Title
Self-normalization: Taming a wild population in a heavy-tailed world

Journal
Applied Mathematics-A Journal of Chinese Universities, 32(3)

ISSN
1005-1031

Authors
Shao, Qi-man
Zhou, Wen-xin

Publication Date
2017-09-01

DOI
10.1007/s11766-017-3552-y

Peer reviewed

Appl. Math. J. Chinese Univ.
2017, 32(3): 253-269

Self-normalization: Taming a wild population in a heavy-tailed world

SHAO Qi-man[1] ZHOU Wen-xin[2]

Abstract. The past two decades have witnessed the active development of a rich probability theory of Studentized statistics or self-normalized processes, typified by Student's t-statistic as introduced by W. S. Gosset more than a century ago, and their applications to statistical problems in high dimensions, including feature selection and ranking, large-scale multiple testing and sparse, high dimensional signal detection. Many of these applications rely on the robustness property of Studentization/self-normalization against heavy-tailed sampling distributions. This paper gives an overview of the salient progress of self-normalized limit theory, from Student's t-statistic to more general Studentized nonlinear statistics. Prototypical examples include Studentized one- and two-sample U-statistics. Furthermore, we go beyond independence and glimpse some very recent advances in self-normalized moderate deviations under dependence.

§1 Introduction

As one of the most important statistics, Student's t-statistic [71] has a wide range of applications in probability, statistics, finance and other fields of science. During the past century, the t-statistic has evolved into much more general Studentized statistics and self-normalized processes, and as noted in [28], it is finding applications today that were never envisaged when it was introduced. The past two decades have also witnessed the significant development of a rich probability theory of self-normalized processes, beginning with weak convergence [55], laws of the iterated logarithm [39] and exponential and moment bounds [38] and culminating in large and moderate deviations for self-normalized sums of both independent and dependent random variables [63, 64, 46, 19]. The main goal of this paper is to provide an overview of the developments of self-normalized limit theory, from Student's t-statistic to more general Studentized nonlinear statistics.

Received: 2017-06-01. Revised: 2017-07-13.
MR Subject Classification: 60F05, 60F10, 62E17, 62E20.
Keywords: Berry-Esseen inequality, Hotelling's T^2-statistic, large deviation, moderate deviation, self-normalization, Student's t-statistic, U-statistic.
Digital Object Identifier(DOI): 10.1007/s11766-017-3552-y.
Supported by Hong Kong RGC GRF 14302515.

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (IID) random variables drawn from X, a real-valued random variable with mean μ and variance $\sigma^2 > 0$. For some prespecified $\mu_0 \in \mathbb{R}$, we consider testing

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0,$$

which is one of the most fundamental hypothesis testing problems in statistics. Without loss of generality, we assume $\mu_0 = 0$; otherwise, it suffices to replace X_i with $X_i - \mu_0$. Define

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i, \quad \hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad S_n = \sum_{i=1}^n X_i \quad \text{and} \quad V_n^2 = \sum_{i=1}^n X_i^2.$$

Using the above notation, we can write the z-statistic and Student's t-statistic as

$$Z_n = \frac{\sqrt{n}\bar{X}_n}{\sigma} \quad \text{and} \quad T_n = \frac{\sqrt{n}\bar{X}_n}{\hat{\sigma}_n} = \frac{S_n/V_n}{\sqrt{\{n-(S_n/V_n)^2\}/(n-1)}}, \tag{1}$$

respectively. Student's t-statistic is one of the statistics which are most commonly used to conduct hypothesis testing for μ when σ is unknown, while Z_n is invoked when σ is known, and to construct the confidence interval.

Under the normality assumption $X \sim N(\mu, \sigma^2)$ and the null hypothesis $H_0 : \mu = \mu_0$, it is known that Z_n has a standard normal distribution and T_n follows Student's t-distribution with $n-1$ degrees of freedom. When normality is violated, based on the central limit theorem (CLT) and the law of large numbers, quite often statisticians recommend using the normal distribution as an approximation to the (unknown) distribution of T_n as long as the sample size is sufficiently large, say $n \geq 30$. This naturally leads to the question of how good this normal approximation can be, or equivalently how accurate the estimated p-value (based on normal calibration) is. In this paper, we review the asymptotic properties of T_n when the distribution of X deviates from the normal distribution and may even have very heavy tails.

Unlike the z-statistic, Student's t-statistic is a highly nonlinear statistic, that is, $T_n = f(X_1, \ldots, X_n)$ for some nonlinear function $f : \mathbb{R}^n \mapsto \mathbb{R}$, which makes the study of its distributional properties much more difficult. A key observation that facilitates analysis is the following equivalence between T_n and S_n/V_n [30]:

$$\{T_n \geq t\} = \left\{\frac{S_n}{V_n} \geq t\sqrt{\frac{n}{n+t^2-1}}\right\}, \quad t \geq 0.$$

This enables us to consider only the distributional properties of the less complex S_n/V_n, which is referred to as the self-normalized sum. It turns out that the limiting properties of S_n/V_n usually require much less stringent moment conditions than those for Z_n and hence provide a much wider practical applicability. The key intuition behind these properties is that erratic fluctuations in S_n tend to be canceled, or at least dampened, by those of V_n, much more so than if V_n were replaced by its population counterpart.

The rest of this paper is organized as follows. In Section 2, we briefly review the history and development of Student's t-statistic. Sections 3 and 4 review the self-normalized limit theory for t-statistics, including weak convergence, Berry-Esseen bounds, large deviations and Cramér-type moderate deviations. In Section 5, we go beyond the t-statistic and focus on more general Studentized nonlinear statistics, typified by the Studentized U-statistic. Some recent normal approximation results for Hotelling's T^2-statistics are given in Section 6. Finally, in Section 7, we mention some recent progress on self-normalized moderate deviation results for Studentized two-sample U-statistics and self-normalized sums of weakly dependent data.

§2 History of Student's t-statistic

The t-statistic was introduced by W. S. Gosset in his work entitled "The probable error of a mean," which he published under the *nom de plume* of "Student" [71]. As a chemist working at the Guinness Brewery in Dublin, Ireland, Gosset was interested in the chemical properties of barley where sample sizes might number as few as three. His first studies resulted in a report, "The Application of the 'Law of Error' to the work of the Brewery" dated November 3, 1904, although it has never been published. A meeting with Professor Karl Pearson in July 1905 had a big impact on Gosset's research. Thanks to Guinness's enlightened policy that allowed technical staff leave for study, Gosset spent the first two terms of the 1906/07 academic year in Karl Pearson's Biometric Laboratory at University College London [78].

To prevent disclosure of confidential information, Guinness prohibited its employees from publishing any papers regardless of the contained information. Gosset pleaded that his mathematical and philosophical conclusions were of no possible practical use to competing brewers, and finally was allowed to publish them under a pseudonym to avoid difficulties with the rest of the staff. The pseudonym "Student" was selected by Christopher Digges La Touche, the Managing Director of Arthur Guinness & Company. Pearson, the co-founder and editor-in-chief of the journal *Biometrika*, helped Gosset with the 1908 paper. Pearson had little appreciation of its importance because the paper addressed the brewer's concern with small samples, while biometricians typically had hundreds of observations and saw no urgency in developing small-sample methods. Student's t-distribution became well known through the work of R. A. Fisher, who called the distribution "Student's distribution" and represented the test value with the letter t [36]. For detailed discussions of the development of t-distributions, see [60], [30], [11] and [78], among others.

Student's t-statistic, along with the t-test, has become one of the most commonly used methods in statistics and related fields. A self-normalized statistic refers to a statistic that is normalized by an estimator of the nuisance parameter instead of the nuisance parameter itself. According to Delaigle, Hall and Jin [28], "Student's t-statistic is finding applications today that were never envisaged when it was introduced more than a century ago. Many of these applications rely on properties, for example robustness against heavy-tailed distributions, that were not explicitly considered until relatively recently."

Motivated by various applications of t-statistics/t-tests to high dimensional statistical analysis, including large-scale multiple testing [34, 13, 54], signal detection [28], classification [33] and feature screening [15, 16], in the following sections we review and explore Berry-Esseen type bounds, moderate and large deviations of Student's t-statistics and some other important self-normalized/Studentized statistics. These results reveal several attractive advantages of self-normalization/Studentization that are indispensable to understanding even common procedures for analyzing high dimensional data and also motivate new methods.

§3 Central limit theorem and Berry-Esseen bounds

3.1 Central limit theorem and invariance principle

Proceeding with the notation introduced in Section 1, we assume that $\mathbb{E}(X) = 0$. Efron [30] may be the first to investigate the limiting behavior of Student's t-statistic T_n, or equivalently the self-normalized sum S_n/V_n, in some special cases. The general research begins with Logan et al. [55], who proved, among many other results, that if X is in the domain of attraction of an α-stable law with $0 < \alpha \le 2$, centered if $\alpha > 1$ and symmetric if $\alpha = 1$, then S_n/V_n converges in distribution to a limit, which is sub-Gaussian. In particular, if X is symmetric, the moments of S_n/V_n also converge to those of this limit. Moreover, the authors conjectured that S_n/V_n is asymptotically normal if and only if X is in the domain of attraction of the normal law, and the only possible nontrivial limiting distributions of S_n/V_n are those obtained when X follows a stable law. The "if" part, as Maller [56] noted, is relatively easy based on Raikov's theorem. See [25] and [40] for more discussions. The "only if" remained open until Giné, Götze and Mason [38] proved the result for the general case of not necessarily symmetric random variables.

Giné, Götze and Mason [38] also showed that if the sequence $\{S_n/V_n\}_{n\ge1}$ is stochastically bounded, then it is uniformly sub-Gaussian, i.e., there exists some constant $c > 0$ such that $\sup_{n\ge1} \mathbb{E}e^{tS_n/V_n} \le 2e^{ct^2}$ for all $t \in \mathbb{R}$. The second conjecture of Logan et al. [55] was addressed by Chistyakov and Götze [22]. In the independent but not necessarily identically distributed case, Mason [57] studied the limiting behaviors for self-normalized triangular arrays. The extension of self-normalized CLT to Donsker-type functional CLT was established by Csörgő, Szyszkowicz and Wang [26], who also proved an invariance principle for self-normalized, self-randomized partial sum processes of independent random variables.

As we assume $\mathbb{E}(X) = 0$, the aforementioned investigation of the asymptotic behaviors for S_n/V_n is specifically related to centralized t-statistics. The limiting behaviors of the non-central Student's t-statistic was discussed in Bentkus et al. [7]. Under the assumption of $\mathbb{E}(X^2) < \infty$, the limiting behaviors of the non-central t-statistic are different under the two scenarios of $\mathbb{E}(X^4) < \infty$ and $\mathbb{E}(X^4) = \infty$.

3.2 Berry-Esseen bounds

Assume that $\mathbb{E}(X) = 0$. The self-normalized CLT states that if X is in the domain of attraction of the normal law, then

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(S_n \ge xV_n) - \{1 - \Phi(x)\}| \to 0 \quad \text{as } n \to \infty.$$

The CLT is useful when x is not too large or when the error is well controlled. There are two ways to measure the normal approximation error. The first is to study the absolute error via a Berry-Esseen bound or an Edgeworth expansion, and the second is to estimate the relative error of $\mathbb{P}(S_n \geq xV_n)$ to $1-\Phi(x)$. In this section, we focus on the former.

Define $b_n = \sup\{x \in \mathbb{R} : \mathbb{E}\{X^2 I(|X| \leq x)\} \geq x^2/n\}$ and

$$\delta_n = n\mathbb{P}(|X| > b_n) + nb_n^{-1}|\mathbb{E}\{XI(|X| \leq b_n)\}| + nb_n^{-3}\mathbb{E}\{|X|^3 I(|X| \leq b_n)\}. \tag{2}$$

Making major progress in this direction, Bentkus, Bloznelis and Götze [5] refined the results of Slavova [70] and Hall [42] and proved the following theorem.

Theorem 3.1. *If X is in the domain of attraction of the normal law, then* $\sup_{x\in\mathbb{R}} |\mathbb{P}(S_n \leq xV_n) - \Phi(x)| \leq C\,\delta_n$*, where $C > 0$ is an absolute constant and δ_n is given in* (2)*.*

We refer to [31] and [58] for explicit constants in both uniform and non-uniform Berry-Esseen bounds. The preceding result was extended to the case of independent but not necessarily identically distributed random variables by Bentkus, Bloznelis and Götze [5] and Shao [65].

The Berry-Esseen bounds provide an upper bound for the rate of convergence in the CLT. To fully characterize the convergence rate, [43] investigated the exact rate and leading term in the CLT. They showed that the rate of convergence of the t-statistic to normality is strictly faster than that for the z-statistic when the second moment is only just finite. In the case of finite third moment, this leading term is asymptotic to its conventional form in an Edgeworth expansion. More results on the Edgeworth expansion for Student's t-statistics can be found in [41] and [10].

Wang and Jing [76] were the first to investigate the non-uniform Berry-Esseen bound for S_n/V_n. Their result was later extended by Robinson and Wang [62], who established an exponential non-uniform bound which was established under optimal moment conditions. The following result is taken from Theorem 3 in [62].

Theorem 3.2. *If X is in the domain of attraction of the normal law, then there exists some $\eta \in (0,1)$ such that $|\mathbb{P}(S_n \leq xV_n) - \Phi(x)| \leq C\,\delta_n\, e^{-\eta x^2/2}$ for all $x \in \mathbb{R}$, where $C > 0$ is an absolute constant and δ_n is given in* (2)*.*

§4 Large and moderate deviations

In this section, we review the self-normalized large and moderate deviation results, which characterize the relative error of the normal approximation. More specifically, a Cramér-type moderate deviation is used to consider the problem of estimating the relative error of the tail probability of T_n against the tail probability of its limiting distribution, that is,

$$\frac{\mathbb{P}(T_n \geq x)}{1-\Phi(x)} \quad \text{for } x \geq 0.$$

Assume that the p-value of the test is $\mathbb{P}(T_n \geq x_0)$. As the exact p-value is usually unknown, it is a common practice to use the limiting tail probability $\mathbb{P}(Z \geq x_0)$ to estimate the p-value. As such, the Cramér-type moderate deviation quantifies the accuracy of the estimated p-value. Moderate deviation results have been successfully applied to multiple hypothesis tests based on

t-statistics [34, 24, 28], feature selection in classification [35] and square-root Lasso for recovery of sparse signals [4].

It is well known that moment conditions or other related conditions are necessary and sufficient for many classical limit theorems for the conventional mean. On the contrary, its Studentized counterpart admits accurate large deviation approximations in heavy-tailed cases where the sampling distribution has only a small number of finite moments. The classical Cramér-Chernoff large deviation [20] states that if $\mathbb{E}(e^{t_0 X}) < \infty$ for some $t_0 > 0$, then for every $x > \mathbb{E}(X)$,

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}(S_n/n \geq x) = \ln \rho(x),$$

where $\rho(x) = \inf_{t\geq 0} e^{-tx}\mathbb{E}(e^{tX})$. The self-normalized large deviation [63], however, holds without any moment assumptions:

$$\lim_{n\to\infty} \mathbb{P}(S_n/V_n \geq x\sqrt{n}\,) = \sup_{b\geq 0} \inf_{t\geq 0} \mathbb{E}e^{t\{bX - x(X^2+b^2)/2\}}$$

for $x > 0$ if $\mathbb{E}(X) = 0$ or $\mathbb{E}(X^2) = \infty$. Moreover, Shao [63] showed that the tail probability of S_n/V_n is Gaussian-like when X 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. Specifically, assuming only a finite second moment, he proved that

$$\lim_{n\to\infty} \frac{\ln \mathbb{P}(T_n \geq x)}{\ln \mathbb{P}(t_{n-1} \geq x)} = \lim_{n\to\infty} \frac{\ln \mathbb{P}(T_n \geq x)}{\ln\{1 - \Phi(x)\}} = 1$$

holds uniformly in x in the interval $0 \leq x \leq o(\sqrt{n})$, where t_{n-1} has a t-distribution with $n-1$ degrees of freedom. These results also lead to a precise constant in Griffin and Kuelbs' self-normalized law of the iterated logarithm [39]. In a subsequent paper, Shao [64] proved a Cramér-type moderate deviation result: if $\mathbb{E}(X) = 0$ and $\mathbb{E}(|X|^3) < \infty$, then

$$\lim_{n\to\infty} \frac{\mathbb{P}(S_n \geq xV_n)}{1 - \Phi(x)} = 1$$

holds uniformly in $0 \leq x \leq o(n^{1/6})$. On the contrary, a finite moment generating function of $|X|^{1/2}$ is necessary for a similar result in relation to the z-statistic. The following large and moderate deviation results for the z-statistic are borrowed from Linnik [52]. Suppose $X, X_1, \ldots, X_n$ are IID random variables with $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) = 1$.

(i) If $\mathbb{E}(e^{t_0|X|^\alpha}) < \infty$ for some $t_0 > 0$ and $0 < \alpha \leq 1$, then

$$\lim_{n\to\infty} \frac{1}{x_n^2} \ln \mathbb{P}(S_n/\sqrt{n} \geq x_n) = -1/2$$

for any sequence $\{x_n\}_{n\geq 1}$ satisfying $x_n \to \infty$ and $x_n = o(n^{\alpha/(4-2\alpha)})$.

(ii) If $\mathbb{E}(e^{t_0|X|^\alpha})$ for some $t_0 > 0$ and $0 < \alpha \leq 1/2$, then

$$\frac{\mathbb{P}(S_n/\sqrt{n} \geq x)}{1 - \Phi(x)} \to 1$$

holds uniformly for $0 \leq x \leq o(n^{\alpha/(4-2\alpha)})$.

(iii) Assume $\mathbb{E}(e^{t_0 X}) < \infty$ for some $t_0 > 0$. Then,

$$\mathbb{P}(S_n/\sqrt{n} \geq x) = \{1 - \Phi(x)\} \exp\left(\frac{x^3 \mathbb{E}X^3}{6\sqrt{n}}\right)\left\{1 + O\left(\frac{1+x}{\sqrt{n}}\right)\right\}$$

holds uniformly for $0 \le x \le o(n^{1/4})$.

The results in Shao [64] were further extended to independent (not necessarily identically distributed) random variables by Jing, Shao and Wang [46] under a Lindeberg-type condition, where the authors actually established more general frameworks and considered applications of the iterated logarithm and Studentized bootstrap to the self-normalized law. For $0 < \delta \le 1$, let

$$d_{n,\delta} = \bigg(\sum_{i=1}^n \mathbb{E}X_i^2\bigg)^{1/2} \Big/ \bigg(\sum_{i=1}^n \mathbb{E}|X_i|^{2+\delta}\bigg)^{1/(2+\delta)}.$$

The following result is a version of Theorems 2.1 and 2.3 in [46].

Theorem 4.1. *Let $X_1, \ldots, X_n$ be independent random variables satisfying $\mathbb{E}(X_i) = 0$, $\mathbb{E}(X_i^2) > 0$ and $\mathbb{E}(|X_i|^{2+\delta})$ for $0 < \delta \le 1$ for all i. Then, there exists an absolute constant $C > 0$ such that*

$$\bigg|\frac{\mathbb{P}(S_n \ge xV_n)}{1 - \Phi(x)} - 1\bigg| \le C\bigg(\frac{1+x}{d_{n,\delta}}\bigg)^{2+\delta} \tag{3}$$

holds for all $0 \le x \le d_{n,\delta}$.

Wang [75] established a refined, second-order Cramér-type moderate deviation theorem for S_n/V_n under the condition when the fourth moments are finite, the best result known to date.

There are several further extensions in the IID setting. For example, Chistyakov and Götze [21] proved the sharpness of the result in [46]. Robinson and Wang [62] proved a Cramér-type result under the optimal condition of X being in the domain of attraction of the normal law. Assuming $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^4) < \infty$, Wang [74] proved that

$$\mathbb{P}(S_n \ge xV_n) = \{1 - \Phi(x)\} \exp\bigg(-\frac{x^3\mathbb{E}X^3}{3\sigma^3\sqrt{n}}\bigg)\bigg\{1 + O\bigg(\frac{1+x}{\sqrt{n}}\bigg)\bigg\}$$

holds uniformly in $0 \le x \le O(n^{1/6})$. The result of Wang [74] was recently extended by Gao, Shao and Shi [37] to a more general self-normalized sum $\sum_{i=1}^n X_i/(\sum_{i=1}^n Y_i^2)^{1/2}$, where $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent random variables.

In addition, Bercu, Gassiat and Rio [8] obtained large and moderate deviation results for self-normalized empirical processes. To obtain a better estimate of $\mathbb{P}(S_n \ge xV_n)$ in a median or large range of x, we need completely different techniques. Jing, Shao and Zhou [47] and Zhou and Jing [79] investigated saddle-point approximations for the tail probability $\mathbb{P}(S_n \ge xV_n)$ in the IID setting when x is very large, that is, $x = c\sqrt{n}$ for some $0 < c < 1$. Still, it remains an open question whether the techniques in [47] and [79] can be used to provide a better approximation for $\mathbb{P}(S_n \ge xV_n)$ in a median range of x, say $O(n^{1/6}) \le x \le O(n^{1/2})$. Jing, Shao and Zhou [48] established a universal self-normalized moderate deviation when X is in the centered Feller class. We refer to [27] for a systematic presentation of the general self-normalized limit theory and its statistical applications, and to [66] and [68] for two comprehensive surveys.

§5 Moderate deviations for Studentized U-statistics

The purpose of this section is to go beyond self-normalized sums and catch a glimpse of more general self-normalized processes.

5.1 Studentized nonlinear statistics

The research on self-normalized processes is motivated by Studentized nonlinear statistics. Nonlinear statistics are the building blocks in various statistical inference problems. Many of them can be written as a partial sum plus a negligible term, for example, due to the Hoeffding decomposition or Bahadur representation. Typical examples include U-statistics, multi-sample U-statistics, L-statistics, random sums and functions of nonlinear statistics. We refer to Chen and Shao [18] for a unified approach to uniform and non-uniform Berry-Esseen bounds for standardized nonlinear statistics.

Assume that the nonlinear statistic of interest can be decomposed as a standardized partial sum of independent random variables plus a remainder, say, $\sigma^{-1}(\sum_{i=1}^n \xi_i + D_{1n})$, where $\xi_1, \ldots, \xi_n$ are independent random variables satisfying

$$\mathbb{E}\xi_i = 0 \text{ for } i = 1, \ldots, n, \quad \text{and} \quad \sum_{i=1}^n \mathbb{E}\xi_i^2 = 1, \tag{4}$$

and where $D_{1n} = D_{1n}(\xi_1, \ldots, \xi_n)$. As σ is typically unknown, a Studentized statistic

$$T_n = \frac{1}{\widehat{\sigma}}\left(\sum_{i=1}^n \xi_i + D_{1n}\right)$$

is more commonly used in practice, where $\widehat{\sigma}$ is an estimator of σ that can be written as $\widehat{\sigma} = \{(\sum_{i=1}^n \xi_i^2)(1 + D_{2n})\}^{1/2}$, where $D_{2n} = D_{2n}(\xi_1, \ldots, \xi_n)$ satisfies $1 + D_{2n} > 0$. Without loss of generality, we assume $\sigma = 1$. Under (4), T_n can be written as

$$T_n = \frac{W_n + D_{1n}}{V_n(1 + D_{2n})^{1/2}}, \tag{5}$$

where $W_n = \sum_{i=1}^n \xi_i$ and $V_n = (\sum_{i=1}^n \xi_i^2)^{1/2}$. The basic observation underpinning (5) is that for a nonlinear statistic that be can written as a partial sum plus a negligible remainder, the corresponding normalizing term should be dominated by a quadratic form. Examples satisfying (5) include the t-statistic, Studentized U-statistics and L-statistics. We refer to Wang, Jing and Zhao [77] and the references therein for more detailed discussions.

Shao and Zhou [69] established a general Cramér-type moderate deviation theorem for T_n in the form of (5), which is reproduced as follows. For $x \geq 1$, write

$$L_{n,x} = \sum_{i=1}^n \delta_{i,x}, \quad I_{n,x} = \mathbb{E}\exp(xW_n - x^2V_n^2/2) = \prod_{i=1}^n \mathbb{E}\exp(\xi_{i,x} - \xi_{i,x}^2/2), \tag{6}$$

where $\delta_{i,x} = \mathbb{E}\xi_{i,x}^2 I(|\xi_{i,x}| > 1) + \mathbb{E}|\xi_{i,x}|^3 I(|\xi_{i,x}| \leq 1)$ with $\xi_{i,x} := x\xi_i$. For $i = 1, \ldots, n$, let $D_{1n}^{(i)}$ and $D_{2n}^{(i)}$ be arbitrary measurable functions of $\{\xi_j\}_{j=1, j\neq i}^n$, such that $\{D_{1n}^{(i)}, D_{2n}^{(i)}\}$ and ξ_i are independent. Moreover, define

$$\begin{aligned} R_{n,x} = I_{n,x}^{-1} \times \Big(& \mathbb{E}\{(x|D_{1n}| + x^2|D_{2n}|)e^{\sum_{j=1}^n(\xi_{j,x} - \xi_{j,x}^2/2)}\} \\ & + \sum_{i=1}^n \mathbb{E}[\min(|\xi_{i,x}|, 1)\{|D_{1n} - D_{1n}^{(i)}| + x|D_{2n} - D_{2n}^{(i)}|\}e^{\sum_{j\neq i}(\xi_{j,x} - \xi_{j,x}^2/2)}]\Big). \end{aligned}$$

Here, we use $\sum_{j\neq i} = \sum_{j=1, j\neq i}^n$ for simplicity. We are now ready to present the main results from [69].

Theorem 5.1. *Let T_n be as in* (5) *under condition* (4). *Then, there exist positive absolute constants C_1–C_4 and c_1 such that*

$$\mathbb{P}(T_n \geq x) \geq \{1 - \Phi(x)\} \exp(-C_1 L_{n,x})(1 - C_2 R_{n,x})$$

and

$$\mathbb{P}(T_n \geq x) \leq \{1 - \Phi(x)\} \exp(C_3 L_{n,x})(1 + C_4 R_{n,x}) + \mathbb{P}(x|D_{1n}| > V_n/4) + \mathbb{P}(x^2|D_{2n}| > 1/4)$$

for all $x \geq 1$ satisfying $\max_{1\leq i\leq n} \delta_{i,x} \leq 1$ and $L_{n,x} \leq c_1 x^2$.

The quantity $L_{n,x}$ in (6) is essentially the same as the factor $\Delta_{n,x}$ in [46], which is the leading term that describes the accuracy of the relative normal approximation error. Theorem 5.1 provides upper and lower bounds of the relative errors when $x \geq 1$. For completeness, the following result covers the case of $0 \leq x \leq 1$ [69]. We refer to [68] for general Berry-Esseen bounds for Studentized nonlinear statistics.

Theorem 5.2. *There exists an absolute constant $C > 1$ such that for all $x \geq 0$,*

$$|\mathbb{P}(T_n \leq x) - \Phi(x)| \leq C\breve{R}_{n,x},$$

where

$$\breve{R}_{n,x} := L_{n,1+x} + \mathbb{E}|D_{1n}| + x\mathbb{E}|D_{2n}| + \sum_{i=1}^{n} \mathbb{E}[|\xi_i I\{|\xi_i| \leq 1/(1+x)\}\{|D_{1n} - D_{1n}^{(i)}| + x|D_{2n} - D_{2n}^{(i)}|\}]$$

for $L_{n,1+x}$ as in (6).

In particular, when $0 \leq x \leq 1$, the quantity $L_{n,1+x}$ satisfies

$$L_{n,1+x} \leq (1+x)^2 \sum_{i=1}^{n} \mathbb{E}\xi_i^2 I(|\xi_i| > 1) + (1+x)^2 \sum_{i=1}^{n} \mathbb{E}\xi_i^2 I(1/2 < |\xi_i| \leq 1) + (1+x)^3 \sum_{i=1}^{n} \mathbb{E}|\xi_i|^3 I(|\xi_i| \leq 1),$$

which can be further bounded, up to a constant, by

$$\sum_{i=1}^{n} \mathbb{E}\xi_i^2 I(|\xi_i| > 1) + \sum_{i=1}^{n} \mathbb{E}|\xi_i|^3 I(|\xi_i| \leq 1).$$

When $D_{1n} = D_{2n} = 0$, T_n reduces to the self-normalized sum of independent random variables, and thus Theorems 5.1 and 5.2 together immediately imply the main result from [46]. Also, D_{1n} and D_{2n} in the definitions of $R_{n,x}$ and $\breve{R}_{n,x}$ can be replaced by any non-negative random variables D_{3n} and D_{4n}, respectively, provided that $|D_{1n}| \leq D_{3n}$, $|D_{2n}| \leq D_{4n}$. Condition (4) implies that ξ_i actually depends on both n and i; that is, ξ_i denotes ξ_{ni}, which is an array of independent random variables.

5.2 Studentized U-statistics

As a prototypical example of the Studentized nonlinear statistic given in (5), the Studentized U-statistic is of particular interest. Based on Theorems 5.1 and 5.2, Shao and Zhou [69] obtained a sharp Cramér-type moderate deviation for Studentized U-statistics under optimal moment conditions.

Let $X_1, X_2, \ldots, X_n$ be a sequence of IID random variables and let $h : \mathbb{R}^m \to \mathbb{R}$ be a symmetric Borel measurable function of m variables, where $2 \le m < n/2$ is fixed. Hoeffding's U-statistic with a kernel h of degree m is defined as [44]

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

which is an unbiased estimate of $\theta = \mathbb{E}h(X_1, \ldots, X_m)$. Let

$$h_1(x) = \mathbb{E}\{h(X_1, X_2, \ldots, X_m)|X_1 = x\},\ x \in \mathbb{R}$$

and

$$\sigma^2 = \mathrm{var}\{h_1(X_1)\}, \quad \sigma_h^2 = \mathrm{var}\{h(X_1, X_2, \ldots, X_m)\}.$$

Assuming $0 < \sigma^2 < \infty$, the standardized non-degenerate U-statistic is given by

$$Z_n^U = \frac{\sqrt{n}}{m\sigma}(U_n - \theta).$$

The U-statistic is one of the most commonly used nonlinear and nonparametric statistics, and its asymptotic theory has been well studied since the seminal work of Hoeffding [44]. However, because σ is usually unknown, the Studentized U-statistic [2], denoted by

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

is of more practical interest, where s_1^2 denotes the leave-one-out Jackknife estimator of σ^2 given by

$$s_1^2 = \frac{(n-1)}{(n-m)^2} \sum_{i=1}^n (q_i - U_n)^2 \quad \text{with}$$

$$q_i = \frac{1}{\binom{n-1}{m-1}} \sum_{\substack{1 \le \ell_1 < \cdots < \ell_{m-1} \le n \\ \ell_j \ne i,\, j=1,\ldots,m-1}} h(X_i, X_{\ell_1}, \ldots, X_{\ell_{m-1}}).$$

In contrast to the standardized U-statistic Z_n^U, few optimal limit theorems are available for Studentized U-statistics in the literature. A uniform Berry-Esseen bound for Studentized U-statistics was proved by Wang, Jing and Zhao [77] for $m = 2$ when $\mathbb{E}|h(X_1, X_2)|^3 < \infty$. Partial results on Cramér-type moderate deviation were obtained in [72], [73] and [51]. Recently, Shao and Zhou [69] established the following sharp Cramér-type moderate deviation theorem for the Studentized U-statistic T_n^U.

Theorem 5.3. *Assume that $\sigma_p := (\mathbb{E}|h_1(X_1) - \theta|^p)^{1/p} < \infty$ for some $2 < p \le 3$. Suppose there are constants $c_0 \ge 1$ and $\tau \ge 0$ such that*

$$\{h(x_1, \ldots, x_m) - \theta\}^2 \le c_0 \left[\tau\sigma^2 + \sum_{i=1}^m \{h_1(x_i) - \theta\}^2\right]. \tag{7}$$

Then, there exist constants $C_1, c_1 > 0$ independent of n such that

$$\frac{\mathbb{P}(T_n^U \ge x)}{1 - \Phi(x)} = 1 + O(1)\left\{(\sigma_p/\sigma)^p \frac{(1+x)^p}{n^{p/2-1}} + (\sqrt{a_m} + \sigma_h/\sigma)\frac{(1+x)^3}{\sqrt{n}}\right\}$$

holds uniformly for $0 \le x \le c_1 \min\{(\sigma/\sigma_p) n^{1/2-1/p}, (n/a_m)^{1/6}\}$, where $|O(1)| \le C_1$ and $a_m =$

$\max(c_0\tau, c_0 + m)$. *In particular,*

$$\frac{\mathbb{P}(T_n^U \geq x)}{1-\Phi(x)} \to 1 \tag{8}$$

holds uniformly in $x \in [0, o(n^{1/2-1/p}))$.

Condition (7) is satisfied for the t-statistic ($h(x_1, x_2) = (x_1 + x_2)/2$ with $c_0 = 2$ and $\tau = 0$), sample variance ($h(x_1, x_2) = (x_1 - x_2)^2/2$, $c_0 = 10$, $\tau = \theta^2/\sigma^2$), Gini's mean difference ($h(x_1, x_2) = |x_1 - x_2|$, $c_0 = 8$, $\tau = \theta^2/\sigma^2$) and one-sample Wilcoxon's statistic ($h(x_1, x_2) = I(x_1 + x_2 \leq 0)$, $c_0 = 1$, $\tau = 1/\sigma^2$). Result (8) was proved earlier by Lai, Shao and Wang [51] for $m = 2$.

§6 Hotelling's T^2-statistics

Testing the equality of two mean vectors μ_1 and μ_2 based on two random samples is a canonical testing problem in multivariate analysis, and it arises in many scientific applications, including genomics, finance and signal processing. Let $\{\mathbf{X}_i\}_{i=1}^{n_1}$ and $\{\mathbf{Y}_j\}_{j=1}^{n_2}$ be two samples of IID d-dimensional random vectors with mean vectors μ_1 and μ_2 and positive definite covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, respectively. Assume the two samples are independent. The classical test for testing

$$H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \mu_1 \neq \mu_2$$

is Hotelling's T^2-test, with the test statistic given by [45]

$$T^2_{n_1,n_2} = (\bar{\mathbf{X}} - \bar{\mathbf{Y}})^{\mathrm{T}}\left(\frac{1}{n_1}\hat{\mathbf{\Sigma}}_1 + \frac{1}{n_2}\hat{\mathbf{\Sigma}}_2\right)^{-1}(\bar{\mathbf{X}} - \bar{\mathbf{Y}}),$$

where $\bar{\mathbf{X}} = (1/n_1)\sum_{i=1}^{n_1}\mathbf{X}_i$ and $\bar{\mathbf{Y}} = (1/n_2)\sum_{j=1}^{n_2}\mathbf{Y}_j$ are the sample means and $\hat{\mathbf{\Sigma}}_1 = (1/n_1)\sum_{i=1}^{n_1}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{\mathrm{T}}$ and $\hat{\mathbf{\Sigma}}_2 = (1/n_2)\sum_{j=1}^{n_2}(\mathbf{Y}_j - \bar{\mathbf{Y}})(\mathbf{Y}_j - \bar{\mathbf{Y}})^{\mathrm{T}}$ are the sample covariance matrices. The properties of Hotelling's T^2-statistic under normality have been well studied in the conventional low dimensional setting. The properties are desirable when the dimension d is fixed [1].

In the one-sample case, the T^2-statistic is defined by

$$T^2_{n_1} = n_1(\bar{\mathbf{X}} - \mu_1)^{\mathrm{T}}\hat{\mathbf{\Sigma}}_1^{-1}(\bar{\mathbf{X}} - \mu_1).$$

If $\{\mathbf{X}_i\}_{i=1}^{n_1}$ is a sample from a multivariate normal population $N(\mu_1, \mathbf{\Sigma}_1)$, then $\{(n_1 - d)/d\}\{T^2_{n_1}/(n_1 - 1)\}$ follows an F-distribution. When d is fixed and if the underlying distribution has a finite second moment, the limiting law of $T^2_{n_1}$ is the chi-squared distribution with degrees of freedom d. Large and moderate deviations (the logarithm of the tail probabilities) were obtained by Dembo and Shao [29].

Recently, Liu and Shao [53] established a Cramér-type moderate deviation theorem for Hotelling's T^2-statistic in both one- and two-sample cases. Specifically, they proved that if $\mathbb{E}(\|\mathbf{X}_1\|_2^{3+\delta}) + \mathbb{E}(\|\mathbf{Y}_1\|_2^{3+\delta}) < \infty$ for some $\delta > 0$ and $n_1 \asymp n_2$, then under $H_0 : \mu_1 = \mu_2$,

$$\frac{\mathbb{P}(T^2_{n_1,n_2} \geq x^2)}{\mathbb{P}(\chi^2_d \geq x^2)} \to 1 \quad \text{as } n \to \infty$$

uniformly for $x \in [0, o(n^{1/6}))$, where $n = n_1 + n_2$ and χ^2_d has a chi-squared distribution with degrees of freedom d. A similar result holds for the one-sample T^2-statistic: if $\mathbb{E}(\|\mathbf{X}_1\|_2^{3+\delta}) < \infty$ for some $\delta > 0$, then

$$\frac{\mathbb{P}(T_{n_1}^2 \geq x^2)}{\mathbb{P}(\chi_d^2 \geq x^2)} \to 1 \quad \text{as } n_1 \to \infty$$

uniformly for $x \in [0, o(n_1^{1/6}))$. As proved in [64] and [46], the preceding results hold under finite third moments when $d = 1$ and the range $[0, o(n^{1/6}))$ ($[0, o(n_1^{1/6}))$ in the one-sample case) is the widest possible. An open question is whether they remain valid for $d \geq 2$ under finite third moments.

Another interesting problem arises when the dimension d is large or proportional to the sample size. Pan and Zhou [59] studied the asymptotic distribution of $T_{n_1}^2$ when $d = d_{n_1}$ satisfies $d \leq n_1$ and $d/n_1 \to c \in (0, 1)$ as $n_1 \to \infty$. We refer to [3] for discussion of the two-sample T^2-statistic T_{n_1,n_2}^2 under the assumption that the coordinates of $\mathbf{X}_1$ and $\mathbf{Y}_1$ are normal random variables.

§7 Recent development

7.1 Studentized two-sample U-statistics

Two-sample U-statistics, typified by the two-sample Mann-Whitney test statistic, have been widely used in a broad range of scientific research. For example, they are commonly used to compare the different (treatment) effects of two groups, such as an experimental group and a control group, in scientifically controlled experiments. Unfortunately, many of these applications rely on a misunderstanding of what is being tested and the implicit underlying assumptions, which were not explicitly considered until relatively recently by Chung and Romano [23]. More importantly, these authors provided evidence for the advantage of using the Studentized statistics both theoretically and empirically. However, due to structural complexities, the theoretical properties of Studentized two-sample U-statistics are lacking in general. Recently, Chang, Shao and Zhou [14] proved a Cramér-type moderate deviation theorem in a general framework for Studentized two-sample U-statistics, with typical examples include the two-sample t-statistic and Studentized Mann-Whitney test statistic. A refined moderate deviation theorem with the second-order accuracy was established for the two-sample t-statistic under a finite fourth moment condition; see Theorem 2.4 therein. In contrast to the one-sample case, the two-sample t-statistic cannot be reduced to a self-normalized sum of independent random variables, and thus the results for self-normalized ratios [46, 74, 75] cannot be directly applied. Instead, Chang, Shao and Zhou [14] modified Theorem 2.1 in [69] to obtain a more precise expansion that could be used to derive a refined result for the two-sample t-statistic.

7.2 Self-normalized limit theory under dependence

There are many variations of asymptotic theories related to self-normalized sums in the literature, and some allow for dependent data; see, for example, [9] and [49]. We refer to [27] and [66] for a comprehensive study and a recent review. Among the existing theories, the Cramér-type moderate deviations of self-normalized sums may be the most useful for constructing simultaneous confidence sets for ultra-high dimensional statistics [34, 54]. It remains as an important open question whether Theorem 4.1 can be generalized to dependent random variables, as such a generalization is useful for ultra-high dimensional statistical inference on dependent data with heavy-tailed marginal distributions.

Recently, Chen et al. [19] showed that the general result (3) is not valid for the range of type $0 \le x \le n^{\rho}$ for any $\rho > 0$ if the dependence of the underlying process $\{X_t\}$ decays algebraically. In this case, only a much narrower range $0 \le x \le (\kappa \log n)^{1/2}$ for some constant $\kappa > 0$ is available; see Section 3 in [19]. Using block versions of V_n instead, the authors established Cramér-type moderate deviations results for self-normalized sums of weakly dependent processes with geometrically decaying dependence, under mild polynomial moment conditions. In particular, three types of self-normalized sums were introduced based on the big-block-small-block scheme, the equal-block and the interlacing scheme, respectively, and the associated Cramér-type moderate deviations were established. In the context of resampling theory for weakly dependent processes, block bootstrap procedures were proposed to adjust for dependence; see, for example, [12], [61] and [50]. However, the accuracy of a tail Gaussian approximation of type (3) has not been studied for dependent data. As shown by Chen et al. [19], due to dependency, the range of Gaussian approximation is narrower than that in the independent case, while under the same moment conditions, it is still wider than their non-Studentized counterparts. A time series two-sample moderate deviation extension was also presented. These results are useful for conducting ultra-high dimensional statistical inferences on dependent data with heavy-tailed marginal distributions, such as multiple hypothesis testing of mean vectors of ultra-high dimensional time series models in one or two samples.

The proof techniques in [19] may be used to extend the additional self-normalized limit theorems in [46], [53] and others surveyed in [66] from independent data to weakly dependent data with finite polynomial moments.

7.3 Non-normal approximation

For general non-normal approximation, Chatterjee and Shao [17] and Shao and Zhang [67] developed a concrete Stein's method to identify the limiting distribution as well as a Berry-Esseen type bound via an exchangeable pair approach. In particular, Shao and Zhang [67] proved a Berry-Esseen type bound of order $O(n^{-3/4})$ for the Curie-Weiss model at the critical temperature. Still, the self-normalized Cramér-type moderate deviation for general non-normal approximation remains open.

References

[1] T W Anderson. *An Introduction to Multivariate Statistical Analysis*, 3rd ed, Wiley, Hoboken, 2003.

[2] J N Arvesen. *Jackknifing U-statistics*, Ann Math Statist, 1969, 40(6): 2076-2100.

[3] Z Bai, H Saranadasa. *Effect of high dimension: by an example of a two sample problem*, Statist Sinica, 1996, 6(2): 311-329.

[4] A Belloni, V Chernozhukov, L Wang. (2011). *Square-root lasso: pivotal recovery of sparse signals via conic programming*, Biometrika, 2011, 98(4): 791-806.

[5] V Bentkus, M Bloznelis, F Götze. *A Berry-Esséen bound for student's statistic in the non-I.I.D. case*, J Theoret Probab, 1996, 9(3): 765-796.

[6] V Bentkus, F Götze. *The Berry-Esseen bound for Student's statistic*, Ann Probab, 1996, 24(1): 491-503.

[7] V Bentkus, B Y Jing, Q M Shao, W Zhou. *Limiting distributions of the non-central t-statistic and their applications to the power of t-tests under non-normality*, Bernoulli, 2007, 13(2): 346-364.

[8] B Bercu, E Gassiat, E Rio. *Concentration inequalities, large and moderate deviations for self-normalized empirical processes*, Ann Probab, 2002, 30(4): 1576-1604.

[9] B Bercu, A Touati. *Exponential inequalities for self-normalized martingales with applications*, Ann Appl Probab, 2008, 18(5): 1848-1869.

[10] M Bloznelis, H Putter. *Second-order and bootstrap approximation to Student's t-statistic*, Theory Probab Appl, 2003, 47(2): 300-307.

[11] J F Box. *Gosset, Fisher, and t distribution*, Amer Statist, 1981, 35(2): 61-66.

[12] P Bühlmann. *Bootstrap for time series*, Statist Sci, 2002, 17(1): 52-72.

[13] H Cao, M R Kosorok. *Simultaneous critical values for t-tests in very high dimensions*, Bernoulli, 2011, 17(1): 347-394.

[14] J Chang, Q M Shao, W X Zhou. *Cramér-type moderate deviations for Studentized two-sample U-statistics with applications*, Ann Statist, 2016, 44(5): 1931-1956.

[15] J Chang, C Y Tang, Y Wu. *Marginal empirical likelihood and sure independence feature screening*, Ann Statist, 2013, 41(4): 2123-2148.

[16] J Chang, C Y Tang, Y Wu. *Local independence feature screening for nonparametric and semi-parametric models by marginal empirical likelihood*, Ann Statist, 2016, 44(2): 515-539.

[17] S Chatterjee, Q M Shao. *Nonnormal approximation by Stein's method of exchangeable pairs with application to the Curie-Weiss model*, Ann Appl Probab, 2011, 21(2): 464-483.

[18] L H Y Chen, Q M Shao. *Normal approximation for nonlinear statistics using a concentration inequality approach*, Bernoulli, 2007, 13(2): 581-599.

[19] X Chen, Q M Shao, W B Wu, L Xu. *Self-normalized Cramér-type moderate deviations under dependence*, Ann Statist, 2016, 44(4): 1593-1617.

[20] H Chernoff. *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann Math Statist, 1952, 23(4): 493-507.

[21] G P Chistyakov, F Götze. *On bounds for moderate deviations for Student's statistic*, Theory Probab Appl, 2004, 48(3): 528-535.

[22] G P Chistyakov, F Götze. *Limit distributions of Studentized means*, Ann Probab, 2004, 32(1A): 28-77.

[23] E Chung, J Romano. *Asymptotically valid and exact permutation tests based on two-sample U-statistics*, J Statist Plann Inference, 2016, 168: 97-105.

[24] S Clarke, P Hall. *Robustness of multiple testing procedures against dependence*, Ann Statist, 2009, 37(1): 332-358.

[25] M Csörgő, L Horváth. *Asymptotic representations of self-normalized sums*, Probab Math Statist, 1988, 9: 15-24.

[26] M Csörgő, B Szyszkowicz, Q Wang. *Donsker's theorem for self-normalized partial sums processes*, Ann Probab, 2003, 31(3): 1228-1240.

[27] V H de la Peña, T L Lai, Q M Shao. *Self-Normalized Processes: Theory and Statistical Applications*, Springer, Berlin, 2009.

[28] A Delaigle, P Hall, J Jin. *Robustness and accuracy of methods for high dimensional data analysis based on Student's t-statistic*, J Roy Statist Soc Ser B, 2011, 73(3): 283-301.

[29] A Dembo, Q M Shao. *Large and moderate deviations for Hotelling's T^2-statistic*, Electron Commun Probab, 2006, 11: 149-159.

[30] B Efron. *Student's t-test under symmetry conditions*, J Amer Statist Assoc, 1969, 64: 1278-1302.

[31] V A Egorov. *Estimation of distribution tails for normalized and self-normalized sums*, J Math Sci, 2005, 127(1): 1717-1722.

[32] C Eisenhart. *On the transition from "Student's" z to "Student's" t*, Amer Statist, 1979, 33(1): 6-10.

[33] J Fan, Y Fan. *High-dimensional classification using features annealed independence rules*, Ann Statist, 2008, 36(6): 2605-2637.

[34] J Fan, P Hall, Q Yao. *To how many simultaneous hypothesis tests can normal, Student's t or bootstrap calibration be applied*, J Amer Statist Assoc, 2007, 102: 1282-1288.

[35] J Fan, J Lv. *A selective overview of variable selection in high dimensional feature space*, Statist Sinica, 2010, 20(1): 101-148.

[36] R A Fisher. *Applications of "Student's" distribution*, Metron, 1925, 5: 90-104.

[37] L Gao, Q M Shao, J S Shi. *Cramér moderate deviations for a general self-normalized sum*, Preprint, 2017.

[38] E Giné, F Götze, D M Mason. *When is the Student t-statistic asymptotically standard normal*, Ann Probab, 1997, 25(3): 1514-1531.

[39] P S Griffin, J D Kuelbs. *Self-normalized laws of the iterated logarithm*, Ann Probab, 1989, 17(4): 1571-1601.

[40] P S Griffin, D M Mason. *On the asymptotic normality of self-normalized sums*, Math Proc Cambridge Philos Soc, 1991, 109(3): 597-610.

[41] P Hall. *Edgeworth expansion for Student's t statistic under minimal moment conditions*, Ann Probab, 1987, 15(3): 920-931.

[42] P Hall. *On the effect of random norming on the rate of convergence in the central limit theorem*, Ann Probab, 1988, 16(3): 1265-1280.

[43] P Hall, Q Wang. *Exact convergence rate and leading term in central limit theorem for Student's t statistic*, Ann Probab, 2004, 32(2): 1419-1437.

[44] W Hoeffding. *A class of statistics with asymptotically normal distribution*, Ann Math Statist, 1948, 19(3): 293-325.

[45] H Hotelling. *The generalization of Student's ratio*, Ann Math Statist, 1931, 2(3): 360-378.

[46] B Y Jing, Q M Shao, Q Wang. *Self-normalized Cramér-type large deviations for independent random variables*, Ann Probab, 2003, 31(4): 2167-2215.

[47] B Y Jing, Q M Shao, W Zhou. *Saddlepoint approximation for Student's t-statistic with no moment conditions*, Ann Statist, 2004, 32(6): 2679-2711.

[48] B Y Jing, Q M Shao, W Zhou. *Towards a universal self-normalized moderate deviation*, Trans Amer Math Soc, 2008, 360: 4263-4285.

[49] M Juodis, A Račkauskas. *A remark on self-normalization for dependent random variables*, Lith Math J, 2005, 45(2): 142-151.

[50] S N Lahiri. *Resampling Methods for Dependent Data*, Springer, New York, 2003.

[51] T L Lai, Q M Shao, Q Wang. *Cramér type moderate deviations for Studentized U-statistics*, ESAIM Probab Stat, 2011, 15: 168-179.

[52] J V Linnik. *On the probability of large deviations for the sums of independent variables*, Proc 4th Berkeley Sympos Math Statist and Prob, Vol II, Univ California Press, Berkeley, 1961, 289-306.

[53] W Liu, Q M Shao. *A Cramér type moderate deviation theorem for Hotelling's T^2-statistic with applications to global tests*, Ann Statist, 2013, 41(1): 296-322.

[54] W Liu, Q M Shao. *Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control*, Ann Statist, 2014, 42(5): 2003-2025.

[55] B F Logan, C L Mallow, S O Rice, L A Shepp. *Limit distributions of self-normalized sums*, Ann Probab, 1973, 1(5): 788-809.

[56] R A Maller. *A theorem on products of random variables with application to regression*, Aust N Z J Stat, 1981, 23(2): 177-185.

[57] D M Mason. *The asymptotic distribution of self-normalized triangular arrays*, J Theoret Probab, 2005, 18(4): 853-870.

[58] S Y Novak. *On self-normalized sums of random variables and the Student's statistic*, Theory Probab Appl, 2005, 49(2): 336-344.

[59] G M Pan, W Zhou. *Central limit theorem for Hotelling's T^2 statistic under large dimension*, Ann Appl Probab, 2011, 21(5): 1860-1910.

[60] E S Pearson. *Some reflections on continuity in the development of mathematical statistics, 1885–1920*, Biometrika, 1967, 54: 341-355.

[61] D N Politis, J P Romano, M Wolf. *Subsampling*, Springer, New York, 1999.

[62] J Robinson, Q Wang. *On the self-normalized Cramér-type large deviation*, J Theoret Probab, 2005, 18(4): 891-909.

[63] Q M Shao. *Self-normalized large deviations*, Ann Probab, 1997, 25(1): 285-328.

[64] Q M Shao. *A Cramér type large deviation result for Student's t statistic*, J Theoret Probab, 1999, 12(2): 385-398.

[65] Q M Shao. *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 University Press, Singapore, 2005, 143-155.

[66] Q M Shao, Q Wang. *Self-normalized limit theorems: a survey*, Probab Surv, 2013, 10: 69-93.

[67] Q M Shao, Z S Zhang. *Identifying the limiting distribution by a general approach of Stein's method*, Sci China Math, 2016, 59(12): 2379-2392.

[68] Q M Shao, K Zhang, W X Zhou. *Stein's method for nonlinear statistics: a brief survey and recent progress*, J Statist Plann Inference, 2016, 168: 68-89.

[69] Q M Shao, W X Zhou. *Cramér type moderate deviation theorems for self-normalized processes*, Bernoulli, 2016, 22(4): 2029-2079.

[70] V V Slavova. *On the Berry-Esseen bound for Student's statistics*, In: *Stability Problems for Stochastic Models*, Lecture Notes in Math, 1155, Springer, Berlin, 1985, 355-390.

[71] Student. *The probable error of a mean*, Biometrika, 1908, 6: 1-25.

[72] M Vandemaele, N Veraverbeke. *Cramér type large deviations for Studentized U-statistics*, Metrika, 1985, 32(1): 165-179.

[73] Q Wang. *Bernstein type inequalities for degenerate U-statistics with applications*, Chin Ann Math, 1998, 19(2): 157-166.

[74] Q Wang. *Limit theorems for self-normalized large deviation*, Electron J Probab, 2005, 10: 1260-1285.

[75] Q Wang. *Refined self-normalized large deviations for independent random variables*, J Theoret Probab, 2011, 24(2): 307-329.

[76] Q Wang, B Y Jing. *An exponential nonuniform Berry-Esseen bound for self-normalized sums*, Ann Probab, 1999, 27(4): 2068-2088.

[77] Q Wang, B Y Jing, L Zhao. *The Berry-Esseen bound for Studentized statistics*, Ann Probab, 2000, 28(1): 511-535.

[78] S L Zabell. *On Student's 1908 article "The probable error of a mean."*, J Amer Statist Assoc, 2008, 103: 1-7.

[79] W Zhou, B Y Jing. *Tail probability approximations for Student's t-statistics*, Probab Theory Related Fields, 2006, 136(4): 541-559.

[1] Department of Statistics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong.
Email: qmshao@cuhk.edu.hk

[2] Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, USA.
Email: wez243@ucsd.edu