A Review on Approximation Approach for the Distribution of the Studentized Mean

Menus Nkurunziza
University of Burundi, Burundi

Lea Vermeire
KU Leuven Campus Kulak, Belgium

This article is distributed under the Creative Commons by-nc-nd Attribution License. Copyright © 2023 Hikari Ltd.

Abstract

In case of a normal population the studentized sample mean has the Student distribution with the sample size minus one as degrees of freedom. Simulation reports in the literature concluded that this distribution may be hold on as an acceptable approximation, and is often better than the classical normal approximation, in case of a general finite variance population provided moderate to large sample size. We provide a mathematical proof and add simulation evidence. A comparison of the performances of both approximations for the true distribution of the studentized statistic is given for the cumulative distribution function and for the quantile function, by asymptotic expansions and by simulations. A population condition such that the student approximation is universally better than the normal approximation is obtained.

Mathematics Subject Classification: 62A01

Keywords: studentized mean; t-statistic; Edgeworth expansion; normal-like distribution; gamma-like population; zero-skewness distribution

1 Introduction

Let $X$ be a population with unknown mean $\mu = E(X)$ and finite variance $\sigma^2 = V(X)$. Given an independent and identically distributed sample $X_i, i =$
1, 2, ..., \( n \) from \( X \), basic statistical inference for the population mean \( \mu \) is based on the studentized sample mean \( T = n^{1/2} (\bar{X} - \mu)/S \), where \( \bar{X} \) is the sample mean and \( S^2 \) is the usual unbiased variance estimator. Two approaches for the distribution of \( T \) populate established introductory textbooks. The widespread two-distribution solution is usually presented by sample size: for large sample size, \( T \approx Z \sim N(0, 1) \), and for small to moderate sample size and a normal or nearly normal population, \( T \approx t_{n-1} \) where \( t_{n-1} \) is Student’s \( t \)-distribution with \( n - 1 \) degrees of freedom. It is found in Barrow (2017); Johnson (2011); Lefevre (2011); Mendenhall and Sincich (1992); Rosner (2010); Wonnacott and Wonnacott (1990); Wackerly, Mendenhall and Scheaffer (2008); Massart et al. (1997) to name a few. A second approach, read in Moore, MacCabe and Craig (2017); Saporta (2006); Vining and Kowalski (2010) among others, is a one-distribution solution: \( T \approx t \) is overall allowed except for small samples from non-normal populations.

The second approach has benefits of pedagogical and accuracy nature. The simplicity of a one-distribution approach might be particularly appreciated in applied courses. As to accuracy, in case of a normal population and any sample size \( t \) is the true distribution of \( T \), while in case of a general population and a large sample, \( t \) and \( Z \) are asymptotically equal approximations for practice.

Rhiel ans Wilkie (1996) investigated the problem for statistical testing, by simulations, and suggested teaching that \( t \) critical values are used instead of \( Z \) critical values. Finner and Dickhaus (2010) give a correction on Chungs method and obtain higher order Edgeworth expansions for the cumulative distribution function of a generalized self-normalized sum with \( T \) as a special case - with respect to a standard normal and to a \( t \)-distribution as approximations, and discuss the rate of convergence for each. Lehmann and Romano (2005) considered Edgeworth expansions of the distribution function of the sample mean to study the rejection probability of the \( t \)-test when normality assumption fails.

In this paper we compare the errors in the \( Z \) and \( t \) approximations for the true distribution of \( T \), as to the cumulative distribution function (cdf) and the quantile function (qf), for small to large sample size, essentially for non-normal populations, and this by asymptotic expansions and simulations. The research is guided by three questions: what is the order of the error in each approximation; is the difference in the approximations relevant; is one of the approximations better than the other?
2 Methods

The performance of the Z and t approximations for the true cdf and the true qf of T is investigated in two ways: a theoretical investigation using asymptotic expansions of the errors in powers of \(n^{-1/2}\), and an empirical investigation using simulations. The theoretical study covers all populations under mild regularity conditions but is limited to second order errors, while the empirical study investigates the true errors but is limited to a representative selection of populations; as such the two studies are complementary. It should be noticed that T is invariant under a location-scale transform of the population \(X\), and thus, results for a population \(X\) hold for its location-scale family \(a + bX\), \(b > 0\), as well.

**Asymptotic expansions** We obtain asymptotic expansions in powers of \(n^{-1/2}\) to order two for the cdf of T with each approximation as germ – known as Edgeworth expansions – as application of the theory developed by Hall (1992). Next an inverting strategy leads to asymptotic expansions of order two for the qf of T again with each approximation as germ – known as Cornish-Fisher expansions. The following terminology will be used. Assume a valid asymptotic expansion

\[
F_n(x) = G_n(x) + n^{-1/2}g_1(x) + n^{-1}g_2(x) + \ldots + n^{-k/2}g_k(x) + o(n^{-k/2})
\]

for a function \(F_n(x)\), for which an approximation \(G_n(x)\), \(F_n(x) - G_n(x) \to 0\) as \(n \to \infty\), is available. Then the right expression is the asymptotic expansion to order \(k\) with germ \(G_n(x)\), the error of the approximation and its asymptotic expansion to order \(k\) is \(\Delta_G(x) := F_n(x) - G_n(x) = n^{-1/2}g_1(x) + n^{-1}g_2(x) + \ldots + n^{-k/2}g_k(x) + o(n^{-k/2})\), the unsigned error is its absolute value, the \(k\)-th order error of the approximation or \(k\)-th order correction on the approximation is the trimmed polynomial \(\Delta_G,k(x) := n^{-1/2}g_1(x) + n^{-1}g_2(x) + \ldots + n^{-k/2}g_k(x)\), the \(k\)-th order error term or \(k\)-th order correction term is the term in \(n^{-k/2}\). Thus the second order error/correction \(\Delta_G,2(x) = n^{-1/2}g_1(x) + n^{-1}g_2(x)\) should not be confused with the second order error/correction term \(n^{-1/2}g_1(x)\).

**Simulations** We compare the true errors of the Z and t approximations for the true distribution of T through simulations of the true distribution of T for a selection of populations and for several sample sizes. Computations are done with the software R. Besides the normal population, chosen as reference, non-normal populations are selected for their deviation from normality in skewness and/or kurtosis. The used distributions are given in Table 1 and show the following characteristics:
normal $N(0, 1)$ : control for the simulation program, the simulation outcome should match the known result $T \sim t$

uniform continuous $U[0, 1]$ : symmetric light tails

$t(5)$ : symmetric heavy tails

$B(5, 2)$ : asymmetric with positive skewness, light tails

exponential $\text{Exp}(1)$ : asymmetric with positive skewness, heavy tails

They are agents of the families shown in column 2 of Table 1. The simulations are done for increasing sample size: $n = 5, 10, 20, 50, 100$. It is known that in the limit, for $n \to \infty$, the two approximations $Z$ and $t$, and the true $T$, are all equal in distribution. For the $Z$ and $t$ distributions the essential functions like cdf and qf are available as functions in R. The true distribution of $T$, for a chosen population $X$ and chosen sample size $n$, is accurately simulated as the empirical distribution obtained from $N = 10^6$ replicate samples. This sample size guarantees, with confidence 95%, a precision of at least $10^{-3}$ for simulated probabilities on $T$. Each sample $X_{j1}, \ldots, X_{jn}$ yields statistics $\bar{X}_j, S_j^2$, and a corresponding value $T_j = n^{1/2}(\bar{X}_j - \mu)/S_j$. Replicate sampling yields $N = 10^6$ random values $T_1, \ldots, T_N$ for $T$. The distribution of $T$ is estimated as the empirical distribution of the values $T_1, \ldots, T_N$. A probability $p = P(T \in B)$, for some measurable set $B \subset \mathbb{R}$, is estimated as the empirical proportion or the relative frequency of the $N = 10^6$ observations $T_j$ that meet the event $B$, i.e. as the average success indicator

$$p \approx \frac{\#(j : T_j \in B)}{N} = \bar{I}, \quad \bar{I} = \frac{\sum_j I_j}{N}, \quad I_j = I(T_j \in B).$$

where $I(A)$ stands for the success indicator of an event $A$, with value 1 if $A$ occurs, and value 0 if not.

### 3 Edgeworth expansions for the cumulative distribution function

Consider an iid sample $X_i, i = 1, 2, \ldots, n$ from a population $X$ with with mean $\mu = \text{E}(X)$ and finite variance $\sigma^2 = \text{V}(X) = \text{E}(X - \mu)^2$. Then $\bar{X} = \sum_i X_i/n$ is the sample mean, $n^{1/2} \frac{\bar{X} - \mu}{\sigma}$ is the standardized (sample) mean, and the studentized (sample) mean is

$$T = n^{1/2} \frac{\bar{X} - \mu}{S}$$ (1)
where $S^2 = \sum_i (X_i - \bar{X})^2 / (n - 1)$ is the classical unbiased estimator of the variance. Mathematical statistics provides basic theorems for the distribution of the statistic $T$. If $X$ is normal $N(\mu, \sigma^2)$, the exact distribution is $T \sim t_{n-1}$. If $X$ is general with finite variance, then $T \overset{d}{\to} Z \sim N(0, 1)$ if $n \to \infty$, as a corollary of the central-limit theorem and Slutsky’s theorem. Applying this for normal $X$ provides $t \overset{d}{\to} Z$. Hence $Z$ and $t$ provide two asymptotic approximations for the distribution of $T$. Asymptotic expansions for their errors on the cdf are now given.

**Theorem 3.1.** Edgeworth expansions for the cumulative distribution function of the studentized mean $T$. The Edgeworth expansions of the cdf $F_T$ with respectively the standard normal cdf $\Phi(x) = F_Z(x) = P(Z \leq x)$ and Student’s $t \overset{d}{\to} Z_{n-1}$ cdf $F_t(x) = P(t \leq x)$ as germ, are given by

\[
F_T(x) = \Phi(x) + n^{-1/2} p_1(x) \varphi(x) + n^{-1} [p_2(x) + p(x)] \varphi(x) + o(n^{-1}) \tag{2}
\]

\[
F_T(x) = F_t(x) + n^{-1/2} p_1(x) \varphi(x) + n^{-1} p_2(x) \varphi(x) + o(n^{-1}) \tag{3}
\]

where

\[
p_1(x) = \frac{1}{6} \gamma (2x^2 + 1) \tag{4a}
\]

\[
p_2(x) = \frac{1}{12} \kappa (x^3 - 3x) - \frac{1}{18} \gamma^2 (2x^3 + 3x^2 - 3x) \tag{4b}
\]

\[
p(x) = \frac{1}{4} (x^3 + x) \tag{4c}
\]

$\gamma = E[(X - \mu) / \sigma^3]$ is the skewness and $\kappa = E[(X - \mu) / \sigma^4] - 3$ is the kurtosis of the population $X$. These asymptotic expansions are valid under the sufficient
conditions: $X$ is absolute continuous – in particular if $X$ has a proper density function – and the fourth moment $E(|X|^4) < \infty$.

**Proof.** An extensive derivation of the Edgeworth expansion, under minimal moment conditions for its validity, for the cdf of a studentized mean with $n$-weight sample variance, $T = n^{1/2}(\bar{X} - \mu)/S$, $\bar{S}^2 = \sum (X_i - \bar{X})^2/n$, is given by Hall (1992). Next to the general structure to order $k$, an explicit expression for the expansion to order 2 is given:

$$F_{\tilde{T}}(x) = \Phi(x) + n^{-1/2} \tilde{p}_1(x) \varphi(x) + n^{-1/2} \tilde{p}_2(x) + o(n^{-1})$$ (5)

where

$$\tilde{p}_1(x) = \frac{1}{6} \gamma (2x^2 + 1)$$ (6a)

$$\tilde{p}_2(x) = \frac{1}{12} \kappa (x^3 - 3x) - \frac{1}{18} \gamma^2 (x^5 + 2x^3 - 3x)$$ (6b)

$$\tilde{p}(x) = -\frac{1}{4} (x^3 + 3x)$$ (6c)

and the sufficient conditions for validity are those reproduced in the theorem. The studentized mean under consideration is an $n$-dependent scale transform of $\bar{T}$:

$$T = (1 - n^{-1/2} \delta x) \tilde{T}. Using the binomial series and Taylor expansion, it follows that

$$F_T(x) = \Phi(x) + n^{-1} p(x) \varphi(x) + o(n^{-1}) = \Phi(x) - \frac{1}{4} (x^3 + x) \varphi(x) n^{-1} + o(n^{-1})$$ (8)

We may notice that this is also the Edgeworth expansion to order two for the cdf of Student’s $t_n$. Finally, $F_T(x) - F_t(x)$, using (2) and (8), yields the second expansion (3) in the theorem.

4 Second order errors in the cdf approximations

4.1 Source of the errors

The errors in the normal and Student approximations for the cdf of $T$ are

$$\Delta_Z(x) = F_T(x) - \Phi(x) \quad \text{and} \quad \Delta_t(x) = F_T(x) - F_t(x),$$ (1)
with asymptotic expansions given in Theorem 3.1. They depend on population characteristics like $\gamma$ and $\kappa$, next to the sample size $n$, and the function argument $x$.

**First order error** The errors for the two approximations in (2) and (3) show as lowest order error

$$\Delta_{Z,1}(x) = \Delta_{t,1}(x) = \frac{1}{6} n^{-1/2} \gamma (2x^2 + 1) \varphi(x),$$

the same for both, a term in $n^{-1/2}$, and it depends on the population only through $\gamma$. Hence, for the cdf of the studentized mean $T$, both approximations $Z$ and $t$ are first order equivalent and their joint first order error is of order $n^{-1/2}$, also called order 1; they differ at most at order $n^{-1}$. The dominant population reason for error in the approximations is skewness in the population. This is no surprise as, for given $n$, the central limit effect for the sample mean suffers most from skewness. For given skewness, the unsigned first order error reaches its maximum at $x = \pm \sqrt{3/2} \approx \pm 1.225$, and is then $0.1256 n^{-1/2} \gamma \approx 1/8 n^{-1/2} |\gamma|$. It will be at most $10^{-2}$ provided $|\gamma| \leq 8\sqrt{n}/100$. Thus if the first order error should be visible typically at most from the third decimal on, and the sample size is $n \geq 25$, then the skewness should be bounded by $|\gamma| \leq 0.40$. Conversely, there is indication that neither of the two approximations, $Z$ or $t$, should be used in the case of small to moderate samples from a skewed population.

**Second order error** The second order errors/corrections on the $Z$ and $t$ approximations for the cdf of $T$ are

$$\Delta_{Z,2}(x) = n^{-1/2} p_1(x) \varphi(x) + n^{-1} [p_2(x) + p(x)] \varphi(x)$$

$$\Delta_{t,2}(x) = n^{-1/2} p_1(x) \varphi(x) + n^{-1} p_2(x) \varphi(x)$$

with the polynomials $p_1, p_2, p$ given above. Thus, the population impact on the second order error is the same for both approximations, and is due to skewness from the first order term on, and kurtosis in the second order term. The difference in the second order errors depends on $x$ and $n$ only.

**4.2 Is the difference in the approximations relevant**

The difference in the two approximations, using (8), may be expressed as :

$$F_t(x) - \Phi(x) = n^{-1} p(x) \varphi(x) + o(n^{-1}), \quad p(x) = -\frac{1}{4} (x^3 + x).$$
It is $O(n^{-1})$ and thus of order 2, which explains why a difference in the two Edgeworth expansions for $F_T$ will be visible at earliest at order 2. The maximum unsigned leading term of difference is reached at two symmetric points $x = \pm \sqrt{1 + \sqrt{2}} \approx \pm 1.554$, and equals $0.158n^{-1}$. Thus, from sample size $n > 15$ on, the second order difference will be irrelevant in the sense of smaller than $10^{-2}$, and thus visible typically at most in the third decimal. Table 2 shows the region for $x$ where the difference in the second order approximations is irrelevant.

<table>
<thead>
<tr>
<th>$n$</th>
<th>region of irrelevant difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$</td>
</tr>
<tr>
<td>10</td>
<td>$</td>
</tr>
<tr>
<td>15</td>
<td>$</td>
</tr>
<tr>
<td>$&gt;15$</td>
<td>the whole real line</td>
</tr>
</tbody>
</table>

4.3 Is one approximation second order better

We say that one approximation is better than the other if its unsigned error is less than or equal to that of the other, i.e. as soon as the other does not improve the approximation. An approximation is second order better than the other if it is better as to second order error.

Theorem 4.1. At given $n$, $t$ is second order better than $Z$ as approximation for the cdf of the studentized mean $T$ in the region $R$ in the $(\gamma, \kappa, x)$-space given by any of the two descriptions:

1. the full format: $R$ is given by

$$ x \left\{ 4\gamma^2(x^5 + 2x^3 - 3x) - 12n^{1/2}\gamma(2x^2 + 1) - 6\kappa(x^3 - 3x) + 9(x^3 + x) \right\} \geq 0 $$

(6)
2. the split format: $R = R_1 \cup R_2$ is the union of the two regions given by

$R_1: x \geq 0$

$$4\gamma^2(x^5 + 2x^3 - 3x) - 12n^{1/2}\gamma(2x^2 + 1) - 6\kappa(x^3 - 3x) + 9(x^3 + x) \geq 0$$  \hspace{1cm} (7)

$R_2: x \leq 0$

$$4\gamma^2(x^5 + 2x^3 - 3x) - 12n^{1/2}\gamma(2x^2 + 1) - 6\kappa(x^3 - 3x) + 9(x^3 + x) \leq 0$$  \hspace{1cm} (8)

At fixed $\kappa$, in the $(\gamma, x)$-plane, the regions $R_1$ and $R_2$ are each others reflection image through the origin, and thus $R$ is symmetric for reflection through the origin.

**Proof.** $t$ is second order better than $Z$ for the cdf of $T$ if $|\Delta_{t,2}(x)| \leq |\Delta_{Z,2}(x)|$. With (3), (4), the condition is

$$|n^{1/2}p_1(x) + p_2(x)| \leq |(n^{1/2}p_1(x) + p_2(x) + p(x)|$$  \hspace{1cm} (9)

with polynomials $p, p_1, p_2$ from (4a)–(4c). From the structure $|a| \leq |a+p|$, with $a = \sqrt{n}p_1 + p_2$ and $p = p(x) = -\frac{1}{4}x(x^2 + 1)$, it is equivalent to $a^2 \leq (a + p)^2$ and then to $p(2a + p) \geq 0$ and to $x(-2a - p) \geq 0$, which gives the full format in item 1. Next, the latter inequality shows the region $R$ is the union of the disjoint regions $x = 0$, $(x < 0, -2a - p \leq 0)$ and $(x > 0, -2a - p \geq 0)$. Absorbing $x = 0$ in the second and third region, gives item 2. Then at fixed $\kappa$ one verifies $R_2 = R_1|_{(\gamma, x)\rightarrow (-\gamma, -x)}$ and $R|_{(\gamma, x)\rightarrow (-\gamma, -x)} = R$.

The full format is suited for direct computer implementation, while the split format is more transparent for the mathematical understanding and reduces the study to $R_1$. The point $x = 0$ is the trivial point where $t$ is second order better than $Z$, as there both approximations take the joint value 0.5, make the same error and the same second order error. We also know that $t$ is universally better than $Z$ in case of a normal population, and we see it confirmed in the second order error for $\gamma = \kappa = 0$. Looking for a converse we obtain the following theorem.

**Theorem 4.2.** Student’s $t$ is universally, that is for all $n \geq 2$ and all $x \in \mathbb{R}$, second order better than $Z$ as approximation for the cdf of the studentized mean $T$, if and only if

$$\gamma = 0, \quad -\frac{1}{2} \leq \kappa \leq \frac{3}{2},$$  \hspace{1cm} (10)

in which case the population will be said to have a normal-like distribution.
Examples of populations satisfying the condition are seen in table 3. Examples reaching the inequality limits are the symmetric distributions Beta($\alpha, \beta$), $\beta = \alpha = 9/2$ with $\kappa = -1/2$, and Student’s $t_8$ with $\kappa = 3/2$.

Proof. The condition inequality (6) has the structure $x(a - n^{1/2}\gamma b) \geq 0$ for suitable $a = a(\gamma, \kappa, x)$ and $b = 12(2x^2 + 1) > 0$, and it should hold for all $n$ and all $x$. Assume $\gamma \neq 0$, for instance $\gamma > 0$. Then at $x > 0$ the condition imposes a bound $\sqrt{n} \leq a/\gamma b$, which is either impossible (bound $< \sqrt{2}$) or unacceptable, and the condition cannot be universally satisfied. Thus the condition requires $\gamma = 0$. Then the inequality reduces to

$$x^2 \{x^2(-2\kappa + 3) + 3(2\kappa + 1)\} \geq 0$$

for all $x$, and thus to $x^2(-2\kappa + 3) + 3(2\kappa + 1) \geq 0$ for all $x \neq 0$. This requires $-2\kappa + 3 \geq 0$ and $2\kappa + 1 \geq 0$, that is $-1/2 \leq \kappa \leq 3/2$. The condition is also sufficient.

Looking at the general condition in theorem 4.1, the border surface of the region $R$ in the $(\gamma, \kappa, x)$-space is given by

$$4\gamma^2(x^5 + 2x^3 - 3x) - 12n^{1/2}\gamma(2x^2 + 1) - 6\kappa(x^3 - 3x) + 9(x^3 + x) = 0.$$  (11)

It is an algebraic surface of degree 7. Given this complexity, we present a casewise investigation of $R$, guided by the skewness/kurtosis deviation from normality of the population, and remembering Pearson’s inequality $\kappa \geq \gamma^2 - 2$ (Pearson, 1916):

- **Zero-skewness** populations : $\gamma = 0$. Then $\kappa \geq -2$. This case covers all symmetric populations and more.
- **Normal-like tails** populations : case $\kappa = 0$. Then $-\sqrt{2} \leq \gamma \leq \sqrt{2}$.
- **Gamma-like** populations : $\kappa = 3\gamma^2/2$. We chose initially the gamma family, as a typical general population with nonzero skewness and nonzero kurtosis, which moreover covers the chisquare and the exponential population. A gamma population, Gamma($\alpha, \beta$), with shape parameter $\alpha$ and rate or inverse scale parameter $\beta$, has skewness $\gamma = 2/\sqrt{\alpha}$, kurtosis $\kappa = 6/\alpha$, and shows the relation $\kappa = 3\gamma^2/2$. It turned out that the second order analysis for the gamma family made use only of this relation.

For each case we state a property on the condition for $t$ second order better than $Z$ as approximation for the cdf of $T$, giving : an expression for the condition by corollary of theorem 4.1, a graph of the corresponding region, a practical description of the region which results from a mathematical analysis.
Figure 1: Case of a zero-skewness population. In color is the non-trivial \((\kappa, x)\)-region where \(t\) is second order better than \(Z\) as approximation for the cdf of \(T\).

of the condition. In two cases the mathematical analysis reduced to the study of a region in the \((\gamma, x)\)-plane essentially through a study of a quadratic form in \(\gamma\), whose results were next converted from a \(\gamma\)-view to an \(x\)-view.

Property 4.3. Case of a zero-skewness population: \(\gamma = 0\). Then \(\kappa \geq -2\).

1. The non-trivial region \(R\) where \(t\) is second order better than \(Z\) as approximation for the cdf of the studentized mean \(T\) is given by

\[
\kappa(x^2 - 3) - \frac{3}{2}(x^2 + 1) \leq 0,
\]

and is a region in the \((\kappa, x)\)-plane, which is independent of \(n\). It is presented in Figure 1.

2. This region \(R\) is symmetric around the \(\kappa\)-axis, the border curve has asymptotes \(\kappa = 3/2\) and \(x = \pm \sqrt{3} \approx \pm 1.732\) and a local maximum \(\kappa(0) = -1/2\) at \(x = 0\). At given \(\kappa\), the \(T\)-region where \(t\) is second order better than \(Z\) as approximation for the cdf of \(T\) is given by
Table 3: Some symmetric distributions and their region $R$ where $t$ is second order better than $Z$ as approximation for the cdf of the studentized mean $T$

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$\kappa$</th>
<th>Region $R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal-like</td>
<td>$-\frac{1}{2} \leq \kappa \leq \frac{3}{2}$</td>
<td>$\mathbb{R}$</td>
</tr>
<tr>
<td>Uniform $U[0,a]$</td>
<td>$\frac{-6}{5}$</td>
<td>$</td>
</tr>
<tr>
<td>Uniform sum $US_m$ &amp; $\frac{-6}{5m}$</td>
<td>$m = 2$ $</td>
<td>x</td>
</tr>
<tr>
<td>&amp;</td>
<td>$m \geq 3$ $\frac{-2}{5} \leq \kappa &lt; 0$</td>
<td>$\mathbb{R}$</td>
</tr>
<tr>
<td>Wigner semicircle</td>
<td>-1</td>
<td>$</td>
</tr>
<tr>
<td>$Beta(\alpha,\alpha)$ &amp; $\frac{-6}{2\alpha+3}$</td>
<td>$0 &lt; \alpha &lt; \frac{9}{2}$ $\frac{-2}{\alpha} &lt; \kappa &lt; -\frac{1}{2}$</td>
<td>$</td>
</tr>
<tr>
<td>&amp;</td>
<td>$\alpha \geq \frac{9}{2}$ $\frac{-1}{2} \leq \kappa &lt; 0$</td>
<td>$\mathbb{R}$</td>
</tr>
<tr>
<td>Student $t_{\nu}, \nu &gt; 4$</td>
<td>$\frac{6}{\nu-3}$</td>
<td>$5 \leq \nu \leq 7$ $\frac{3}{2} &lt; \kappa &lt; \frac{1}{2}$</td>
</tr>
<tr>
<td>&amp;</td>
<td>$\nu = 5$ $\frac{-2}{3} &lt; \kappa &lt; 0$</td>
<td>$</td>
</tr>
<tr>
<td>&amp;</td>
<td>$\nu = 6$ $\frac{-2}{5} &lt; \kappa &lt; 0$</td>
<td>$</td>
</tr>
<tr>
<td>&amp;</td>
<td>$\nu = 7$ $\frac{-2}{7} &lt; \kappa &lt; 0$</td>
<td>$</td>
</tr>
<tr>
<td>&amp;</td>
<td>$\nu \geq 8$ $0 &lt; \kappa \leq \frac{3}{2}$</td>
<td>$\mathbb{R}$</td>
</tr>
<tr>
<td>Laplace</td>
<td>3</td>
<td>$</td>
</tr>
<tr>
<td>Logistic</td>
<td>6/5</td>
<td>$\mathbb{R}$</td>
</tr>
</tbody>
</table>

$$-2 \leq \kappa \leq -\frac{1}{2} : |x| \geq a, \quad a = a(\kappa) \uparrow \begin{cases} 0 & \text{to} \frac{3}{\sqrt{7}} \text{ as } \kappa \downarrow \left[-\frac{1}{2} \text{ to } -2\right] \end{cases} ;$$

$$-\frac{1}{2} \leq \kappa \leq \frac{3}{2} : \quad \text{all } x \in \mathbb{R} ;$$

$$\frac{3}{2} < \kappa : |x| \leq a, \quad a = a(\kappa) \downarrow \begin{cases} +\infty & \text{to} \sqrt{3} \text{ as } \kappa \uparrow \left[\frac{3}{2} \text{ to } +\infty\right] \end{cases} .$$

where

$$a(\kappa) = \left(\frac{3(2\kappa + 1)}{2\kappa - 3}\right)^{1/2}, \quad \kappa > \frac{3}{2} \text{ or } \kappa < -\frac{1}{2}. \quad (13)$$

Table 3 gives the region for some symmetric populations from practice.

**Conclusion 4.4.** Case of a zero-skewness population. The $x$-region where $t$ is second order better than $Z$ as approximation for the cdf of $T$ depends on the population kurtosis and not on the sample size. $t$ is uniformly second order better than $Z$ for a normal-like population as obtained in theorem 4.2. $t$ is not second order better than $Z$ in two cases: a narrow center around $x = 0$ for a
population with low tails \((\kappa < -\frac{1}{2})\), the tails of \(T\) for a population with heavy tails \((\kappa > \frac{3}{2})\).

\[\text{Figure 2: Case of a normal-like tails population (}\kappa = 0\text{), and several sample sizes } n. \text{ In color is the } (\gamma, x)\text{-region where } t\text{ is second order better than } Z\text{ as approximation for the cdf of } T.\]

**Property 4.5.** Case of a normal-like tails population : \(\kappa = 0\). Then \(-\sqrt{2} \leq \gamma \leq \sqrt{2}\).

1. The region \(R\) where \(t\) is second order better than \(Z\) is given by

\[x \left( \gamma^2 (x^5 + 2x^3 - 3x) - 3\gamma \sqrt{n}(2x^2 + 1) + \frac{9}{4}(x^3 + x) \right) \geq 0 \quad (14)\]

For any fixed \(n\), it represents a region in the \((\gamma, x)\)-plane. The region is presented in Figure 2. The region is symmetric for reflection through the origin.

2. The half region \(R_1\) in the right half plane \(x \geq 0\) is represented by the inequality

\[Q_n(\gamma, x) := \gamma^2 (x^5 + 2x^3 - 3x) - 3\gamma \sqrt{n}(2x^2 + 1) + \frac{9}{4}(x^3 + x) \geq 0. \quad (15)\]
At given skewness $\gamma$ the $x$-region is given by

\begin{align*}
-\sqrt{2} \leq \gamma \leq 0 : & \quad x \geq 0 \\
0 < \gamma \leq \sqrt{2} : & \quad x \geq x_n(\gamma)
\end{align*}

(16)

where $x_n(\gamma)$ is the unique positive root of $Q_n(\gamma, x)$. As $\gamma$ grows from 0 to $\sqrt{2}$ the border point $x_n(\gamma)$ grows from 0 to a maximum $x_{0,n} := \max_\gamma x_n(\gamma)$ and then goes down to $x_{1,n} := x_n(\sqrt{2}) > 1$. The maximum $x_{0,n}$ is the positive zero of the discriminant of $Q_n(\gamma, x)$ or of

$$
\Delta_n(x) = n(2x^2 + 1)^2 - x^2(x^2 + 3)(x^4 - 1),
$$

(18)

and it is reached at the value $\gamma = \gamma_{0,n}$ which is the double root of $Q_n(\gamma, x_{0,n})$.

The full region $R$ is the union of $R_1$ and its reflection through the origin, $R_2$.

Table 4 lists the values $x_{0,n}$, $x_{1,n}$ for a selection of sample sizes.

**Table 4**: Values $x_{0,n}$ and $x_{1,n}$ in property 4.5 for a selection of sample sizes $n$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$x_{0,n}$</th>
<th>$x_{1,n}$</th>
<th>$n$</th>
<th>$x_{0,n}$</th>
<th>$x_{1,n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.981</td>
<td>1.835</td>
<td>10</td>
<td>2.378</td>
<td>2.100</td>
</tr>
<tr>
<td>6</td>
<td>2.078</td>
<td>1.901</td>
<td>15</td>
<td>2.649</td>
<td>2.273</td>
</tr>
<tr>
<td>7</td>
<td>2.163</td>
<td>1.959</td>
<td>20</td>
<td>2.860</td>
<td>2.404</td>
</tr>
<tr>
<td>8</td>
<td>2.241</td>
<td>2.011</td>
<td>25</td>
<td>3.035</td>
<td>2.509</td>
</tr>
<tr>
<td>9</td>
<td>2.312</td>
<td>2.058</td>
<td>50</td>
<td>3.649</td>
<td>2.865</td>
</tr>
</tbody>
</table>

**Conclusion 4.6.** Case of a population with normal-like tails, $\kappa = 0$. Then $-\sqrt{2} \leq \gamma \leq \sqrt{2}$. The $x$-domain where $t$ is second better than $Z$ as approximation for the cdf of the studentized mean $T$ depends on the population skewness and the sample size. In case of zero skewness, $t$ is uniformly better, as seen in property 4.3. In case of positive skewness – typically a population with a tail to the right, which is the more practical skewed case – $t$ is better in the full left tail $x \leq 0$ and in a right tail $x \geq x_n(\gamma) > 0$. At fixed $\gamma$, this right tail shrinks as $n$ grows. For growing $\gamma$, under chosen $n$, this right tail first shrinks from $x > 0$ to $x \geq x_{0,n}$ and then expands to $x \geq x_{1,n} > 1$. As an example, consider $n = 10$; if $\gamma = 0.2$ the right tail where $t$ is second order better than $Z$ is $x \geq 1.355$, and if $\gamma = 1$ it is the more extreme tail $x \geq 2.261$. A similar property holds for the case of negative skewness.
Figure 3: Case of a gamma-like population. $(\gamma, x)$-region where $t$ is second order better than $Z$ as approximation for the cdf of $T$, for a selection of sample sizes.

Property 4.7. Case of a gamma-like population, $\kappa = \frac{3}{2} \gamma^2$.

1. The region $R$ where $t$ is second order better than $Z$ as approximation for the cdf of the studentized mean $T$ is given by

$$x \left( \gamma^2(4x^5 - x^3 + 15x) - 12\sqrt{n}\gamma(2x^2 + 1) + 9(x^3 + x) \right) \geq 0.$$  \hspace{1cm} (19)

As $\gamma > 0$, it is for every fixed $n$ a region in the upper half $(\gamma, x)$-plane. $R$ is presented in Figure 3 for selected sample sizes.

2. $R = R_1 \cup R_2$, where $R_2 : x \leq 0$ is the full left quarter plane, and $R_1$ is the region in the right quarter plane $x \geq 0$ given by

$$Q_n(\gamma, x) := \gamma^2(4x^5 - x^3 + 15x) - 12\sqrt{n}\gamma(2x^2 + 1) + 9(x^3 + x) \geq 0.$$  \hspace{1cm} (20)

At fixed $\gamma > 0$, this condition reduces to $x \geq x_n(\gamma)$ where $x_n(\gamma)$ is the unique positive root of $Q_n(\gamma, x)$. As $\gamma$ grows from 0 to $\infty$ the border value $x_n(\gamma)$ grows from 0 to a maximum $x_{0,n} := \max_{\gamma} x_n(\gamma)$ and then goes down to $x_n(\infty) = 0$. The maximum $x_{0,n}$ is the positive zero of the discriminant of $Q_n(\gamma, x)$ or of

$$\Delta_n(x) = 4n(2x^2 + 1)^2 - (4x^5 - x^3 + 15x)(x^3 + x)$$  \hspace{1cm} (21)

and it is reached at the value $\gamma = \gamma_{0,n}$ which is the double root of $Q_n(\gamma, x_{0,n})$. 

Table 5: Values of $x_{0,n}$ in property 4.7 for for selected sample sizes $n$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$x_{0,n}$</th>
<th>$n$</th>
<th>$x_{0,n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2.046</td>
<td>10</td>
<td>2.483</td>
</tr>
<tr>
<td>6</td>
<td>2.157</td>
<td>15</td>
<td>2.764</td>
</tr>
<tr>
<td>7</td>
<td>2.253</td>
<td>20</td>
<td>2.978</td>
</tr>
<tr>
<td>8</td>
<td>2.337</td>
<td>25</td>
<td>3.154</td>
</tr>
<tr>
<td>9</td>
<td>2.414</td>
<td>50</td>
<td>3.761</td>
</tr>
</tbody>
</table>

As an example, consider a one-parameter gamma population $X \sim \text{Gamma}(\alpha)$ with shape parameter $\alpha > 0$. Then $\gamma = 2/\sqrt{\alpha}$. The region for $t$ second order better than $Z$ as approximation for the cdf of $T$ is the region $R = R_1 \cup R_2$ in the $(\alpha, x)$-halfplane with $\alpha > 0$, where $R_2 : x \leq 0$ and $R_1$ is the region in the right quarter plane $x \geq 0$ satisfying

$$9\alpha(x^3 + x) - 24\sqrt{n}\sqrt{\alpha}(2x^2 + 1) + 4(4x^5 - x^3 + 15x) \geq 0. \quad (22)$$

The region $R$ is presented in Figure 4.

Figure 4: Case of a Gamma($\alpha$) population. $(\alpha, x)$-region where $t$ is second order better than $Z$ as approximation for the cdf of $T$, for a selection of sample sizes.

**Conclusion 4.8.** Case of a gamma-like population, $\kappa = 3\gamma^2/2$. The $x$-domain where $t$ is second better than $Z$ as approximation for the cdf of the studentized mean $T$ depends on the population skewness and the sample size. For any $\gamma > 0$, $t$ is better in the full left tail $x \leq 0$ and in a right tail.
\[
x \geq x_n(\gamma) > 0.\]  
At fixed \(\gamma\), this right tail shrinks as \(n\) grows. For growing \(\gamma\), under chosen \(n\), this right tail first shrinks from \(x \geq 0\) to \(x \geq x_{0,n} > 0\) and then expands to \(x > 0\). As an example, consider \(n = 10\); if \(\gamma = 2\) the right tail where \(t\) is second order better than \(Z\) is \(x \geq 1.99\), and if \(\gamma = 6\) it is the wide tail \(x \geq 1.05\).

5 Simulation of the errors in the cdf approximations

Simulation of the errors  The true errors of the \(Z\) and \(t\) approximations for the true cdf of \(T\), \(\Delta_Z(x)\) and \(\Delta_t(x)\), defined by (1), are presented in Figure 5, in a panel graph for a selection of populations (normal, skewed and tailed) and a selection of sample sizes (\(n = 5, 10, 20, 50, 100\)), each with \(x\)-interval \([-4, 4]\). The true cdf of \(T\) is simulated as the empirical cdf, based on \(n = 10^6\) random values of \(T\), as described in Section 2. This empirical cdf \(F_T\), and the normal and Student cdf’s, \(\Phi\) and \(F_t\), are available as functions in R. Then the error functions (1) were plotted for \(x \in [-4, 4]\) based on 200 \(x\)-values.

The figure illustrates known convergence theorems, \(T \xrightarrow{d} Z\), \(t \xrightarrow{d} Z\); for the exponential distribution the convergence of \(T\) is very slow, due to the high population skewness. Next the figure confirms the theoretical results obtained from the second order errors: in the approximations \(t\) and \(Z\) for the cdf of \(T\),
the main source of error is skewness in the population; $t$ is close to or better than $Z$ for the cdf of $T$ almost always, except in the case of small samples from a very skewed population.

**Evaluation of the approximations** Figure 6 presents two indicators for the $x$-region where $t$ is better than $Z$ as approximation for the cdf of $T$. The tolerant indicator $B_n$ is defined by: $B_n(x) = 1$ as soon as any of the next three conditions is satisfied:

1. $t$ is better than $Z$ as to error for the cdf of $T$: $|\Delta_t(x)| \leq |\Delta_Z(x)|$;

2. the $t$-error is irrelevant: $|\Delta_t(x)| < 10^{-2}$, i.e. it affects typically at most the third decimal of the cdf;

3. the difference in the $t$ and $Z$ approximations is irrelevant: $|F_t(x) - \Phi(x)| < 10^{-2}$, i.e. the difference between the two approximations will be visible at most in the third decimal;

otherwise $B_n(x) = 0.8$. A strict indicator $C_n$ is defined by $C_n(x) = 0.6$ if the above condition (1) is satisfied, otherwise $C_n(x) = 0.4$.

$C_n$ shows the simulations match the theoretical results for the errors. $B_n$ combines the errors and their relevance, and strengthens the conclusion that $t$ is not overclassed by $Z$, except in the right tail of $T$ in case of a small sample
from a positively skewed population, and similarly in the left tail of $T$ in case of a negatively skewed population, that is in a case where none of the two approximations is acceptable.

**Conclusion 5.1.** Combining the theoretical and the simulation analyses, there is strong indication in favor of the one-distribution approach for the studentized mean in practice: $t$ is always acceptable, except for small samples from a non-normal population up to moderate samples from an explicit skewed population. In the exception case, none of the two approximations is acceptable. Moreover, up to second order error, $t$ is uniformly – for all $n$ and all $x$ – the best iff the population is normal-like in the sense of skewness $\gamma = 0$ and kurtosis $-\frac{1}{2} \leq \kappa \leq \frac{3}{2}$.

## 6 Cornish-Fisher expansions for the quantile function

The two approximations $t$ and $Z$ for the distribution of the studentized mean $T$ provide each approximations for the quantiles of $T$. In this section we compare these approximations. As the quantile function of a rv is the inverse of the cumulative distribution function studied above, we may expect a similar analysis. Therefore, we present only key expressions and corresponding figures; a detailed analysis may be obtained from the authors.

**Theorem 6.1.** *Cornish-Fisher expansions for the quantile function of the studentized mean.* Let $0 < \alpha < 1$. The Cornish-Fisher expansions of the studentized mean quantile $T_\alpha = Q_T(\alpha) = F_T^{-1}(\alpha)$ with respectively the normal quantile $z_\alpha = Q_Z(\alpha) = \Phi^{-1}(\alpha)$ and the Student $t = t_{n-1}$ quantile $t_\alpha = Q_t(\alpha) = F_t^{-1}(\alpha)$ as germ are given by

\[
T_\alpha = z_\alpha - n^{-1/2} p_1(z_\alpha) - n^{-1} [q_2(z_\alpha) + p(z_\alpha)] + o(n^{-1}) \tag{1}
\]

\[
T_\alpha = t_\alpha - n^{-1/2} p_1(z_\alpha) - n^{-1} q_2(z_\alpha) + o(n^{-1}) \tag{2}
\]

with the polynomials $p_1$ and $p$ as before, and $q_2$ as a new polynomial,

\[
p_1(z) = \frac{1}{6} \gamma (2z^2 + 1)
\]

\[
q_2(z) = \frac{1}{12} \kappa (z^3 - 3z) - \frac{1}{72} \gamma^2 (20z^3 - 5z) \tag{3}
\]

\[
p(z) = -\frac{1}{4} (z^3 + z).
\]

Under sufficient moment conditions, each expansion is valid for $\alpha$ in the region $n^{-2} < \alpha < 1 - n^{-2}.$
Thus, for a typical $\alpha = 0.01, 0.025, 0.05$ the condition requires respectively
$n > 10, 6, 4$.

Proof. We propose an expression

$$T_\alpha = z + n^{-1/2} a_1(z) + n^{-1} a_2(z) + o(n^{-1}), \quad z = z_\alpha$$  \hspace{1cm} (4)

where the $a_j$ are unknown polynomials, that will be obtained as a corollary of
the Edgeworth expansion for $F_T$. The proposed structure provides

$$\alpha = F_T(T_\alpha) = F_T(z_\alpha + n^{-1/2} \delta), \quad \delta = a_1 + n^{-1/2} a_2 + o(n^{-1/2}).$$

Using the Taylor expansion for $F_T$, followed by the Edgeworth expansion (2),
and equalizing corresponding $n-$powers, we find

$$a_1 = -p_1 \quad \text{and} \quad a_2 = -p_2 - p - a_1 p_1' + z a_1 p_1 + \frac{1}{2} z a_1^2,$$

which leads to the first expansion (1) in the theorem.

We apply this result for a normal population $X$, and obtain a Cornish-Fisher
expansion for the quantile of Student’s $t \sim t_{n-1}$:

$$t_\alpha = z_\alpha - n^{-1} p(z_\alpha) + o(n^{-1}) = z_\alpha + \frac{1}{4} (z_\alpha^2 + z_\alpha) n^{-1} + o(n^{-1}).$$  \hspace{1cm} (5)

This is the expansion up to order 2 for Student’s $t_n$ quantile as well.
Finally $T_\alpha - t_\alpha$ provides the second expansion (2) in the theorem.
For the studentized mean with $n$-weight variance estimator, Hall (1992) arrived
at the structure condition $n^{-c} < \alpha < 1 - n^{-c}$ for a valid expansion to order
c > 0. We used the same construction for $T_\alpha$ to order 2, which provides the
validity condition $n^{-2} < \alpha < 1 - n^{-2}$ for the expansion of $T_\alpha$ with normal
ger; this condition is inherited by the expansion for $t_\alpha$, and further for $T_\alpha$
with the $t$-germ.

7 Second order errors in the quantile approximations

For the $Z$ and $t$ quantiles as approximations for the true quantile $T_\alpha$ we denote
the errors,

$$\Delta Z(\alpha) = T_\alpha - z_\alpha \quad \text{and} \quad \Delta t(\alpha) = T_\alpha - t_\alpha$$  \hspace{1cm} (6)

and, from the theorem, the first order errors/corrections,

$$\Delta Z,1(\alpha) = \Delta t,1(\alpha) = -n^{-1/2} p_1(z_\alpha) = -\frac{1}{6} \gamma n^{-1/2} \left\{ 2 \left[ \Phi^{-1}(\alpha) \right]^2 + 1 \right\}$$  \hspace{1cm} (7)

as well as the second order errors/corrections,

$$\Delta Z,2(\alpha) = -n^{-1/2} p_1(z_\alpha) - n^{-1} \left[ q_2(z_\alpha) + p(z_\alpha) \right]$$  \hspace{1cm} (8)
$$\Delta t,2(\alpha) = -n^{-1/2} p_1(z_\alpha) - n^{-1} q_2(z_\alpha)$$  \hspace{1cm} (9)
It should be noticed that the errors in the approximations for the $\alpha$-quantile $T_\alpha$ depend on population characteristics, sample size and $\alpha$; the given error expressions happen to depend on $\alpha$ through $z_\alpha = \Phi^{-1}(\alpha)$.

**Source of the errors** Both approximations are equivalent as to their lowest order error, which is of order $n^{-1/2}$, and is caused by population skewness. In the second order errors, up to order $n^{-1}$, the population impact is the same for both and is due to skewness in the first order and kurtosis in second order. The difference in the errors depends on $\alpha$ and $n$ only.

**Is the difference in the approximations relevant** The unsigned difference in the second order approximations is

$$|t_\alpha - z_\alpha| = \left| \frac{1}{4} |z_\alpha^3 + z_\alpha| n^{-1} \right|. \quad (10)$$

The difference is of order $n^{-1}$. At fixed $n$ it is increasing in the tails at order $O(|z_\alpha|^3)$, which means a higher $n$ is required to control the difference. At given $\alpha$ the difference will be less than $10^{-1}$ if $n > \frac{3}{2} |z_\alpha^3 + z_\alpha|$, and the relative difference $|t_\alpha - z_\alpha|/|z_\alpha|$ is at most 5% if $n > 5|z_\alpha^2 + 1|$. At $\alpha \in \{0.05; 0.95\}$ these conditions are $n > 15$ and $n > 18$, while $\alpha \in \{0.025; 0.975\}$ gives $n > 23$ and $n > 24$.

**Is one approximation second order better** The $t$-quantile is second order better than the $Z$-quantile as approximation for the true $T$-quantile if $|\Delta_{t,2}(\alpha)| \leq |\Delta_{Z,2}(\alpha)|$ which, by (8), (9) and (3) leads to the following theorem.

**Theorem 7.1.** At given $n$, $t$ is second order better than $Z$ as approximation for the $\alpha$-quantile of the studentized mean $T$, in the region $R$ in the $(\gamma, \kappa, \alpha)$-space given by any of the two descriptions

1. the full format: $R$ is given by

$$z_\alpha \left( 5\gamma^2(4z_\alpha^3 - z_\alpha) - 12n^{1/2}\gamma(2z_\alpha^2 + 1) - 6\kappa(z_\alpha^3 - 3z_\alpha) + 9(z_\alpha^3 + z_\alpha) \right) \geq 0 \quad (11)$$
2. the split format: \( R = R_1 \cup R_2 \) is the union of the two regions given by

\[
R_1 : 0 \leq \alpha \leq \frac{1}{2}
\]

\[
5\gamma^2(4z^3_{\alpha} - z_{\alpha}) - 12n^{1/2}\gamma(2z^2_{\alpha} + 1) - 6\kappa(z^3_{\alpha} - 3z_{\alpha}) + 9(z^3_{\alpha} + z_{\alpha}) \leq 0
\]

(12)

\[
R_2 : \frac{1}{2} \leq \alpha \leq 1
\]

\[
5\gamma^2(4z^3_{\alpha} - z_{\alpha}) - 12n^{1/2}\gamma(2z^2_{\alpha} + 1) - 6\kappa(z^3_{\alpha} - 3z_{\alpha}) + 9(z^3_{\alpha} + z_{\alpha}) \geq 0
\]

(13)

At fixed \( \kappa \) the sections \( R_1 \) and \( R_2 \) in the \((\gamma, \alpha)\)-plane are each others reflection through the point \((\gamma, \alpha) = (0, \frac{1}{2})\), i.e. \( R_2 = R_1|_{(\gamma, \alpha) \leftrightarrow (-\gamma, 1-\alpha)} \), and thus the \( R \)-section is symmetric for reflection through the point \((0, \frac{1}{2})\).

A study of the quantile errors is done, in parallel to that on the cdf errors, for some special cases, with results below. Obviously \( z_{\alpha} = 0 \), that is \( \alpha = 0.5 \), gives the trivial solution. Also, \( t \) is universally – for all \( n \) and all \( \alpha \) – better than \( Z \) iff the population is normal-like in the sense of equation (10).

**Zero-skewness population** : \( \gamma = 0 \). Then \( \kappa \geq -2 \). The condition (11) for the region \( R \) reduces to:

\[
(2\kappa - 3)z^2_{\alpha} \leq 3(2\kappa + 1).
\]

(14)

and has the same form as (12), which provides directly similar conclusions. In particular the condition is satisfied for all \( \alpha \) as soon as \(-\frac{1}{2} \leq \kappa \leq \frac{3}{2} \), i.e. for normal-like populations. Figure 7 presents the region.

**Normal-like tails population** : \( \kappa = 0 \). Then \(-\sqrt{2} \leq \gamma \leq \sqrt{2} \). \( R \) is a region in the \((\gamma, \alpha)\)-plane, in the belt \( 0 \leq \alpha \leq 1 \), given by:

\[
z_{\alpha} \left(5\gamma^2(4z^3_{\alpha} - z_{\alpha}) - 12n^{1/2}\gamma(2z^2_{\alpha} + 1) + 9(z^3_{\alpha} + z_{\alpha}) \right) \geq 0.
\]

(15)

It is shown in Figure 7.

**Gamma-like population** : \( \kappa = \frac{3}{2}\gamma^2 \). The region \( R \) is now given by

\[
z_{\alpha} \left(11\gamma^2(z^3_{\alpha} + 2z_{\alpha}) - 12n^{1/2}\gamma(2z^2_{\alpha} + 1) + 9(z^3_{\alpha} + z_{\alpha}) \right) \geq 0
\]

(16)

and is presented in Figure 7 for selected sample sizes \( n \).
Figure 7: Region where \( t_\alpha \) is second order better than \( z_\alpha \) as approximation for the studentized mean quantile \( T_\alpha \). Top: case of a zero-skewness population, \((\alpha, \kappa)\)-region. Middle: case of a zero-kurtosis population, \((\gamma, \alpha)\)-region. Bottom: gamma-like population, \((\gamma, \alpha)\)-region.
Figure 8: Errors of the Z and t approximations for the quantile function of the studentized mean $T$.

Figure 9: Indicators for the $\alpha$-region where $t$ is better than $Z$ as approximation for the quantile function of $T$. 
8 Simulation of the errors in the approximations of the qf

Simulation of the errors The error functions $\Delta_Z(\alpha) = (Q_T - Q_Z)(\alpha) = T_\alpha - z_\alpha$ and $\Delta_t(\alpha) = (Q_T - Q_t)(\alpha) = T_\alpha - t_\alpha$ of the Z and t approximations for the true qf of $T$ are presented in Figure 8, in a panel graph for the selection of populations and sample sizes used for the cdf-study in Section 5. They were obtained as line plots in the software R, based on 200 $\alpha$-values. The quantile function $Q_T = F_T^{-1}$ is accurately estimated as the empirical quantile function of the $N = 10^6$ simulated data $T_j$ presented in Section 2. This empirical quantile function $Q_T$, the normal quantile function $Q_Z = \Phi^{-1}$ and the Student quantile function $Q_t = F_t^{-1}$ with $n - 1$ degrees of freedom, are available functions in the software R. The error graphs were obtained in R as line plots based on 200 $\alpha$-values.

Evaluation of the approximations Figure 9 presents two indicators for the $\alpha$-region where $t_\alpha$ is better than $Z_\alpha$ as approximation for the quantile $T_\alpha$. The tolerant indicator $B_n$ is defined by : $B_n(\alpha) = 1$ as soon as any of the next three conditions is satisfied:

(1) $t$ is better than $Z$ as to error for the qf of $T$ : $|\Delta_t(\alpha)| \leq |\Delta_Z(\alpha)|$;

(2) the $t$-error is irrelevant : $|\Delta_t(\alpha)| < 10^{-1}$, i.e. it affects typically at most the second decimal of the quantile;

(3) the difference in the $t$ and $Z$ approximations is irrelevant : $|t_\alpha - z_\alpha| < 10^{-1}$, i.e. the difference between the two approximations will be visible at most in the second decimal;

otherwise $B_n(\alpha) = 0.8$. A strict indicator $C_n$ is defined by $C_n(\alpha) = 0.6$ if the above condition (1) is satisfied, otherwise $C_n(\alpha) = 0.4$. $C_n$ shows the simulations match the theoretical results for the errors. $B_n$ combines the errors and their relevance, and strengthens the conclusion that $t$ is not overclassed by $Z$, except in the right tail of $T$ in case of a small sample from a positively skewed population, and similarly in the left tail of $T$ in case of a negatively skewed population.

9 Conclusion Given the theoretical and the simulation results, we see scientific evidence, next to pedagogical evidence, to recommend the $t$-rule for practice: the distribution of the studentized mean $T$ is Student’s $t$ with $n - 1$ degrees of freedom
in case of a normal population, and this distribution is acceptable in general except for small samples from a noticeable skewed population.

In more detail, the paper compares the errors of the two classical approximation models for the distribution of \( T \), namely the standard normal \( Z \) and Student’s \( t \) with \( n - 1 \) degrees of freedom, and this for the error on the cumulative distribution function and for the error on the quantile function. For each, expressions for the asymptotic expansions to order two – Edgeworth expansions and Cornish-Fisher expansions – are obtained, and simulations of the full errors are given for a representative selection of populations. Conclusions are listed below, essentially and mostly in terms of the cumulative distribution functions.

1. The difference between the \( t \) and \( Z \) cumulative distribution functions is of order \( O(n^{-1}) \), and the same is true for their quantile functions. From the lowest order differences we conclude that the difference between the two approximations will be visible typically at most from the third decimal on in the cumulative probabilities if \( n > 15 \).

2. The first order or \( O(n^{-1/2}) \) error of \( t \) and \( Z \) as approximations for the cdf of \( T \) is the same, and depends on the population through skewness. This is no surprise, as \( t \) and \( Z \) are first order equal – they differ only at order \( n^{-1} \) – and are symmetric. As a rule of thumb based on the first order error: each of the two approximations will be correct to at least the second decimal for the cdf provided \( |\gamma| \leq 8\sqrt{n}/100 \). In particular, none of the two approximations should be used in case of small to moderate samples from a skewed population.

3. In the second order errors the part on population causes of error is the same for both approximations, and the dominant population causes of error are skewness in first order and kurtosis in second order.

4. We obtain expressions and graphs for the \((\gamma, \kappa, x)\)-domain where \( t \) is second order better than \( Z \) as approximation for the cumulative distribution function of \( T \). It is found that \( t \) is mostly – or at least often – better than \( Z \). In particular, \( t \) is uniformly second order better than \( Z \) as approximation for the cdf of \( T \), iff and only if the population is normal-like in the sense of skewness \( \gamma = 0 \) and kurtosis \(-1/2 \leq \kappa \leq 3/2 \). Simulation of the errors based on various types of populations confirms the theoretical findings.

5. Finally, by a combination of the region where \( t \) is better than \( Z \) as to error, and a tolerance region for irrelevance of the \( t \)-error or irrelevance
of the difference between the two approximations, the simulations show full support of the t-rule.

Acknowledgments. The authors thank the VLIR UOS University Development Cooperation, Belgium for support during this research.

References


Received: November 29, 2022; Published: March 29, 2023