RELATIVE ERRORS IN CENTRAL LIMIT THEOREMS FOR STUDENT'S t STATISTIC, WITH APPLICATIONS

Qiying Wang and Peter Hall

University of Sydney and University of Melbourne

Abstract: Student's t statistic is frequently used in practice to test hypotheses about means. Today, in fields such as genomics, tens of thousands of t-tests are implemented simultaneously, one for each component of a long data vector. The distributions from which the t statistics are computed are almost invariably non-normal and skew, and the sample sizes are relatively small, typically about one thousand times smaller than the number of tests. Therefore, theoretical investigations of the accuracy of the tests would be based on large-deviation expansions. Recent research has shown that in this setting, unlike classical contexts, weak dependence among vector components is often not a problem; independence can safely be assumed when the significance level is very small, provided dependence among the test statistics is short range. However, conventional large-deviation results provide information only about the accuracy of normal and Student's t approximations under the null hypothesis. Power properties, especially against sparse local alternatives, require more general expansions where the data no longer have zero mean, and in fact where the mean can depend on both sample size and the number of tests. In this paper we derive this type of expansion, and show how it can be used to draw statistical conclusions about the effectiveness of many simultaneous t-tests. Similar arguments can be used to derive properties of classifiers based on high-dimensional data.

Key words and phrases: Genomics, classification, family-wise error rate, large-deviation expansion, signal detection, simultaneous hypothesis testing, sparsity.

1. Introduction and Summary

1.1. Statistical motivation.

Student's t statistic, computed from a dataset $X_1, \ldots, X_n$ drawn from the distribution of X, has the form

$$T = \frac{(S_n - n\mu)}{\sigma_n}, \qquad (1.1)$$

where

$$S_n = \sum_{i=1}^{n} X_i, \quad \sigma_n^2 = \frac{n-1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2, \quad \bar{X} = \frac{S_n}{n}. \qquad (1.2)$$

A variety of contemporary problems in statistics involve large deviations of T. They include multiple hypothesis testing and classification, and, either directly or implicitly, sparse signal detection. In the first two of these problems, data analysis can involve making tens of thousands of decisions, each based on a version of T computed from as few as tens of data values. There are strong connections between the methodology used to solve the first two problems and that employed to solve the third.

The fact that the number of decisions made is many times larger than the number of data on which each decision is based, typically means that the theoretical background to methodology is built on large-deviation theory. For example, if in a multiple hypothesis testing problem the test statistics can fairly be viewed as independent; if a Normal or Student's t approximation is made to the distribution of the test statistic, despite the sampling distribution having nonzero skewness; and if there are p tests, each founded on a t-statistic computed from a sample of size n; then, in order that it be possible to hold the family-wise error rate (FWER) at a given level, say 5%, it is necessary and sufficient that $\log p = o(n^{1/3})$ as n and p increase. See Hall (2006). The result is derived from large-deviation expansions of Shao (1999), Jing, Shao and Wang (2003) and Wang (2005). It continues to hold in the case of weak dependence among the test statistics, for example if the $n \times p$ data array has independent columns but rows that are computed from time-series; and it is valid too if FWER is replaced by false discovery rate (FDR). See Clarke and Hall (2007).

Results such as these provide crucial information about the robustness of FWER and FDR against widely-used approximations based on Normal or Student's t distributions. However, they offer no insight into how the approximations affect the power of multiple hypothesis tests, or the sensitivity of classifiers. In particular, the arguments needed to address FWER and FDR are all undertaken in the case $\mu = E(X)$ in (1.1), and so require only conventional large-deviation results; but the calculations needed to elucidate power and sensitivity properties have $\mu = E(X) + \delta_n$, where δ_n is a perturbation that generally depends on sample size. In this paper we derive large-deviation results in the perturbation setting, and apply them to multiple hypothesis testing and classification problems.

1.2. Probabilistic background

Motivated partly by contemporary statistical applications, the last decade has seen a very substantial increase in work on properties of self-normalised sums. A portion of it has focused on weak convergence, and includes deriving conditions under which convergence to normality occurs. In particular, Giné, Götze and Mason (1997) proved that when $\mu = E(X)$ the statistic T, at (1.1), converges in distribution to a standard normal random variable if and only if

the distribution of X is in the domain of attraction of the normal law, written $X \in \text{DAN}$. This attractive result verified part of a conjecture of Logan, Mallows, Rice and Sheep (1973) and was later extended to self-normalised sum processes by Csörhő, Szyszkowicz and Wang (2003, 2004). Chistyakov and Götze (2004) recently established the remainder of the conjecture by showing that Student's t statistic has a non-trivial limit distribution if and only if X is in the domain of attraction of a stable law with exponent $\gamma \in (0, 2]$. Mason (2005) considered the asymptotic distribution of self-normalised triangular arrays. Bentkus, Jing, Shao and Zhou (2007) investigated the asymptotic behaviors of the non-central t-statistic and their applications to the power of t-tests.

The absolute error in the central limit theorem for Student's t statistic was obtained by Bentkus and Götze (1996), who gave bounds of general Berry-Esseen type when the data are independent and identically distributed. See also Chibisov (1980, 1984) and Slavov (1985). Bentkus, Bloznelis and Götze (1996) extended Bentkus and Götze's arguments to non-identically distributed summands. Using a leading-term approach, Hall and Wang (2004) derived the exact convergence rate in the central limit theorem, up to terms of order $n^{-1/2}$ (where n denotes sample size), or order n^{-1} when the sampled distribution satisfied Cramér's continuity condition. (Hall (1987) had previously established Edgeworth expansions under minimal moment conditions.) Robinson and Wang (2005) considered exponential, non-uniform Berry-Esseen bounds under only $X \in \text{DAN}$, improving an earlier result of Wang and Jing (1999). Saddle-point and large-deviation approximations for Student's t statistic have been given by Shao (1997, 1999), Jing et al. (2003), Jing, Shao and Zhou (2004), Wang (2005), Robinson and Wang (2005) and Zhou and Jing (2006). Related research includes that of van Zwet (1984), Friedrich (1989), Putter and VAN Zwet (1998), Bentkus, Götze and VAN Zwet (1997), Wang, Jing and Zhao (2000) and Bloznelis and Putter (1998, 2002).

1.3. Summary of main results

Assume that

$$E|X|^3 < \infty\,, \quad E\big(X^2\big) = \sigma^2 \neq 0\,, \quad E(X) = 0\,. \tag{1.3}$$

In place of T at (1.1), define $U = (S_n + c\,\sigma)/\sigma_n$, where S_n and σ_n are as at (1.2) and c may depend on n. Our aim is to describe the relative error in a normal approximation to the distribution of U; that is, to investigate the ratio $P(U > x)/\{1 - \Phi(u)\}$, where Φ denotes the standard normal distribution function and $u = u(c, n, x)$ is a real number chosen to ensure that the ratio is close to 1 in some sense.

We shall show that an appropriate choice of u is $u = 2\,\gamma\,x$, where

$$\gamma = \tfrac{1}{2}\left\{1 - c\,(x\sqrt{n})^{-1}\right\}\,. \tag{1.4}$$

In this case, $P(U > x)/\{1 - \Phi(u)\}$ equals, to first order,

$$\Psi_{n,\gamma}(x) = \exp\left\{\gamma^2\,(\frac{4\gamma}{3} - 2)\,x^3\,\frac{E(X^3)}{\sigma^3\sqrt{n}}\right\}. \tag{1.5}$$

2. Main Results

Put $\rho_n = E|X|^3/(\sqrt{n}\,\sigma^3)$, $\tau = \sigma/(1+x)$ and

$$\Delta_{n,x} = n^{-\frac{1}{2}}\,\tau^{-3}\,E|X|^3\,I_{\{|X|>\sqrt{n}\,\tau\}} + n^{-1}\,\tau^{-4}\,E|X|^4\,I_{\{|X|\leq\sqrt{n}\,\tau\}}\,.$$

Define γ and $\Psi_{n,\gamma}$ by (1.4) and (1.5).

Theorem 1. *There exists an absolute constant $A_0 > 0$ such that, if (1.3) holds,*

$$\frac{P(S_n + c\,\sigma \geq x\,\sigma_n)}{1 - \Phi(2\,\gamma\,x)} = \Psi_{n,\gamma}(x)\,\exp(O_1\,\Delta_{n,x})\,\{1 + O_2\,(1+x)\,\rho_n\}, \tag{2.1}$$

uniformly in $|c| \leq x\,\sqrt{n}/5$ and $0 \leq x \leq \rho_n^{-1}/A_0$, where O_1 and O_2 are bounded by an absolute constant.

Theorem 1 is closely related to a Cramér-type large-deviation property for the so-called self-normalised sum $(S_n + c\,\sigma)/V_n$, where $V_n^2 = \sum_{j\leq n} X_j^2$. In fact, the theorem is implied by the following result.

Theorem 2. *There exists an absolute constant $A_0 > 0$ such that, if (1.3) holds,*

$$\frac{P(S_n + c\,\sigma \geq x\,V_n)}{1 - \Phi(2\,\gamma\,x)} = \Psi_{n,\gamma}(x)\,\exp(O_1\,\Delta_{n,x})\,\{1 + O_2\,(1+x)\,\rho_n\}, \tag{2.2}$$

uniformly in $|c| \leq x\,\sqrt{n}/5$ and $0 \leq x \leq \rho_n^{-1}/A_0$, where O_1 and O_2 are bounded by an absolute constant.

3. Statistical Implications

To appreciate the implications of (2.1), consider first the following, relatively conventional large-deviation expansion of the distribution of T, derivable from results of Wang (2005): if (1.3) holds,

$$\frac{P(T > x)}{1 - \Phi(x)} = \Psi_n(x)\,\exp(O_1\,\Delta_{n,x})\,\{1 + O_2\,(1+x)\,\rho_n\}, \tag{3.1}$$

uniformly in $0 \leq x \leq \rho_n^{-1/2}/A_0$, where $\Psi_n(x) = \exp\{-(x^3/3)\,EX^3/\sqrt{n}\}$ and the functions O_1 and O_2 are bounded by an absolute constant. (It is assumed throughout this section that $\sigma = 1$.) We shall show how (3.1) can be used to determine a critical point, x_p say, for each of p simultaneous tests of the null

hypothesis that $\mu = 0$, against the alternative $\mu > 0$, when the respective tests are based on p independent samples of size n drawn from populations having potentially different values of μ.

Provided $E|X|^3 < \infty$ and

$$x = o\!\left(n^{\frac{1}{6}}\right), \tag{3.2}$$

it follows directly from (3.1) that the normal approximation to the distribution of T is accurate in relative terms:

$$\frac{P(T > x)}{1 - \Phi(x)} \to 1 \tag{3.3}$$

as $n \to \infty$. Moreover, (3.2) is necessary for (3.3) if $E|X|^3 < \infty$ and $E(X^3) \neq 0$. Therefore, if we wish to simultaneously test p hypotheses that $\mu = 0$, each for a sample of size n, then a family-wise error rate of approximately α can be achieved by taking the critical point x_p for each test to be the solution of $\Phi(x_p)^p = 1 - \alpha$. Since (3.2) is necessary and sufficient for (3.3), then the assertion that the family-wise error rate of the simultaneous test equals α in the limit as n and p increase, is correct if and only if

$$\log p = o\!\left(n^{\frac{1}{3}}\right). \tag{3.4}$$

A similar argument, but based on (2.1) rather than (3.1), can be used to determine the power of such tests against local alternatives. For simplicity we assume that the n data in the sample drawn from the ith population, and from which the ith t statistic T_i is computed, are all independent and distributed as $X + \mu_i$, where the distribution of X has zero mean and does not depend on i. (The case where both n and the distribution of X depend on i can also be treated using the argument below.) Student's t test rejects the hypothesis H_{0i} that $\mu_i = 0$, in favour of H_{1i} that $\mu_{1i} > 0$, if $T_i > x_p$. Define $\gamma_i = \frac{1}{2}\left(1 - n^{1/2}\mu_i x_p^{-1}\right)$. Provided (3.4) holds or, equivalently, assuming that (3.2) is true for $x = x_p$; and supposing too that $\mu_i = \nu_i/(x_p\sqrt{n})$, where the nonnegative constants ν_i are bounded above by $x_p^2/5 \sim (2/5)\log p$ and satisfy

$$\frac{1}{p}\sum_{i=1}^{p}\left(1 - \nu_i x_p^{-2}\right)^{-1}\exp\left(\nu_i - \tfrac{1}{2}\nu_i^2 x_p^{-2}\right) \to \rho \in [0, \infty] \tag{3.5}$$

as $p \to \infty$; it follows from (2.1) that the probability that H_{0i} is rejected for at least one values of i equals

$$1 - \prod_{i=1}^{p}\Phi(2\gamma_i x_p) + o(1) \to \pi(\alpha) \equiv 1 - (1 - \alpha)^\rho. \tag{3.6}$$

Result (3.6) encompasses three cases of special interest: (i) $\pi(\alpha) = \alpha$ (i.e., $\rho = 1$), which amounts to the alternative hypotheses H_{1i} not being detectable using the family-wise error rate approach, because nonzero values of μ_i are either too sparse or too small to be detectable; (ii) $\alpha < \pi(\alpha) < 1$ (i.e., $1 < \rho < \infty$), in which case the tradeoff between sparsity and size is at a critical point where the presence of nonzero μ_i's is just noticeable; and (iii) $\pi(\alpha) = 1$ (i.e., $\rho = \infty$), where the presence of nonzero μ_i's is relatively obvious. To appreciate the different circumstances that lead to cases (i), (ii) and (iii), suppose that there are just $q = C_1\,p^\xi$ nonzero values of ν_i, each of them equal to $\eta \log p$, where $\xi, \eta \geq 0$ and $C_1 > 0$ is a constant. Decreasing ξ and increasing η correspond to increasing sparsity and increasing the size, respectively, of departures from the null hypothesis. In the case of this parametrisation the series on the left-hand side in (3.5) is essentially proportional to $C_2\,p^{\xi+\eta-(\eta^2/4)-1} + 1 - C_1\,p^{\xi-1}$, where $C_2 = (1 - \frac{1}{2}\eta)^{-1}\,C_1$, and so cases (i), (ii) and (iii) correspond to $\xi + \eta - \eta^2/4 < 1$, $\xi + \eta - \eta^2/4 = 1$ and $\xi + \eta - \eta^2/4 > 1$, respectively.

Results of Clarke and Hall (2007) show that these results remain valid in the case of short-range dependence among test statistics, for example for a moving average of finite order. They remain valid too if a Student's t approximation, with $n-1$ degrees of freedom, is used in place of a normal approximation. This follows from the fact that the Student's t approximation to the normal distribution enjoys greater accuracy than the normal approximation to the distribution of T when skewness is nonzero.

Similar arguments, again based on Theorem 1, can be used in the context of higher-criticism approaches to (a) signal detection, and (b) classification (see e.g., Donoho and Jin (2004) and Delaigle and Hall (2007)). In particular, the distribution of the signal in case (a), and the marginal distributions of data components in case (b), can be permitted to have the distribution of a Student's t statistic computed from non-normal data. In the context of (b) the main requirements are that marginal distributions have uniformly bounded moments of order $4 + \epsilon$ for some $\epsilon > 0$, and that the length, p, of the vector satisfy a uniform version of (3.4): $\log p = o(\min_{1 \leq i \leq p} n_i^{1/3})$, where n_i denotes the size of the sample used to construct the t statistic corresponding to the ith vector component. This assumption is realistic. For example, in genomic problems the ith vector component is generally an empirical measure of the extent to which the ith gene is "switched-on," and can be a t statistic.

4. Proofs of Theorems

We assume, throughout the proofs in Section 4 and without loss of generality, that $\sigma^2 = EX^2 = 1$ and $A_0 \geq 256$. We denote by $A, A_1, A_2, \ldots, C_0, C_1, C_2, \ldots$ absolute positive constants, which may be different at each appearance. We only

give the proof of Theorem 2 since Theorem 1 is essentially a corollary of that result. The detailed proof of Theorem 1 may be found in the on-line supplement of this paper.

Proof of Theorem 2. Write $Y_j = X_j I_{|X_j| \leq \sqrt{n}\tau}$, where $\tau = 1/(1+x)$. Then,

$$P(S_n + c \geq x\, V_n) = p_{1n} + p_{2n}, \tag{4.1}$$

where $p_{1n} = P(S_n + c \geq x\, V_n,\ X_j = Y_j,\ \text{all } j = 1, \ldots, n)$ and

$$p_{2n} = P(S_n + c \geq x\, V_n,\ X_j \neq Y_j,\ \text{some } j = 1, \ldots, n).$$

First we treat the case $0 \leq x \leq 16$, where

$$p_{2n} \leq n\, P(|X| \geq \sqrt{n}\tau) \leq A\, \rho_n. \tag{4.2}$$

Since $|(1+y)^{1/2} - (1 + y/2)| \leq y^2$ then, for $y \geq -1$,

$$p_{1n} = P\left[\sum_{j=1}^{n} Y_j + c \geq x\, B_n \left\{ 1 + B_n^{-2} \sum_{j=1}^{n} \left(Y_j^2 - EY_j^2 \right) \right\}^{\frac{1}{2}} \right]$$

$$= P\left\{ \sum_{j=1}^{n} (Z_j - EZ_j) + \frac{x\theta}{B_n^3} \sum_{1 \leq i \neq j \leq n} W_i W_j \geq x B_n - c - n\, EZ_1 \right\}, \tag{4.3}$$

where $|\theta| \leq 1$, $W_j = Y_j^2 - EY_j^2$, $B_n^2 = EY_1^2$ and

$$Z_j = Y_j - \frac{x}{2B_n} W_j + \frac{x\theta}{B_n^3} W_j^2.$$

Simple calculations, using the fact that $EX^2 = 1$, show that $B_n^2 = (1 + O_1 n^{-1/2} E|X|^3)$, $\mathrm{var}(Z_1) = 1 + O_2\, n^{-1/2} E|X|^3$ and $n\, EZ_1 = O_3$, for $0 \leq x \leq 16$, where O_1, O_2 and O_3 are bounded by an absolute constant. We also have $E|Z_1|^3 \leq A\, E|X|^3$ and $E|W_1 W_2 / B_n|^{5/3} \leq A\, E|X|^3$. By virtue of these facts it follows from Theorem 2.1 of Wang et al. (2000) with minor modifications that, for $0 \leq x \leq 16$,

$$\left| p_{1n} - \{1 - \Phi(x - \frac{c}{\sqrt{n}})\} \right| \leq A\, \rho_n. \tag{4.4}$$

This, together with (4.1)–(4.2), implies (2.2) for $0 \leq x \leq 16$ and $|c| \leq x\, \sqrt{n}/5$.

Next we treat the case $x \geq 16$. First note that, whenever $|t - t_0| \leq A\, (s + \Delta_{n,x})/x^2$ where $1/3 \leq t_0 \leq 2/3$ and $|s| \leq A$, we have

$$1 - \Phi(2tx) \leq \{1 - \Phi(2t_0 x)\} \exp\{A\, (\Delta_{n,x} + s)\}, \tag{4.5}$$

$$x^{-1} \exp\left(-2t_0^2 x^2 \right) \leq 2\sqrt{2\pi}\, \{1 - \Phi(2t_0 x)\}, \tag{4.6}$$

$$\Psi_{n,t}(x) \leq \Psi_{n,t_0}(x) \exp\{A\, (\Delta_{n,x} + s)\}, \tag{4.7}$$

for $2 \leq x \leq \rho_n^{-1}/A$. Indeed, (4.7) is obvious and (4.5)–(4.6) follow from the inequality (see, for example, Revuz and Yor (1990, p.30)): for $y > 0$,

$$\frac{y}{\sqrt{2\pi}\,(1+y^2)}\,e^{-\frac{y^2}{2}} \leq 1 - \Phi(y) \leq \frac{1}{\sqrt{2\pi}\,y}\,e^{-\frac{y^2}{2}}\,. \qquad (4.8)$$

We also need following three propositions, the proofs of which will be given in the on-line supplement of this paper.

Proposition 1. *Assume (1.3) holds, take $h = x/\sqrt{n}$ and let θ be bounded by an absolute constant, A. Then there exists an absolute constant $A_0 > 0$ such that, for $4 \leq x \leq \rho_n^{-1}/A_0$ and $|\delta_{1n}| \leq x^2/2$,*

$$\frac{P\big(2hS_n - h^2V_n^2 + \theta h^4 Q_n \geq x^2 + \delta_{1n}\big)}{1 - \Phi(2\,\lambda_1\,x)} = \Psi_{n,\lambda_1}(x)\exp(O_1\Delta_{n,x})(1 + O_2 x\rho_n),(4.9)$$

where $Q_n = \sum_j X_j^4\,I_{|X_j| \leq \sqrt{n}\tau}$, $\lambda_1 = \{1+\delta_{1n}/(2x^2)\}/2$ and O_1 and O_2 are bounded by an absolute constant.

Proposition 2. *If (1.3) holds then there exists an absolute constant $A_0 > 0$ such that, for $4 \leq x \leq \rho_n^{-1}/A_0$ and $|\delta_{2n}| \leq x\sqrt{n}/4$,*

$$P(S_n \geq x\,V_n + \delta_{2n}) \leq \{1 - \Phi(2\lambda_2\,x)\}\,\Psi_{n,\lambda_2}(x)\,\exp\{A\,(\Delta_{n,x} + 1)\}\,, \quad (4.10)$$

where $\lambda_2 = \{1 + \delta_{2n}/(x\sqrt{n})\}/2$. Moreover, for $5 \leq x \leq \rho_n^{-1}/A_0$ and $|\delta_{2n}| \leq x\sqrt{n}/4$,

$$P\big\{S_{n-1} \geq (x^2 - 1)^{\frac{1}{2}}\,V_{n-1} + \delta_{2n}\big\}$$
$$\leq \{1 - \Phi(2\lambda_2\,x)\}\,\Psi_{n,\lambda_2}(x)\,\exp\{A\,(\Delta_{n,x} + 1)\}\,. \quad (4.11)$$

Let ς_n and ψ_n be functions of two and three variables, respectively, and put $T_m = n^{-1/2}\sum_{j \leq m}\varsigma_j$ and $\Lambda_{n,m} = n^{-2}\sum_{1 \leq k \leq m-1}\sum_{k+1 \leq j \leq n}\psi_{k,j}$, where $\varsigma_j = \varsigma_n(x, X_j)$ and $\psi_{k,j} = \psi_n(x, X_k, X_j)$.

Proposition 3. *If the following conditions are satisfied:*

$$\left|E\varsigma_j^2 - 1 + \frac{x}{\sqrt{n}}EX^3\right| \leq C_1\,\frac{\Delta_{n,x}}{(1+x)^2}\,, \qquad (4.12)$$

$$\left|E\varsigma_j^3 - EX^3\right| \leq C_2\,\sqrt{n}\,\frac{\Delta_{n,x}}{(1+x)^3}\,, \qquad (4.13)$$

$$E\varsigma_j \;=\; 0\,, \quad |\varsigma_j| \leq C_3\,\frac{\sqrt{n}}{(1+x)}\,, \quad E\varsigma_j^4 \;\leq\; C_4\,EX^4\,I_{|X| \leq \sqrt{n}\tau}\,, \qquad (4.14)$$

$$E(\psi_{k,j}\,|\,X_k) = E(\psi_{k,j}\,|\,X_j) = 0\,, \quad \text{for } k \neq j\,, \qquad (4.15)$$

$$E|\psi_{k,j}| \leq C_5\,|x|\,, \quad E|\psi_{k,j}|^{\frac{3}{2}} \;\leq\; C_6\,|x|^{\frac{3}{2}}\big(E|X|^3\big)^2\,. \qquad (4.16)$$

then there exists an absolute constant $A_0 > 0$ such that, for $4 \leq x \leq \rho_n^{-1}/A_0$ and $|\delta_{3n}| \leq x/4$,

$$P\Big(T_n + \Lambda_{n,n} \geq x + \delta_{3n}\Big)$$

$$\leq \{1 - \Phi(2\lambda_3\,x)\}\,\Psi_{n,\lambda_3}(x)\,\exp(A\,\Delta_{n,x})\,(1 + A\,x\,\rho_n) + A\,(x\rho_n)^{\frac{3}{2}}\,, \quad (4.17)$$

where $\lambda_3 = (1 + \delta_{3n}/x)/2$.

We are now ready to prove (2.2). By the Cauchy-Schwarz inequality, $x\,V_n \leq (x^2 + h^2 V_n^2)/2h$, where $h = x/\sqrt{n}$. This, together with Proposition 1 with $\theta = 0$ and $\delta_{1n} = -2hc$, implies that

$$P(S_n + c \geq xV_n) \geq P\big(2\,h\,S_n - h^2\,V_n^2 \geq x^2 - 2\,h\,c\big)$$
$$= \{1 - \Phi(2\gamma x)\}\,\Psi_{n,\gamma}(x)\,\exp\big(O_1\,\Delta_{n,x}\big)\,\big(1 + O_2\,x\,\rho_n\big)\,,$$

which implies the lower bound in (2.2). In order to derive the upper bound there we continue to use the notation p_{1n} and p_{2n} defined at (4.1). Using (5.7) of Jing et al. (2003) (see also Wang and Jing (1999)), it follows from (4.11) with $\delta_{2n} = -c$ that, for $16 \leq x \leq \rho_n^{-1}/A_0$,

$$p_{2n} \leq nP(|X| \geq \sqrt{n}\tau)\,P\big\{S_{n-1} + c \geq (x^2 - 1)^{\frac{1}{2}}\,V_{n-1}\big\}$$
$$\leq A\,\Delta_{n,x}\,\{1 - \Phi(2\gamma x)\}\,\Psi_{n,\gamma}(x)\,\exp(A\,\Delta_{n,x})\,. \quad (4.18)$$

By virtue of (4.1) and (4.18), and following from the same lines as in the proof of Theorem 1.2 of Wang (2005) (see (3.16) and (3.17) there), the upper bound of (2.2) will follow if we prove that there exists an absolute constant $A_0 > 0$ such that, for $16 \leq x \leq \rho_n^{-1}/A_0$,

$$p_{1n} \leq \{1 - \Phi(2\gamma x)\}\,\Psi_{n,\gamma}(x)\,\exp(A\,\Delta_{n,x})\,\big(1 + A\,x\,\rho_n\big) + A\,e^{-3x^2}, \quad (4.19)$$

$$p_{1n} \leq \{1 - \Phi(2\gamma x)\}\,\Psi_{n,\gamma}(x)\,\exp(A\,\Delta_{n,x})\,\big(1 + A\,x\,\rho_n\big) + A\,(x\rho_n)^{\frac{3}{2}}\,, \quad (4.20)$$

where γ is as at (1.4).

We first prove (4.20). Without loss of generality, n is so large that $EX^2\,I_{|X|\geq\sqrt{n}\tau} \leq 1/32$. Let $\varsigma_j = \sqrt{n}\,(Z_j - EZ_j)/B_n$, $\psi_{k,j} = 2\,x\,\theta\,B_n^4\,W_k\,W_j/n^2$, $\delta_{3n} = -c/B_n - nEZ_1/B_n$ and $\lambda_3 = (1 + \delta_{3n}/x)/2$, where notations Z_j, W_j and B_n are as in (4.3). Tedious but simple calculations show that ς_j and $\psi_{k,j}$ satisfy (4.12)–(4.16) in Proposition 3, $|\lambda_3 - \gamma| \leq A\Delta_{n,x}/x^2$ and $|\delta_{3n}| \leq x/4$ whenever $|c| \leq x\,\sqrt{n}/5$. Hence, by (4.3), Proposition 3, and (4.5) and (4.7) with $s = 0$, we have

$$p_{1n} = P\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varsigma_j + \frac{1}{n^2}\sum_{k=1}^{n-1}\sum_{j=k+1}^{n}\psi_{k,j} \geq x + \delta_{3n}\Big)$$

$$\leq \{1 - \Phi(2\lambda_3 x)\}\,\Psi_{n,\lambda_3}(x)\,\exp(A\Delta_{n,x})\,(1 + Ax\rho_n) + A\,(x\rho_n)^{\frac{3}{2}}$$

$$\leq \{1 - \Phi(2\gamma x)\}\,\Psi_{n,\gamma}(x)\,\exp(A\Delta_{n,x})\,(1 + Ax\rho_n) + A\,(x\rho_n)^{\frac{3}{2}}\,,$$

which implies (4.20).

To prove (4.19), let $W_j = Y_j^2 - EY_j^2$, $D_n^2 = \sum_{j \leq n} EW_j^2$ and $V_n^{*2} = \sum_{j \leq n} Y_j^2$. Using $(1+y)^{1/2} \geq 1 + y/2 - y^2$ for any $y \geq -1$ and recalling $h = x/\sqrt{n}$, we have

$$
\begin{aligned}
p_{1n} &= P\Big[S_n + c \geq x\sqrt{n}\Big\{1 + \frac{V_n^2 - n}{n}\Big\}^{\frac{1}{2}}, X_j = Y_j, j = 1, \ldots, n\Big] \\
&\leq P\Big[S_n + c \geq x\sqrt{n}\Big\{1 + \frac{1}{2n}(V_n^2 - n) - \frac{1}{n^2}(V_n^2 - n)^2\Big\}, X_j = Y_j, j = 1, \ldots, n\Big] \\
&\leq P\Big[2\,h\,S_n - h^2 V_n^2 + 2\,c\,h \geq x^2\Big\{1 - \frac{(V_n^{*2} - n)^2}{n^2}\Big\}\Big] \\
&\leq P\Big[2\,h\,S_n - h^2 V_n^2 + 2\,c\,h \geq x^2\Big\{1 - \frac{2(V_n^{*2} - EV_n^{*2})^2}{n^2} - 2EX^2 I_{|X| \geq \sqrt{n}\tau}\Big\}\Big] \\
&\leq P\Big[|\sum_{k=1}^{n} W_k| \geq \sqrt{6}x\,\Big\{4D_n + (\sum_{k=1}^{n} W_k^2)^{\frac{1}{2}}\Big\}\Big] \\
&\quad + P\Big[2\,h\,S_n - h^2 V_n^2 + 2\,c\,h \geq x^2\big\{1 - 24x^2 n^{-2} \sum W_k^2 - \Delta_n^*(x)\big\}\Big] \\
&\leq 8e^{-3x^2} + P\Big[2\,h\,S_n - h^2 V_n^2 + 48\,h^4 \sum Y_k^4 \geq x^2 + \delta_{1n}(x)\Big] \\
&:= 8e^{-3x^2} + K_n(x), \tag{4.21}
\end{aligned}
$$

where $\Delta_n^*(x) = 2EX^2 I_{|X| \geq \sqrt{n}\tau} + 192x^2 n^{-1} EX^4 I_{|X| \leq \sqrt{n}\tau}$, $\delta_{1n} = -2ch + 240\,\Delta_{n,x}$, and we have used Lemma 6.4 of Jing et al. (2003) in the last inequality. Write $\lambda_1 = \{1 + \delta_{1n}/(2x^2)\}/2$. Note that $\Delta_{n,x} \leq (1+x)^3 \rho_n$ and recall that $|c| \leq x\sqrt{n}/5$. We have $|\lambda_1 - \gamma| \leq A\Delta_{n,x}/x^2$ and $|\delta_{1n}| \leq x^2/2$ for $16 \leq x \leq \rho_n^{-1}/1500$. Hence it follows easily from Proposition 1 that

$$
\begin{aligned}
K_n(x) &\leq \big\{1 - \Phi(2\lambda_1\,x)\big\}\,\Psi_{n,\lambda_1}(x)\,\exp\big(A\Delta_{n,x}\big)\big(1 + Ax\rho_n\big) \\
&\leq \big\{1 - \Phi(2\gamma\,x)\big\}\,\Psi_{n,\gamma}(x)\,\exp\big(A\Delta_{n,x}\big)\big(1 + Ax\rho_n\big).
\end{aligned}
$$

Taking this estimate into (4.21), we obtain the required (4.19). The proof of Theorem 2 is now complete.

Acknowledgement

References

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.

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

Bentkus, V., Götze, F. and VAN Zwet, W.R. (1997). An Edgeworth expansion for symmetric statistics. *Ann. Statist.* **25**, 851-896.

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.

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. (Edited by B. Grigelionis et al.), 81-98. Vilnius, Utrecht: VSP/TEV.

Bloznelis, M. and Putter, H. (2002). Second Order and Bootstrap Approximation to Student's Statistic. *Theory Probab. Appl.* **47**, 374-381.

Chibisov, D. M. (1980). Asymptotic expansion for the distribution of statistics admitting a stochastic expansion. I. *Teor. Veroyatnost. i Primenen.* **25**, 745-756.

Chibisov, D. M. (1984). Asymptotic expansions and deficiencies of tests. In: *Proceedings of the International Congress of Mathematicians* **1** (Edited by Z. Ciesielski and C. Olech), 1063-1079. PWN, Warsaw.

Chistyakov, G. P. and Götze, F. (2004). Limit distributions of Studentized means. *Ann. Probab.* **32**, 28-77.

Clarke, S. and Hall, P. (2007). Robustness of multiple testing procedures against dependence. Manuscript.

Csörhő, M., Szyszkowicz, B. and Wang, Q. (2003). Donsker's theorem for self-normalized partial sums processes. *Ann. Probab.* **31**, 1228-1240.

Csörhő, M., Szyszkowicz, B. and Wang, Q. (2004). On weighted approximations and strong limit theorems for self-normalized partial sums processes. In *Asymptotic methods in Stochastics,* 489-521. Fields Inst. Commun. **44**, Amer. Math. Soc., Providence, RI.

Delaigle, A. and Hall, P. (2007). Using threshold methods to extend higher-criticism classification to non-normal, dependent vector components. Manuscript.

Donoho, D. L. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32**, 962-994.

Friedrich, K. O. (1989). A Berry-Esseen bound for functions of independent random variables. *Ann. Statist.* **17**, 170-183.

Giné, E., Götze, F. and Mason, D. (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Probab.* **25**, 1514-1531.

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

Hall, P. (2006). Some contemporary problems in statistical science. In *Madrid Intelligencer* (Edited by F. Chamizo and A. Quirós), 38-41. Springer, New York.

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.

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.

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.

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.

Mason, D. M. (2005). The asymptotic distribution of self-normalized triangular arrys. *J. Theoret. Probab.* **18**, 853-870.

Putter, H. and VAN Zwet, H. R. (1998). Empirical Edgeworth expansions for symmetric statistics. *Ann. Statist.* **26**, 1540-1569.

Revuz, D. and Yor, M. (1999). *Continuous Martingale and Brownian Motion.* Springer.

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

Shao, Q.-M. (1997). Self-normalized large deviations. *Ann. Probab.* **25**, 285-328.

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

Slavov, V. V. (1985). On the Berry-Esseen bound for Student's statistic. In: *Stability Problems for Stochastic Models*, (Edited by V. V. Kalashnikov and V. M. Zolotarev) *Lecture Notes in Math.* **1155**, 355-390. Springer, Berlin.

VAN Zwet, W. R. (1984). A Berry-Esseen bound for symmetric statistics. *Z. Wahrsch. Verw. Gebiete* **66**, 425-440.

Wang, Q. (2005). Limit Theorems for self-normalized large deviation. *Electronic J. Probab.* **10**, 1260-1285.

Wang, Q. and Jing, B.-Y. (1999). An exponential nonuniform Berry-Esseen bound for self-normalized sums. *Ann. Probab.* **27**, 2068–2088.

Wang, Q., Jing, B.-Y. and Zhao, L. C. (2000). The Berry-Esseen bound for Studentized statistics. *Ann. Probab.* **28**, 511-535.

Zhou, W. and Jing, B.-Y. (2006). Tail probability approximations for Student's *t* statistics. *Probab. Theory Relat. Fields*, **136**, 541-559.

School of Mathematics & Statistics, The University of Sydney, NSW 2006, Australia.

E-mail: qiying@maths.usyd.edu.au

Department of Mathematics & Statistics, The University of Melbourne, VIC 3010, Australia.

E-mail: halpstat@ms.unimelb.edu.au