3.3 argues that $L_n(x)$ is a leading term in an expansion of the distribution of S_n/V_n. Indeed, it was proved in [35] that $\delta_{1n} \to 0$ and

$$\sup_x |L_n(x)| \asymp \delta_{1n}, \quad \text{as } n \to \infty. \qquad (3.3)$$

Here $a_n \asymp b_n$ denotes that $0 < \liminf_{n\to\infty} a_n/b_n \leq \limsup_{n\to\infty} a_n/b_n < \infty$. There exist examples of distributions in the domain of attraction of the normal law, having zero mean, and for which any given one of the four components in the definition of δ_{1n}, dominates all the others along a subsequence. It follows that none of the terms of which δ_{1n} is comprised can be dropped if we require a full account of the rate of convergence in the central limit theorem. Together, properties (3.2) and (3.3) give concise results about the rate of convergence in the central limit theorem. For example, assuming that X is in the domain of attraction of the normal law and $E(X) = 0$, we have

(a) $\sup_x |P(S_n/V_n \leq x) - \Phi(x)| + n^{-1/2} \asymp \delta_{1n} + n^{-1/2}$, where $n^{-1/2}$ can be replaced by n^{-1} if Cramér's condition is satisfied.
(b) $E|X|^{2(r+1)} < \infty$ if and only if

$$\sum_{n=1}^{\infty} n^{r-1} \sup_x \left| P(S_n/V_n \leq x) - \Phi(x) \right| < \infty, \quad \text{for } 0 \leq r < 1/2.$$

(c) $P(|X| > x) = O(x^{-2(r+1)})$ as $x \to \infty$, if and only if

$$\sup_x \left| P(S_n/V_n \leq x) - \Phi(x) \right| = O\left(n^{-r}\right), \quad \text{for } 0 < r < 1/2.$$

If additionally $E(|X|^3) < \infty$ and $E(X^2) = 1$, then we also have

$$\sup_x \left| n^{1/2} L_n(x) - \tfrac{1}{6}\gamma\left(2x^2 + 1\right)\phi(x) \right| \to 0 \qquad (3.4)$$

where $\gamma = E(X^3)$, as $n \to \infty$.

Formula (3.4) shows that in the case of finite third moments, the leading term is asymptotic to its conventional form in an Edgeworth expansion. Consequently, if $E|X|^3 < \infty$ and the distribution of X is nonlattice, then

$$\sup_x \left| P(S_n/V_n \le x) - \Phi(x) - F_n(x) \right| = o(n^{-1/2}),$$

where $F_n(x) = \frac{1}{6\sqrt{n}} \frac{EX^3}{\sigma^3}(2x^2 + 1)\phi(x)$. More results on Edgeworth expansion for Student's t statistics can be found in Hall (1987) and Bloznelis and Putter (1998, 2002) [9, 10].

Wang and Jing (1999) [69] was the first to investigate the non-uniform Berry-Esseen bound for S_n/V_n. The result given by [69] was extended in Robinson and Wang (2005) [54], where an exponential non-uniform bound was established under the optimal moment conditions. The following result comes from Theorem 3 of [54].

Theorem 3.4. *If X is in the domain of attraction of the normal law, then there exist $0 < \eta < 1$ such that*

$$\left| P\bigl(S_n/V_n \le x\bigr) - \Phi(x) \right| \le A\,\delta_n \exp\bigl(-\eta x^2/2\bigr) \tag{3.5}$$

for all $x \in R$ and $n \ge 1$, where δ_n is defined as in (3.1) and A is an absolute constant.

The constant η in Theorem 3.4 may depend on the distribution of X and cannot be replaced by an absolute constant. For example, let $X_1, \ldots, X_n$ be iid random variables from the distribution $P(X = 1) = 1 - P(X = -p/(1-p)) = p$, where $0 < p < 1$. It is readily seen that $EX = 0$, and for $x = \sqrt{n}$ and $p \ge 1/2$

$$P(S_n \ge xV_n) \ge P(S_n = n) \ge P(X_1 = 1, \ldots, X_n = 1) = p^n = e^{x^2 \log p}.$$

Since $\log p < 0$ and $\log p \uparrow 0$ as $p \uparrow 1$, (3.5) cannot be true for an absolute constant. Corollary 2.3 of [69] provided a similar result to (3.5) under $E|X|^{10/3} < \infty$. However the corollary misused the concept of absolute constant.

It is possible to replace the η in (3.5) by 1 if we restrict the x in a narrow range or if we require X_j to be symmetric random variables. Indeed it follows from Theorem 4.1 below, that, if $EX = 0$ and $E|X|^3 < \infty$, then

$$\left| P\bigl(S_n/V_n \le x\bigr) - \Phi(x) \right| \le A\,(1 + x^2)e^{-x^2/2}E|X|^3/(\sigma^3\sqrt{n}),$$

for $0 \le x \le \sigma n^{1/6}/(E|X|^3)^{1/3}$, where $\sigma^2 = EX^2$ and A is an absolute constant. Furthermore if X is a symmetric random variable around zero with $E|X|^3 < \infty$, then

$$\left| P\bigl(S_n/V_n \le x\bigr) - \Phi(x) \right| \le A\,\min\{(1 + x^2)E|X|^3/(\sigma^3\sqrt{n}), 1\}e^{-x^2/2}.$$

See Wang and Jing (1999) [69] as well as Chistyakov and Götze (2003) [11].

4. Relative errors in the central limit theorems

Section 3 reviewed the absolute error in the central limit theorem for S_n/V_n. This section considers the relative error of $P(S_n/V_n \geq x)$ to $1 - \Phi(x)$, that is, the Cramér-type moderate deviation for S_n/V_n, which is another approach for estimating the errors in the normal approximation in Section 2. In this regard, Jing, Shao and Wang (2003) [37] refined Shao (1999) [57], Wang and Jing (1999) [69] as well as Chistyakov and Götze (2003) [11], and obtained the following result. Let

$$B_n^2 = \sum_{i=1}^n EX_i^2 \quad \text{and} \quad \mathcal{L}_{kn} = B_n^{-k} \sum_{i=1}^n E|X_i|^k, \quad \text{for } k \geq 3.$$

Theorem 4.1. *If* $X_1, X_2, \ldots$ *are independent random variables with* $EX_j = 0$ *and* $0 < E|X_j|^3 < \infty$*, then*

$$P(S_n/V_n \geq x) = \left\{1 - \Phi(x)\right\}\left\{1 + O_1(1+x)^3 \mathcal{L}_{3n}\right\} \tag{4.1}$$

for $0 \leq x \leq \mathcal{L}_{3n}^{-1/3}$*, where* O_1 *is bounded by an absolute constant.*

[37] actually established more general frameworks and considered applications to the self-normalized law of the iterated logarithm and the studentized bootstrap. There are several further extensions in i.i.d. settings. Using an example, Chistyakov and Götze (2004a) [12] proved that the result in [37] is sharp. Robinson and Wang (2005) [54] established a Cramér type result under optimal moment condition, that is, under the assumption that X is in the domain of attraction of the normal law. Assuming $EX^4 < \infty$, Wang (2005) [66] as well as Wang and Hall (2009) [68] proved that $P(S_n/V_n \geq x)/\{1 - \Phi(x)\}$ equals, to first order, $\exp\{-x^3 EX^3/(3\sqrt{n}\sigma^3)\}$ where $\sigma^2 = EX^2$. Explicitly, the following Theorem 4.2 is from Theorem 1.1 of Wang (2005) [66].

Theorem 4.2. *Assume that* $EX = 0$ *and* $EX^4 < \infty$*. Then, for* $x \geq 0$ *and* $x = O(n^{1/6})$,

$$\frac{P(S_n \geq xV_n)}{1 - \Phi(x)} = \exp\left\{-\frac{x^3 EX^3}{3\sqrt{n}\sigma^3}\right\}\left[1 + O\left(\frac{1+x}{\sqrt{n}}\right)\right], \tag{4.2}$$

If in addition $EX^3 = 0$*, then, for* $0 \leq x \leq O(n^{1/6})$,

$$P(S_n \geq xV_n) - (1 - \Phi(x)) = O(n^{-1/2}e^{-x^2/2}). \tag{4.3}$$

We mention that the proofs of Theorems 4.1 and 4.2 (other related results in the cited articles as well) depends heavily on the following ideas and facts. First of all, by the Cauchy inequality $xV_n \leq (x^2 + b^2V_n^2)/(2b)$, where $b := b_x = x/B_n$, we have

$$P(S_n \geq xV_n) \geq P\left\{S_n \geq (x^2 + b^2V_n^2)/(2b)\right\} = P(2bS_n - b^2V_n^2 \geq x^2). \tag{4.4}$$

Since $2bS_n - b^2 V_n^2 = \sum_{i=1}^n Y_i$, where $Y_i = 2bX_i - b^2 Y_i^2$ are independent having $Ee^{tY_i} < \infty$ for any $t > 0$, the lower bound of (4.1) was obtained from the fact that:

$$P(2bS_n - b^2 V_n^2 \geq x^2) = \{1 - \Phi(x)\}\Big\{1 + O(1)(1 + x)^3 \mathcal{L}_{3n}\Big\} \tag{4.5}$$

for $0 \leq x \leq \mathcal{L}_{3n}^{-1/3}$. By the conjugate method, it is easy to prove (4.5) and it is also possible to provide a more concise result.

A truncation technique is used in establishing the upper bound of (4.1). Let $\tau := \tau_{n,x} = B_n / \max\{1, x\}$ and define $\bar{X}_i = X_i I(|X_i| \leq \tau)$. We may write

$$P(S_n \geq xV_n) = p_{1n} + p_{2n}, \tag{4.6}$$

where $p_{1n} = P(S_n \geq xV_n, X_i = \bar{X}_i, \text{ all } i = 1, \ldots, n)$ and

$$p_{2n} = P(S_n \geq xV_n, X_i \neq \bar{X}_i, \text{ some } i = 1, \ldots, n).$$

It is relatively easy to show that

$$p_{2n} \leq A(1 + x)^3 \mathcal{L}_{3n}\{1 - \Phi(x)\}, \tag{4.7}$$

for $0 \leq x \leq \mathcal{L}_{3n}^{-1/3}$. Therefore the key step in the proof of (4.1) is to establish the result:

$$p_{1n} \leq \{1 - \Phi(x)\}\Big\{1 + A(1 + x)^3 \mathcal{L}_{3n}\Big\} \tag{4.8}$$

for $0 \leq x \leq \mathcal{L}_{3n}^{-1/3}$. This was done in Jing, Shao and Wang (2003) [37] (and other related results in the cited articles) by separating the x into "small" and "large" cases. Explicitly in [37] we proved

$$p_{1n} \leq \{1 - \Phi(x)\}\Big\{1 + A(1 + x)^3 \mathcal{L}_{3n}\Big\} + e^{-3x^2} \tag{4.9}$$

and

$$p_{1n} \leq \{1 - \Phi(x)\}\Big\{1 + A(1 + x)^3 \mathcal{L}_{3n}\Big\} + A(x\mathcal{L}_{3n})^{4/3}, \tag{4.10}$$

for $0 \leq x \leq \mathcal{L}_{3n}^{-1/3}$. By using (4.9) we obtain (4.8) when x is "large", and by using (4.10) we obtain (4.8) when x is "small". The proofs of both (4.9) and (4.10) are difficult and very technical.

Observe that p_{1n} may be estimated by the following (since $(1 + y)^{1/2} \geq 1 + y/2 - y^2$ for $y \geq -1$)

$$\begin{aligned}
p_{1n} &= P\Big(S_n \geq xB_n\big[1 + \frac{1}{B_n^2}(V_n^2 - B_n^2)\big]^{1/2}, X_i = \bar{X}_i, \text{ all } i = 1, \ldots, n\Big) \\
&\leq P\Big(S_n \geq xB_n\big\{1 + \frac{1}{2B_n^2}(V_n^2 - B_n^2) - \frac{1}{B_n^4}(V_n^2 - B_n^2)^2\big\}, \\
&\qquad X_i = \bar{X}_i, \text{ all } i = 1, \ldots, n\Big) \\
&\leq P\Big(2bS_n - b^2 V_n^2 \geq x^2\big[1 - 2B_n^{-4}(\bar{V}_n^2 - B_n^2)^2\big]\Big), \tag{4.11}
\end{aligned}$$

where $\bar{V}_n = \sum_{i=1}^n \bar{X}_i^2$. This, together with (4.4), heuristically provides the fact that $P(S_n/V_n \geq x)$ is "close enough" to $P(2bS_n - b^2V_n^2 \geq x^2)$ under certain moment conditions, since $B_n^{-4}E(\bar{V}_n^2 - B_n^2)^2$ is small. As noticed before, it is relatively easy to derive a concise estimate for $P(2bS_n - b^2V_n^2 \geq x^2)$. Based on these observations, Wang (2011) [67] provided an alternative proof of (4.8). Furthermore he gave a more concise estimate of $P(S_n/V_n \geq x)$ instead of the (4.1), which is stated as follows.

Let's start with some notation and basic facts. Let $Y_i = 2bX_i - (bX_i)^2$ and write $m(\lambda) = \sum_{i=1}^n \log Ee^{\lambda Y_i}$. Note that $m(\lambda)$ is well-defined and $m''(\lambda) = \sum_{i=1}^n var(2bZ_i - b^2Z_j^2) > 0$ for all $\lambda > 0$ and $x \neq 0$, where $Z_1, \ldots, Z_n$ are independent random variables with Z_j having distribution function $V_j(u)$ defined by

$$V_j(u) = E\{e^{\lambda Y_j}I(Y_j \leq u)\}/Ee^{\lambda Y_j}. \tag{4.12}$$

Hence, for each $x \neq 0$, $m'(\lambda)$ is a strictly increasing function for $\lambda > 0$. Furthermore, there exists an absolute constant A_0 such that for all $x > 0$ satisfying $x \leq \frac{1}{3}B_n/(\max_i E|X_i|^3)^{1/3}$ and $x \leq \mathcal{L}_{3n}^{-1}/A_0$, the equation $m'(\lambda) = x^2$ has a unique solution $\lambda_0 > 0$. For this defined λ_0, [67] established the following result. Write

$$\Psi_x(t, \delta) = \exp\{m(t) + x^2(2t^2 - t) - t\delta\},$$
$$\widetilde{\Psi}_x(t, c) = \exp\{m(t) + \frac{1}{2}(x + c)^2 - t(x^2 + 2xc)\},$$

where $\delta = \delta(x)$ and $c = c(x)$ are defined later.

Theorem 4.3. *If $X_1, X_2, \ldots$ are independent random variables with $EX_j = 0$ and $0 < E|X_j|^4 < \infty$, then*

$$P(S_n/V_n \geq x)$$
$$= \Psi_x(\lambda_0, 0)\{1 - \Phi(2\lambda_0 x)\}\{1 + O_1(1 + x)\mathcal{L}_{3n} + O_2(1 + x^4)\mathcal{L}_{4n}\}, \tag{4.13}$$

or equivalently,

$$P(S_n/V_n \geq x)$$
$$= \widetilde{\Psi}_x(\lambda_0, 0)\{1 - \Phi(x)\}\{1 + O_1(1 + x)\mathcal{L}_{3n} + O_2(1 + x^4)\mathcal{L}_{4n}\}, \tag{4.14}$$

for $0 \leq x \leq \mathcal{L}_{4n}^{-1/4}$, where O_1 and O_2 are bounded by an absolute constant.

Since $\max_i E|X_i|^3 \leq (\sum_{i=1}^n E|X_i|^4)^{3/4}$ and $\mathcal{L}_{3n}^2 \leq \mathcal{L}_{4n}$, the condition $x \leq \mathcal{L}_{4n}^{-1/4}$ implies that $x \leq B_n/(\max_i E|X_i|^3)^{1/3}$ and $x \leq \mathcal{L}_{3n}^{-1}$, provided $\mathcal{L}_{3n} \leq 1$. Theorem 4.3 is a corollary of the following general framework. Write

$$\Delta_{n,x} = \tau^{-3}\sum_{i=1}^n E\{|X_i|I(|X| \geq \tau)\} + \tau^{-4}\sum_{i=1}^n E\{|X_i|^4I(|X_i| \leq \tau)\},$$

where $\tau = B_n/\max\{1, x\}$ and denote λ_1 for the solution of the equation $m'(\lambda) = x^2 + 2xc$.

Theorem 4.4. *If $X_1, X_2, \ldots$ are independent random variables with $EX_j = 0$ and $0 < E|X_j|^3 < \infty$, then*

$$P(S_n \geq xV_n + c\,B_n)$$
$$= \Psi_x(\lambda_1, \, 2xc)\left\{1 - \Phi(2\,\lambda_1\,x)\right\}e^{O_1\Delta_{n,x}}\left\{1 + O_2(1+x)\mathcal{L}_{3n}\right\}, \qquad (4.15)$$

or equivalently,

$$P(S_n \geq xV_n + c\,B_n)$$
$$= \widetilde{\Psi}_x(\lambda_1, \, c)\left\{1 - \Phi(x+c)\right\}e^{O_1\Delta_{n,x}}\left\{1 + O_2(1+x)\mathcal{L}_{3n}\right\}, \qquad (4.16)$$

uniformly for $|c| \leq x/5$ and for all $x > 0$ satisfying $x \leq \frac{1}{3}B_n/(\max_i E|X_i|^3)^{1/3}$ and $x \leq \mathcal{L}_{3n}^{-1}/A_0$, where O_1 and O_2 are bounded by an absolute constant.
Consequently, we also have

$$P(S_n \geq xV_n + c\,B_n)$$
$$= \left\{1 - \Phi(x+c)\right\} \Psi_x^* \, e^{O_1\Delta_{n,x}}\left\{1 + O_2(1+x)\mathcal{L}_{3n}\right\}, \qquad (4.17)$$

where $\gamma = \frac{1}{2}(1 + c/x)$ and $\Psi_x^* = \exp\left\{\gamma^2(4\gamma/3 - 2)x^3 B_n^{-3} \sum_{i=1}^n EX_i^3\right\}$, [1] *uniformly for $|c| \leq x/5$ and for all $x > 0$ satisfying $x \leq \frac{1}{3}B_n/(\max_i E|X_i|^3)^{1/3}$ and $x \leq \mathcal{L}_{3n}^{-1}/A_0$, where O_1 and O_2 are bounded by an absolute constant.*

Theorem 4.4 provides a concise estimate of $P(S_n/V_n \geq x)$ instead of (4.1). It also improves the general result in Jing, Shao and Wang (2003) [37]. Assuming that $X_1, X_2, \ldots, X_n$ are i.i.d. random variables, result (4.17) gives Theorem 4.2 and Theorem 1 of Wang and Hall (2009) [68]. It would be interesting to see whether or not it is possible to remove the error term $O_1\Delta_{n,x}$ in (4.16). If the x is in a small range [for instance $0 \leq x \leq O(n^{1/6})$ in the i.i.d. settings], then the term $O_1\Delta_{n,x}$ is smaller than that of $x\mathcal{L}_{3n}$, and hence can be removed (see Theorem 4.3). However it plays a part when the x is in a median or large range [for instance $O(n^{1/6}) \leq x \leq o(\sqrt{n})$ in the i.i.d. settings]. In terms of the difference between (4.4) and (4.11), the current truncation technique seems to raise an error term like $O_1\Delta_{n,x}$.

Based on these facts, in order to obtain a better estimate of $P(S_n/V_n \geq x)$ in a median or large range for the x, we may have to use a completely different technique. Jing, Shao and Zhou (2004) [38] [also see Zhou and Jing (2006) [71]] investigated a saddle-point approximation for the tail probability $P(S_n/V_n \geq x)$ in a very large range for the x, that is, $x = c\sqrt{n}$ with $0 < c < 1$ in the i.i.d. settings (see the review in the next section). It is not clear at the moment if the technique in [38] or [71] can be employed to provide a better approximation for $P(S_n/V_n \geq x)$ in a median range for the x [i.e., $O(n^{1/6}) \leq x \leq o(\sqrt{n})$ in the i.i.d. settings]. We note that [38] derived their results without imposing any moment conditions on X, but we do require a moment condition to establish a

---

[1]There is a typo in the definition of Ψ_x^*, presented in (1.17) of Wang (2011) [67].

better approximation for $P(S_n/V_n \geq x)$ in a small range for the x. For instance, a finite third moment is necessary to establish

$$P(S_n/V_n \geq x) = \{1 - \Phi(x)\}\{1 + O(1 + x)\mathcal{L}_{3n}\},$$

for $0 \leq x \leq C$, where C is a constant. By taking consideration of this fact, some significant modifications might be necessary even if the technique in [38] would work to provide a better approximation for $P(S_n/V_n \geq x)$ in a median range for the x. On the other hand, the problem may be solvable by the method developed in Shao and Zhou (2012) [61] where a new randomized concentration inequality is obtained to establish a Cramér type moderate theorem for self-normalized non-linear statistics (see Section 6).

5. Cramér-Chernoff type large deviation and saddle-point approximation

Section 4 considered the relative error of $P(S_n/V_n \geq x)$ to $1 - \Phi(x)$ when $0 \leq x \leq o(n^{1/2})$. In the i.i.d. settings, assuming $EX = 0$ and $E|X|^3 < \infty$, it follows from Theorem 4.4 that

$$\log P(S_n/V_n \geq x_n) \quad \sim \quad -x_n^2/2, \tag{5.1}$$

whenever $x_n \to \infty$ and $x_n = o(n^{1/2})$. It is interesting to notice that the result (5.1) may be proved directly under fewer moment conditions. See Shao (1997) [55] and Jing, Shao and Zhou (2008) [38]. Indeed, Shao (1997) [55] showed that if $EX = 0$ and the distribution of X is in the domain of attraction of the normal law, then (5.1) holds true. Furthermore, [38] established the following more general Theorem 5.1.

Denote the support of X by C_s, that is,

$$C_s = \{x : P\{X \in (x - \epsilon, x + \epsilon)\} > 0, \quad \text{for any } \epsilon > 0\}.$$

Say X is in the centered Feller class if $X \in \mathcal{F}_\theta$ for some $0 \leq \theta < \infty$, where

$$\mathcal{F}_\theta = \Big\{ X : \lim_{a \to \infty} \frac{a^2\{P(|X| > a) + a^{-1}|EXI(|X| \leq a)|\}}{EX^2 I(|X| \leq a)} = \theta \Big\}.$$

Then we have

Theorem 5.1. *Suppose that*

(i) $C_s \cap R^+ \neq \phi$ *and* $C_s \cap R^- \neq \phi$*, where* $R^+ = \{x : x > 0\}$ *and* $R^- = \{x : x < 0\}$*;*
(ii) $EX = 0$ *or* $EX^2 = \infty$*.*

Also assume that X is in the centered Feller class. Then, for any sequence $\{x_n, n \geq 1\}$ satisfying $x_n \to \infty$ and $x_n = o(n^{1/2})$,

$$\log P(S_n/V_n \geq x_n) \quad \sim \quad -n\,\lambda(x_n^2/n), \tag{5.2}$$

where $\lambda(x) = \inf_{b>0} \sup_{t \geq 0}(tx - \log E \exp\{t(2bX - b^2 X^2)\})$*.*

The result (5.2) also holds true when $x_n = \epsilon\sqrt{n}$ for some $\epsilon > 0$. In fact, in this situation, the condition that X is in the centered Feller class is not necessary. The following theorem proved in [55] claims this statement.

Theorem 5.2. *Assume that either $EX = 0$ or $EX^2 = \infty$. Then,*

$$\log P(S_n/V_n \geq \epsilon\sqrt{n}) \quad \sim \quad -n\,\lambda(\epsilon^2), \tag{5.3}$$

for $\epsilon > EX/(EX^2)^{1/2}$, where $\lambda(x)$ is defined as in Theorem 5.1 and $EX/(EX^2)^{1/2}$ is interpreted to be zero if $EX^2 = \infty$ and $0/0$ to be ∞.

Note that no conditions are required for the result (5.3), since it is natural to assume $EX = 0$ if $EX^2 < \infty$. By adding some smoothing conditions for the distribution of X, [38] derived the saddle-point approximation for the tail probability $P(S_n/V_n \geq \epsilon\sqrt{n})$ without any moment conditions. Indeed, it is proved in [38] that, if the distribution function $F(x)$ of X is continuous, then, for any $0 < \epsilon < 1$, the equation:

$$s + \frac{2ta}{\epsilon^2} = 0, \quad \frac{\partial K(s,t)}{\partial s} = a, \quad \frac{\partial K(s,t)}{\partial t} = a^2/\epsilon^2$$

has a unique solution (s_0, t_0, a_0) such that $s_0 > 0, t_0 < 0$ and $a_0 > 0$, where $K(s,t) = \log Ee^{sX+tY}$. Write

$$\Lambda_0(\epsilon) = s_0\,a_0 + t_0\,a_0^2/\epsilon^2 - K(s_0, t_0),$$
$$\Lambda_1(\epsilon) = 2t_0/\epsilon^2 + \left(1, 2a_0/\epsilon^2\right)\Delta^{-1}\left(1, 2a_0/\epsilon^2\right)',$$

where $K_{ss}(s,t) = \frac{\partial^2 K(s,t)}{\partial s^2}$, $K_{st}(st) = \frac{\partial^2 K(s,t)}{\partial s\,\partial t}$, $K_{tt}(s,t) = \frac{\partial^2 K(s,t)}{\partial t^2}$ and

$$\Delta = \left(\begin{array}{cc} K_{ss}(s_0,t_0) & K_{st}(s_0,t_0) \\ K_{st}(s_0,t_0) & K_{tt}(s_0,t_0) \end{array}\right).$$

Jing, Shao and Zhou (2004) [38] established

Theorem 5.3. *Assume that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|Ee^{isX+itX^2}|^r\,ds\,dt < \infty$ for some $r > 1$. Then, for any $0 < \epsilon < 1$,*

$$P(S_n/V_n \geq \epsilon\sqrt{n}) = 1 - \Phi(\sqrt{n}\omega) - \frac{\phi(\sqrt{n}\omega)}{\sqrt{n}}\left(\frac{1}{\omega} - \frac{1}{\nu} + O(n^{-1})\right), \tag{5.4}$$

where $\omega = \sqrt{2\Lambda_0(\epsilon)}$ and $\nu = -t_0\,(\det\Delta)^{1/2}\Lambda_1(\epsilon)^{1/2}$.

The smoothing condition in Theorem 5.3 was reduced to (X, X^2) being a strongly nonlattice vector in Zhou and Jing (2006) [71]. As mentioned in Section 4, it would be interesting to consider the extension of (5.4) to the situation where $0 < \epsilon := \epsilon_n \to 0$, as in Theorems 4.3 and 4.4.

6. Other self-normalized limit theorems

Sections 2–5 reviewed current developments in the investigations for the self-normalized sums S_n/V_n, along the lines related to central limit theorems. This section will briefly mention other important self-normalized limit theorems.

6.1. Self-normalized laws of the iterated logarithm

Griffin and Kuelbs (1989) [31] was the first to investigate the laws of the iterated logarithm for the self-normalized sums. The following beautiful result is from their Theorem 1.

Theorem 6.1. *If $EX = 0$ and X is in the domain of attraction of the normal law, then*

$$\limsup_{n \to \infty} \frac{S_n}{V_n (\log \log n)^{1/2}} = \sqrt{2}, \quad a.s. \tag{6.1}$$

Griffin and Kuelbs (1989) [31] and later Shao (1997) [55] also discussed the laws of the iterated logarithm under the condition that X is in the domain of attraction of the stable law. More currently, Jing, Shao and Zhou (2008) [39] established a general result under the condition that X is in the centered Feller class defined as in Section 5. The following Theorem 6.2 is from their Theorem 1.2. Note that Theorem 6.2 also extends those results given by Giné and Mason (1998) [30].

Theorem 6.2. *Under the conditions of Theorem 5.1,*

$$\limsup_{n \to \infty} \frac{S_n}{V_n (\log \log n)^{1/2}} = \frac{1}{\sqrt{t_0}}, \quad a.s. \tag{6.2}$$

where $t_0 = \lim_{x \to 0^+} t_x$, and (t_x, b_x) is the solution of the following equations:

$$
\begin{aligned}
E\big[b(2X - bX^2) \exp\{tb(2X - bX^2)\}\big] &= x\, E \exp\{tb(2X - bX^2)\}, \\
E\big[(X - bX^2) \exp\{tb(2X - bX^2)\}\big] &= 0.
\end{aligned}
$$

There are other extensions for the self-normalized laws of the iterated logarithm. For instance, Csörgö and Hu (2013) [16] established a strong approximation result, Csörgö, Hu and Mei (2013) [17] obtained a strassen-type law of the iterated logarithm, Dembo and Shao (1998, 2006) [23, 24] considered self-normalized laws of the iterated logarithm under space R^d, de la Pena, Klass and Lai (2000, 2004) [50, 51] investigated the laws of the iterated logarithm for self-normalized martingales. The later also derived other results for self-normalized processes.

6.2. Darling-Erdós type theorem and maximum of self-normalized sum

CsSzW (2003a) [18] and later Wang (2004) [65] investigated the asymptotic behavior in the distribution of the maximum of self-normalized sums, $\max_{1 \le k \le n} S_k / V_k$. The following Darling-Erdős type result Theorem 6.3 comes from Wang's Theorem 1. Write $l(x) = EX^2 I_{(|X| \le x)}$, $a(n) = (2 \log \log n)^{1/2}$ and

$$b(n) = 2 \log \log n + \frac{1}{2} \log \log \log n - \frac{1}{2} \log(4\pi).$$

Theorem 6.3. *Suppose that $l(x)$ is a slowly varying function at ∞, satisfying $l(x) \leq c_1 \exp\{c_2(\log x)^\beta\}$ for some $c_1 > 0, c_2 > 0$ and $0 \leq \beta < 1/2$. Then, for every $t \in R$, we have that*

$$\lim_{n \to \infty} P\Big\{a(n) \max_{1 \leq k \leq n} S_k/V_k \leq t + b(n)\Big\} = \exp(-e^{-t}). \tag{6.3}$$

Self-normalization significantly reduces the moment conditions in comparison to the classical result. Indeed, the classical Darling-Erdős theorem shows that,

$$\lim_{n \to \infty} P\Big\{a(n) \max_{1 \leq k \leq n} S_k/(\sigma\sqrt{k}) \leq t + b(n)\Big\} = \exp(-e^{-t}), \tag{6.4}$$

where $\sigma^2 = EX^2$, if and only if $EX = 0$ and

$$EX^2 I_{(|X| \geq x)} = o\Big((\log\log x)^{-1}\Big), \qquad as \ \ x \to \infty.$$

See Einmahl (1989) [28]. The necessary condition for the result (6.3) remains an open question.

The asymptotic behavior of $\max_{1 \leq k \leq n} S_k/V_n$ is different from $\max_{1 \leq k \leq n} S_k/V_k$. This claim can be justified by the following result, coming from Theorem 1 of Liu, Shao and Wang (2012) [43].

Theorem 6.4. *If $E|X|^3 < \infty$, then*

$$\lim_{n \to \infty} \frac{P\big(\max_{1 \leq k \leq n} S_k \geq x \, V_n\big)}{1 - \Phi(x)} = 2 \tag{6.5}$$

uniformly for $x \in [0, \ o(n^{1/6}))$.

[43] actually established more general results which improved those by Hu, Shao and Wang (2009) [36]. It should be mentioned that Theorem 6.4 is comparable to the large deviation result for the maximum of partial sum given in Aleshkyavichene (1979) [1]. However the latter requires a finite exponential moment condition. If we are only interested in a Chernoff type large deviation, the third moment condition required in Theorem 6.4 can be reduced significantly. Indeed, [36] proved the following theorem.

Theorem 6.5. *If X is in the domain of attraction of the normal law, then*

$$\lim_{n \to \infty} x_n^{-2} \log P\big(\max_{1 \leq k \leq n} S_k \geq x_n \, V_n\big) = -\frac{1}{2} \tag{6.6}$$

for any $x_n \to \infty$ with $x_n = o(\sqrt{n})$.

6.3. Limit theorems for Hotelling's T^2 statistic

Let $\boldsymbol{X}, \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be a sequence of independent and identically distributed (i.i.d.) nondegenerate R^d-valued random vectors with mean $\boldsymbol{\mu}$, where $d \geq 1$. Let

$$SS_n = \sum_{i=1}^n \boldsymbol{X}_i, \ \boldsymbol{V}_n = \sum_{i=1}^n (\boldsymbol{X}_i - SS_n/n)(\boldsymbol{X}_i - SS_n/n)'$$

Define Hotelling's T^2 statistic by

$$T_n^2 = (SS_n - n\boldsymbol{\mu})' \boldsymbol{V}_n^{-1}(SS_n - n\boldsymbol{\mu}). \tag{6.7}$$

Here when $\boldsymbol{V}_n$ is not full rank, i.e., $\boldsymbol{V}_n$ is degenerate, $\boldsymbol{x}' \boldsymbol{V}_n^{-1} \boldsymbol{x}$ is defined as

$$\boldsymbol{x}' \boldsymbol{V}_n^{-1} \boldsymbol{x} = \sup_{||\boldsymbol{\theta}||=1, \boldsymbol{\theta}'\boldsymbol{x}>0} \frac{(\boldsymbol{\theta}'\boldsymbol{x})^2}{\boldsymbol{\theta}'\boldsymbol{V}_n\boldsymbol{\theta}}$$

and $0/0$ is interpreted as ∞.

The T^2-statistic is a natural extension of the Student's t at the vector space, and is used for testing hypotheses about mean $\boldsymbol{\mu}$ and for obtaining confidence regions for unknown $\boldsymbol{\mu}$. When $\boldsymbol{X}$ has a normal distribution $N(\boldsymbol{\mu}, \Sigma)$, it is known that $(n-d)T_n^2/(dn)$ is distributed as an F-distribution with d and $n-d$ degrees of freedom (see, e.g., Anderson (2003) [3]). When the distribution of $\boldsymbol{X}$ is not normal, it was proved by Sepanski (1994) [64] that the limiting distribution of T_n^2 as $n \to \infty$ is a χ^2-distribution with d degrees of freedom. Dembo and Shao (2006) [24] established the following large and moderate deviation results for the T^2-statistic.

Theorem 6.6.
(i) Assume that $\boldsymbol{\mu} = 0$. For $\alpha \in (0,1)$, let

$$K(\alpha) = \sup_{b \geq 0} \sup_{||\boldsymbol{\theta}||=1} \inf_{t \geq 0} E \exp\left(t\big(b\boldsymbol{\theta}'\boldsymbol{X} - \alpha((\boldsymbol{\theta}'\boldsymbol{X})^2 + b^2)/2\big)\right).$$

Then, for all $x > 0$,

$$\lim_{n \to \infty} P\left(T_n^2 \geq x\, n\right)^{1/n} = K(\sqrt{x/(1+x)}).$$

(ii) Let $\{x_n,\ n \geq 1\}$ be a sequence of positive numbers with $x_n \to \infty$ and $x_n = o(n)$ as $n \to \infty$. Assume that $h(x) := E||\boldsymbol{X}||^2 1\{||\boldsymbol{X}|| \leq x\}$ is slowly varying and

$$\liminf_{x \to \infty} \inf_{\boldsymbol{\theta} \in R^d, ||\boldsymbol{\theta}||=1} E(\boldsymbol{\theta}'\boldsymbol{X})^2 1\{||\boldsymbol{X}|| \leq x\}/h(x) > 0. \tag{6.8}$$

If $\boldsymbol{\mu} = 0$, then

$$\lim_{n \to \infty} x_n^{-1} \ln P\left(T_n^2 \geq x_n\right) = -\frac{1}{2}.$$

(iii) Assume that $h(x) := E||\boldsymbol{X}||^2 1\{||\boldsymbol{X}|| \leq x\}$ is slowly varying and 6.8 is satisfied. If $\boldsymbol{\mu} = 0$, then

$$\limsup_{n \to \infty} \frac{T_n^2}{2 \log \log n} = 1 \quad a.s.$$

Theorem 6.6 was further refined by Liu and Shao (2012) [42] as follows.

Theorem 6.7. *Suppose that $E||\boldsymbol{X}_1 - \boldsymbol{\mu}||^{3+\delta} < \infty$ for some $\delta > 0$. Then*

$$\frac{P\big(T_n^2 \geq x^2\big)}{P\big(\chi^2(d) \geq x^2\big)} \to 1 \qquad (6.9)$$

uniformly for $x \in [0, o(n^{1/6}))$.

We also refer to [23, 24] and [42] for other limit theorems related to the T^2-statistic. Because of its usefulness in statistical inferences, it would be interesting to find further sharp results for the T^2-statistic as in Sections 2–5. For instance, we conjecture that the result (6.9) still holds if only $E||\boldsymbol{X}_1 - \boldsymbol{\mu}||^3 < \infty$.

6.4. Limit theorems for studentized non-linear statistics

Non-linear statistics are used in various statistical inference problems. It is known that many of them can be written as a partial sum of independent random variables plus a negligible term. Typical examples include U-statistics, multi-sample U-statistics, L-statistics, random sums and functions of non-linear statistics. Since the Standardized non-linear statistics often involve some unknown nuisance parameters, the Studentized analogues are commonly used in practice.

Let $\xi_1, \ldots, \xi_n$ be independent random variables satisfying $E\xi_i = 0$. Assume the non-linear statistic of interest can be decomposed as a standardized partial sum of $\{\xi_i\}$, e.g., W_n, plus a remainder, say, D_1. Then the Studentized analogues can be written as

$$T_n = \frac{W_n + D_1}{V_n(1 + D_2)^{1/2}}, \qquad (6.10)$$

where

$$W_n = \sum_{i=1}^n \xi_i, , \quad V_n = \Big(\sum_{i=1}^n \xi_i^2\Big)^{1/2},$$

D_1 and D_2 are measurable functions of $\{\xi_i\}, 1 \leq i \leq n$. Examples satisfying (6.10) include the t-statistic, Studentized U-statistics and L-statistics. There are many works on the asymptotic distribution theory for the Studentized non-linear statistics T_n. We only refer to Wang, Jing and Zhao (2000) [70] as well as Shao, Zhang and Zhou (2012) [62] for general Berry-Esseen bounds under minor moment conditions.

A general Cramér type moderation for the Studentized statistics T_n was established in Shao and Zhou (2012) [61]. For $1 \leq i \leq n$ and $x \geq 0$, let

$$\delta_{i,x} = (1 + x)^2 E\xi_i^2 I_{\{(1+x)|\xi_i|>1\}} + (1 + x)^3 E|\xi_i|^3 I_{\{(1+x)|\xi_i|\leq 1\}}$$

and

$$\Delta_{n,x} = \sum_{i=1}^n \delta_{i,x}, \qquad I_{n,0} = Ee^{xW_n - x^2 V_n^2/2} = \prod_{i=1}^n Ee^{x\xi_i - x^2 \xi_i^2/2}.$$

Let $D_1^{(i)}$ and $D_2^{(i)}$, for each $1 \le i \le n$, be arbitrary measurable functions of $\{\xi_j, 1 \le j \le n, j \ne i\}$. Set for $x > 0$,

$$
\begin{aligned}
R_{n,x} = & \ I_{n,0}^{-1} \times \Big\{ E\big[(x|D_1| + x^2|D_2|)e^{\sum_{j=1}^{n}(x\xi_j - x^2\xi_j^2/2)}\big] \\
& + \sum_{i=1}^{n} E\big[\min(x|\xi_i|, 1)(|D_1 - D_1^{(i)}| + x|D_2 - D_2^{(i)}|)e^{\sum_{j \ne i}^{n}(x\xi_j - x^2\xi_j^2/2)}\big] \Big\}.
\end{aligned}
$$

Theorem 6.8. *Let T_n be defined in (6.10) and $\sum_{i=1}^{n} E\xi_i^2 = 1$. Then there is an absolute constant $A \ (> 1)$ such that*

$$
e^{O(1)\Delta_{n,x}}\big(1 - A\,R_{n,x}\big) \le \frac{P(T_n \ge x)}{1 - \Phi(x)} \tag{6.11}
$$

and

$$
\begin{aligned}
P(T_n \ge x) \ \le \ & \big(1 - \Phi(x)\big)e^{O(1)\Delta_{n,x}}\big(1 + A\,R_{n,x}\big) \\
& + P\big(|D_1|/V_n > 1/(4x)\big) + P\big(|D_2| > 1/(4x^2)\big) \tag{6.12}
\end{aligned}
$$

for all $x > 1$ satisfying $\max_{1 \le i \le n} \delta_{i,x} \le 1$ and $\Delta_{n,x} \le (1 + x)^2/A$, where $|O(1)| \le A$.

As a direct but nontrivial consequence of Theorem 6.8, [61] provided a sharp Cramér moderate deviation for Studentized U-statistics under optimal moment conditions, which improved an earlier result by Lai, Shao and Wang (2011) [67, 41].

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables and let $h : \mathcal{R}^m \to \mathcal{R}$ be a Borel measurable symmetric function of m variables, where $2 \le m < n/2$. Consider Hoeffding's U-statistic with kernel h

$$
U_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}).
$$

Let $\theta = Eh(X_1, \ldots, X_m)$, $g(x) = Eh(x, X_2, \ldots, X_m)$ and $\sigma^2 = var(g(X_1))$. The Studentized U-statistic [see e.g., Arvesen (1969) [4]] is defined as

$$
T_n = \frac{\sqrt{n}}{m s_1}(U_n - \theta),
$$

where s_1^2 is the Jacknife estimator of σ^2,

$$
\begin{aligned}
s_1^2 & = \frac{(n-1)}{(n-m)^2} \sum_{i=1}^{n}(q_i - U_n)^2, \\
q_i & = \binom{n-1}{m-1}^{-1} \sum_{(l_1, \ldots, l_m) \in \mathcal{C}_{m-1,i}} h(X_i, X_{l_1}, \ldots, X_{l_{m-1}})
\end{aligned}
$$

and

$$
\mathcal{C}_{m-1,i} = \{(l_1, \ldots, l_{m-1}) : 1 \le l_1 < \cdots < l_{m-1} \le n, l_j \ne i \text{ for } 1 \le j \le m-1\}.
$$

Theorem 6.9. *Let* $2 < p \leq 3$ *and assume* $0 < \sigma_p = (E|g(X_1) - \theta|^p)^{1/p} < \infty$. *Suppose that there are constants* $c_0 \geq 1$ *and* $\tau \geq 0$ *such that*

$$(h(x_1, \ldots, x_m) - \theta)^2 \leq c_0 \Big(\tau \sigma^2 + \sum_{i=1}^{m} (g(x_i) - \theta)^2 \Big). \tag{6.13}$$

Then there exists a constant $A > 1$ *only depending on* m *such that*

$$\frac{P(T_n \geq x)}{1 - \Phi(x)} = 1 + O(1) \Big\{ \frac{\sigma_p^p (1+x)^p}{\sigma^p n^{(p-2)/2}} + c_0 (1+\tau) \frac{(1+x)^3}{n^{1/2}} \Big\}, \tag{6.14}$$

for any $0 < x < \min\{\sigma n^{(p-2)/(2p)}/\sigma_p, n^{1/6}/(c_0(1+\tau))^{1/6}\}/A$, *where* $|O(1)| \leq A$. *In particular, we have*

$$\frac{P(T_n \geq x)}{1 - \Phi(x)} \to 1 \tag{6.15}$$

uniformly in $0 \leq x \leq o(n^{(p-2)/(2p)})$.

References

[1] ALESHKYAVICHENE, A. K. (1979). Probabilities of large deviations for the maximum of sums of independent random variables. *Theory Probab. Appl.*, **24**, 16–33. MR0522234

[2] ALESHKYAVICHENE, A. K. (1979). Probabilities of large deviations for the maximum of sums of independent random variables. II. *Theory Probab. Appl.*, **24**, 318–331. MR0532445

[3] ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis.* Third edition. Wiley-Interscience. MR1990662

[4] ARVESEN, J. N. (1969). Jackknifing U-statistics. *Ann. Math. Statist.*, **40**, 2076–2100. MR0264805

[5] BALAN, R. and KULIK, R. (2009). Weak invariance principle for mixing sequences in the domain of attraction of normal law. *Studia Sci. Math. Hungar.*, **46**, 329–343. MR2657020

[6] BENTKUS, V., BLOZNELIS, M. and GÖTZE, F. (1996). A Berry-Esséen bound for Student's statistic in the non-i.i.d. case. *J. Theoret. Probab.*, **9**, 765–796. MR1400598

[7] BENTKUS, V. and GÖTZE, F. (1996). The Berry-Esséen bound for Student's statistic. *Ann. Probab.*, **24**, 491–503. MR1387647

[8] BENTKUS, V., JING, B.-Y., SHAO, Q.-M. and ZHOU, W. (2007). Limit distributions of non-central t-statistic and their applications to the power of t-test under non-normality. *Bernoulli*, **13**, 346–364. MR2331255

[9] BLOZNELIS, M. and PUTTER, H. (1998). One term Edgeworth expansion for Student's t statistic. *Probability Theory and Mathematical Statistics: Proceedings of the Seventh Vilnius Conference* (B. Grigelionis et al., eds.), Vilnius, Utrecht: VSP/TEV, 1999, 81–98.

[10] Bloznelis, M. and Putter, H. (2002). Second order and Bootstrap approximation to Student's statistic. *Theory Probab. Appl.*, **47**, 374–381. MR2003205

[11] Chistyakov, G. P. and Götze, F. (2003). Moderate deviations for Student's statistic. *Theory Probab. Appl.*, **47**, 415–428. MR1975426

[12] Chistyakov, G. P. and Götze, F. (2004a). On bounds for moderate deviations for Student's statistic. *Theory Probab. Appl.*, **48**, 528–535. MR2141355

[13] Chistyakov, G. P. and Götze, F. (2004b). Limit distributions of Studentized means. *Ann. Probab.*, **32**, no. 1A, 28–77. MR2040775

[14] Choi, Y.-K. and Moon, H.-J. (2010). Self-normalized weak limit theorems for a ϕ-mixing sequence. *Bull. Korean Math. Soc.*, **47**, 1139–1153. MR2560606

[15] Csörgő, M. and Horváth, L. (1988). Asymptotic representations of self-normalized sums. *Probability and Mathematical Statistics*, **9**, no. 1, 15–24. MR0945676

[16] Csörgő, M. and Hu, Z. (2013). A strong approximation of self-normalized sums. *Sci. China Math.*, **56**, 149–160. MR3016589

[17] Csörgő, M., Hu, Z. and Mei, H. (2013). Strassen-type law of the iterated logarithm for self-normalized sums. *J. Theoret. Probab.*, **26**, 311–328. MR3055805

[18] Csörgő, M., Szyszkowicz, B. and Wang, Q. (2003a). Darling-Erdős theorems for self-normalized sums. *Ann. Probab.*, **31**, 676–692. MR1964945

[19] Csörgő, M., Szyszkowicz, B. and Wang, Q. (2003b). Donsker's theorem for self-normalized partial sums processes. *Ann. Probab.*, **31**, 1228–1240. MR1988470

[20] Csörgő, M., Szyszkowicz, B. and Wang, Q. (2004). On weighted approxiamtions and strong limit theorems for self-normalized partial sums processes. In *Asymptotics Methods in Stochastics: Festschrift for M. Csörgö* (L. Horváth and B. Szyszkowicz, eds.), Fields institute communications (Vol. 44), 489–521. MR2108953

[21] Csörgő, M., Szyszkowicz, B. and Wang, Q. (2008). On weighted approximations in $D[0;1]$ with applications to self-normalized partial sum processes. *Acta Math. Hungar.*, **121**, 333–357. MR2461439

[22] de la Peña, V., Lai, T. L. and Shao, Q. M. (2009). *Self-normalized Processes: Theory and Statistical Applications*. Springer, New York. MR2488094

[23] Dembo, A. and Shao, Q.-M. (1998). Self-normalized moderate deviations and LILs. *Stochastic Process. Appl.*, **75**, 51–65. MR1629018

[24] Dembo, A. and Shao, Q.-M. (2006). Large and moderate deviations for Hotelling's t^2-statistic. *Elect. Comm. in Probab.*, **11**, 149–159. MR2240708

[25] Efron, B. (1969). Student's t-test under symmetry conditions. *J. Amer. Statist. Assoc.*, 1278–1302. MR0251826

[26] Egorov, V. A. (1996). On the asymptotic behavior of self-normalized sums of random variables, *Teor. Veroyatnost. i Primenen.*, **41**, 643–650. MR1450081

[27] EGOROV, V. A. (2002). An estimate for the tail of the distribution of normalized and self-normalized sums. *J. Math. Sci. (N. Y.)*, **127**, 1717–1722. MR1976748

[28] EINMAHL, U. (1989). The Darling-Erdős theorem for sums of i.i.d. random variables, *Probab. Th. Rel. Fields*, **82**, 241–257. MR0998933

[29] GINÉ, E., GÖTZE, F. and MASON, D. (1997). When is the Student *t*-statistic asymptotically standard normal? *Ann. Probab.*, **25**, 1514–1531. MR1457629

[30] GINÉ, E. and MASON, D. (1998). On the LIL for self-normalized sums of iid random variables. *J. Theoret. Probab.*, **18**, 351–370. MR1622575

[31] GRIFFIN, P. S. and KUELBS, J. D. (1989). Self-normalized laws of the iterated logarithm. *Ann. Probab.*, **17**, 1571–1601. MR1048947

[32] GRIFFIN, P. S. and MASON, D. M. (1991). On the asymptotic normality of self-normalized sums. *Math. Proc. Cambridge Philos. Soc.*, **109**, 597–610. MR1094756

[33] HALL, P. (1987). Edgeworth expansion for Student's *t* statistic under minimal moment conditions. *Ann. Probab.*, **15**, 920–931. MR0893906

[34] HALL, P. (1988). On the effect of random norming on the rate of convergence in the central limit theorem. *Ann. Probab.*, **16**, 1265–1280. MR0942767

[35] HALL, P. and WANG, Q. (2004). Exact convergence rate and leading term in central limit theorem for student's *t* statistic. *Ann. Probab.*, **32**, 1419–1437. MR2060303

[36] HU, Z., SHAO, Q.-M. and WANG, Q. (2009). Cramér type large deviations for the maximum of self-normalized sums. *Electronic Jour. of Probab.*, **14**, 1181–1197. MR2511281

[37] JING, B.-Y., SHAO, Q.-M. and WANG, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann. Probab.*, **31**, 2167–2215. MR2016616

[38] JING, B.-Y., SHAO, Q.-M. and ZHOU, W. (2004). Saddlepoint approximation for Student's *t*-statistic with no moment conditions. *Ann. Statist.*, **6**, 2679–2711. MR2153999

[39] JING, B.-Y., SHAO, Q.-M. and ZHOU, W. (2008). Towards a universal self-normalized moderate deviation. *Trans. Amec. Math. Soci.*, **360**, 4263–4285. MR2395172

[40] KULIK, R. (2006). Limit theorems for self-normalized linear processes. *Statist. Probab. Lett.*, **76**, 1947–1953. MR2329238

[41] LAI, T. L., SHAO, Q.-M., and WANG, Q. (2011). Cramér type moderate deviations for Studentized U-statistics. *ESAIM – Probability and Statistics*, **15**, 168–179. MR2870510

[42] LIU, W. and SHAO, Q.-M. (2012). Cramér moderate deviation for Hotelling's T^2 statistic with applications to global tests. *Preprint*.

[43] LIU, W., SHAO, Q.-M. and WANG, Q. (2012). Self-normalized Cramér type moderate deviations for the maximum of sums. *Bernoulli*, forthcoming. MR3079304

[44] Logan, B. F., Mallows, C. L., Rice, S. O. and Sheep, L. A. (1973). Limit distributions of self-normalized sums. *Ann. Probab.*, **1**, 788–809. MR0362449

[45] Maller, R. A. (1981). A theorem on products of random variables with application to regression. *Australian Journal of Statistics*, **23**, 177–185. MR0636133

[46] Mason, D. M. (2005). The asymptotic distribution of self-normalized triangular arrys. *Journal of Theoret. Probab.*, **18**, 853–870. MR2289935

[47] Novak, S. Y. (2004). On self-normalized sums of random variables and the Student's statistic. *Theory Probab. Appl.*, **49**, 336–344. MR2144306

[48] O'Brien, G. L. (1980). A limit theorem for sample maxima and heavy branches in Galton-Watson trees. *J. Appl. Probab.*, **17**, 539–545. MR0568964

[49] De la Pena, V. H. and Giné, E. (1999). *Decoupling: From Dependence to Independence.* Springer, New York. MR1666908

[50] De la Pena, V. H., Klass, M. J. and Lai, T. L. (2000). Moment bounds for self-normalized martingales. In *High Dimensional Probability II* (E. Giné, D. M. Mason and J. A. Wellner, eds.), Birkhauser, Boston, 1–11. MR1857343

[51] De la Pena, V. H., Klass, M. J. and Lai, T. L. (2004). Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *Ann. Probab.*, **32**, 1902–1933. MR2073181

[52] De la Pena, V. H., Lai, T. L. and Shao, Q.-M. (2009). *Self-normalized Processes: Limit Theorey and Statistical Applications.* Springer, New York. MR2488094

[53] Prohorov, Yu. V. (1956). Covergence of random processes and limit theorems in probability theory. *Theor. Probab. Appl.*, **1**, 157–214. MR0084896

[54] Robinson, J. and Wang, Q. (2005). On the self-normalized Cramér-type large deviation. *Journal of Theoret. Probab.*, **18**, 891–909. MR2300002

[55] Shao, Q.-M. (1997). Self-normalized large deviations. *Ann. Probab.*, **25**, 285–328. MR1428510

[56] Shao, Q.-M. (1998). Recent development in self-normalized limit theorems. In: *Asymptotic Methods in Probability and Statistics* (B. Szyszkowicz, ed.), Elsevier Science, 467–480. MR1661499

[57] Shao, Q.-M. (1999). A Cramér type large deviation result for Student's *t*-statistic. *J. Theoret. Probab.*, **12**, 385–398. MR1684750

[58] Shao, Q.-M. (2004). Recent progress on self-normalized limit theorems. In: *Probability, Finance And Insurance*, World Sci. Publ., River Edge, NJ, 50–68. MR2189198

[59] Shao, Q.-M. (2005). An explicit Berry-Esseen bound for Student's *t*-statistic via Stein's method. In *Stein's Method and Applications*, Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap., 5, Singapore Univ. Press, Singapore, 143–155. MR2205333

[60] Shao, Q.-M. (2010). Stein's method, self-normalized limit theory and applications. In *Proceedings of the International Congress of Mathematicians*, Hyderabad, India, 1–25.

[61] SHAO, Q. M. and ZHOU, W. (2012). Cramér type moderate deviation theorems for Studentized non-linear statistics. *Preprint.*

[62] SHAO, Q. M., ZHANG, K. and ZHOU, W. (2012). Berry-Esseen bounds for Studentized non-linear statistics via Stein's method. *Preprint.*

[63] SLAVOV, V. V. (1985). On the Berry-Esseen bound for Student's statistic. In *Stability Problems for Stochastic Models* (V. V. Kalashnikov and V. M. Zolotarev, eds.), Lecture Notes in Math., 1155, Springer, Berlin, 355–390. MR0825335

[64] SEPANSKI, S. (1994). Asymptotics for Multivariate t-statistic and Hotelling's T^2-statistic under infinite second moments via bootstrapping. *J. Multivariate Anal.*, **49**, 41–54. MR1275042

[65] WANG, Q. (2004). On Darling Erdös type theorems for self-nornmalized sums, In *Asymptotic Methods in Stochastics: Festschrift for Miklos Csorgo* (Lajos Horvath and Barbara Szyszkowicz, eds.), International Conference on Asymptotic Methods in Stochastics. Fields Institute Communications, American Mathematical Society, 523–530. MR2108954

[66] WANG, Q. (2005). Limit theorems for self-normalized large deviation. *Electronic Journal of Probab.*, **10**, 1260–1285. MR2176384

[67] WANG, Q. (2011). Refined self-normalized large deviations for independent random variables. *Journal of Theor. Probab.*, **24**, 307–329. MR2795041

[68] WANG, Q. and HALL, P. (2009). Relative errors in central limit theorem for Student's t statistic, with applications. *Statistica Sinica*, **19**, 343–354. MR2487894

[69] WANG, Q. and JING, B.-Y. (1999). An exponential nonuniform Berry-Esseen bound for self-normalized sums. *Ann. Probab.*, **27**, 2068–2088. MR1742902

[70] WANG, Q., JING, B.-Y. and ZHAO, L. (2000). The Berry-Esseen bound for Studentized statistics. *Ann. Probab.*, **28**, 511–535. MR1756015

[71] ZHOU, W. and JING, B.-Y. (2006). Tail probability approximations for Student's t statistics. *Probab. Theory Relat. Fields*, **136**, 541–559. MR2257135