Exact bounds on the closeness between the Student and standard normal distributions

Iosif Pinelis

Abstract. Upper bounds on the Kolmogorov distance (and, equivalently in this case, on the total variation distance) between the Student distribution with $p$ degrees of freedom ($\text{SD}_p$) and the standard normal distribution are obtained. These bounds are in a certain sense best possible, and the corresponding relative errors are small even for moderate values of $p$. The same bounds hold on the closeness between $\text{SD}_p$ and $\text{SD}_q$ with $q > p$. Comparisons with known bounds are made.

Mathematics Subject Classification. 62E17, 60E15, 62E20, 62E15.

Received June 13, 2013. Revised May 19, 2014.

1. Summary and discussion

The density and distribution functions of Student’s distribution with $p$ degrees of freedom ($\text{SD}_p$) are given, respectively, by the formulas

$$f_p(x) := \frac{\Gamma \left( \frac{p+1}{2} \right) \left( 1 + \frac{x^2}{p} \right)^{-\frac{p+1}{2}}}{\sqrt{\pi p} \Gamma \left( \frac{p}{2} \right)}$$

and

$$F_p(x) := \int_{-\infty}^{x} f_p(u) \, du$$

for all real $p > 0$ and all real $x$. Let us extend definitions (1.1) and (1.2) by continuity to $p = \infty$, so that $f_\infty =: \varphi$ and $F_\infty =: \Phi$ are the density and distribution functions of the standard normal distribution (SND).

It is a textbook fact that the $\text{SD}_p$ is close to the SND when $p$ is large, say in the sense that $f_p(x) \xrightarrow{p \to \infty} f_\infty(x)$ for each real $x$. By Scheffé’s theorem [8], this implies the convergence of the total variation distance

$$d_{TV}(p) = \frac{1}{2} \int_{-\infty}^{\infty} |f_p(x) - \varphi(x)| \, dx$$

(1.3)

to 0 as $p \to \infty$. In fact, the convergence of the $\text{SD}_p$ to the SND is presented in [8] as the motivating case.

Keywords and phrases. Student’s distribution, standard normal distribution, Kolmogorov distance, total variation distance, probability inequalities.

1 Department of Mathematical Sciences, Michigan Technological University, Houghton, 49931 Michigan, USA. ipinelis@mtu.edu

Article published by EDP Sciences © EDP Sciences, SMAI 2015
Consider also the Kolmogorov distance

\[ d_{Ko}(p) := \sup_{x \in \mathbb{R}} |F_p(x) - \Phi(x)| \]

between the SD\(_p\) and SND. It is clear that, for any two probability distributions, the Kolmogorov distance between them is no greater than twice the total variation distance, and hence the convergence of the latter distance to 0 implies that of the former.

However, in the present case one can say more. For any \( p \) and \( q \) in the interval \((0, \infty]\), let \( d_{Ko}(p, q) \) and \( d_{TV}(p, q) \) denote, respectively, the Kolmogorov distance and the total variation distance between SD\(_p\) and SD\(_q\), so that \( d_{Ko}(p) = d_{Ko}(p, \infty) \) and \( d_{TV}(p) = d_{TV}(p, \infty) \).

**Proposition 1.1.**

(i) For all \( p \) and \( q \) such that \( 0 < p < q \leq \infty \)

\[ \frac{1}{2} d_{TV}(p, q) = d_{Ko}(p, q) = \max_{x \in (0, \infty)} (F_q(x) - F_p(x)). \]  

(ii) Moreover, for each \( p \in (0, \infty) \) the distance \( d_{Ko}(p, q) \) is strictly increasing in \( q \in [p, \infty] \), and for each \( q \in (0, \infty) \) the distance \( d_{Ko}(p, q) \) is strictly decreasing in \( p \in (0, q] \). In particular,

\[ d_{Ko}(p, q) < d_{Ko}(p, \infty) = d_{Ko}(p) \quad \text{(1.5)} \]

for all \( p \) and \( q \) such that \( 0 < p \leq q < \infty \), and \( d_{Ko}(p) \) is strictly decreasing in \( p \in (0, \infty) \).

Statement (ii) holds as well with \( d_{TV} \) in place of \( d_{Ko} \).

This proposition and the other results stated in this section will be proved in Section 2.

The motivation for this study comes from the discussion in [4]. In turn, the paper [4] was motivated by developments of [6].

**Theorem 1.2.** For any real \( p \geq 4 \)

\[ \frac{1}{2} d_{TV}(p) = d_{Ko}(p) < \frac{C}{p}, \quad \text{(1.6)} \]

where

\[ C := \frac{1}{4} \sqrt{\frac{7 + 5\sqrt{2}}{\pi e^{1+\sqrt{2}}} \approx 0.158 \ldots} \quad \text{(1.7)} \]

Moreover,

\[ \lim_{p \to \infty} p d_{Ko}(p) = C, \quad \text{(1.8)} \]

so that the constant \( C \) is the best possible one in (1.6).

**Corollary 1.3.** For all \( p \) and \( q \) such that \( 4 \leq p < q \leq \infty \)

\[ \frac{1}{2} d_{TV}(p, q) = d_{Ko}(p, q) < \frac{C}{p}. \quad \text{(1.9)} \]

Graphs suggest that the bound \( C/p \) is very close to \( d_{Ko}(p) \) for all \( p \geq 4 \), even in terms of the relative error (see [5], Figs. 1 and 2). Moreover, by (1.8), the relative error \( \frac{C/p - d_{Ko}(p)}{d_{Ko}(p)} - 1 \) of the upper bound \( C/p \) goes to 0 as \( p \to \infty \). This convergence appears to be rather fast. E.g., for \( p = 12 \), the relative and absolute errors of the bound \( C/p \) are less than 1.5% and \( 2 \times 10^{-4} \). One may as well note that, if the distance \( d_{Ko}(p) \) is considered as a kind of “initial” error – of the approximation of the Student distribution by the SND, then the relative error \( \frac{C/p - d_{Ko}(p)}{d_{Ko}(p)} - 1 \) is a relative error “of the second order”, in the sense that it is the relative error of the estimate \( C/p \) of the initial error \( d_{Ko}(p) \). Comparisons with results in [1, 2, 7, 9] can also be found in [5].
2. Proofs

Introduce

\[ r_{p,q}(x) := \frac{f_p(x)}{f_q(x)} \quad \text{and} \quad r_p(x) := r_{p,\infty}(x) = \frac{f_p(x)}{\varphi(x)}. \]  

(2.1)

**Lemma 2.1.** For each pair \((p, q)\) such that \(0 < p < q \leq \infty\)

(i) the ratio \(r_{p,q}(x)\) decreases in \(x \in [0, 1]\) from \(r_{p,q}(0) < 1\), and then increases in \(x \in [1, \infty)\) to \(\infty\); therefore,

(ii) there is a unique point \(x_{p,q} \in (0, \infty)\) (which is in fact greater than 1) such that

\[ f_p(x) < f_q(x) \text{ for all } x \in [0, x_{p,q}), \]

\[ f_p(x) = f_q(x_{p,q}), \]

\[ f_p(x) > f_q(x) \text{ for all } x \in (x_{p,q}, \infty), \]  

and hence

\[ d_{Ko}(p, q) = F_q(x_{p,q}) - F_p(x_{p,q}). \]  

(2.3)

**Proof of Lemma 2.1.** Take indeed any \(p\) and \(q\) such that \(0 < p < q \leq \infty\). A key observation here (borrowed from [3]) is that \(r_{p,q}(x)\) decreases in \(x \in [0, 1]\) and increases in \(x \in [1, \infty)\). Moreover, by Lemma 2.1 in [3], \(f_p(0)\) increases in \(p > 0\) and hence \(r_{p,q}(0) < 1\). On the other hand, it is easy to see that \(r_{p,q}(x) \to \infty\) as \(x \to \infty\). This completes the proof of part (i) of Lemma 2.1, which in turn implies that there is a unique \(x_{p,q} \in (0, \infty)\) such that \(r_{p,q}(x) < 1\) for \(x \in [0, x_{p,q})\), \(r_{p,q}(x_{p,q}) = 1\), and \(r_{p,q}(x) > 1\) for \(x \in (x_{p,q}, \infty)\) (at that necessarily \(x_{p,q} > 1\)).

In other words, one has the relations (2.2). Since \((F_q - F_p)' = f_q - f_p\), one now sees that \(F_q(x) - F_p(x)\) increases in \(x \in [0, x_{p,q}]\) from 0 to \(F_q(x_{p,q}) - F_p(x_{p,q}) > 0\), and then decreases in \(x \in [x_{p,q}, \infty)\) to 0. So, (2.3) follows by the symmetry of the Student and standard normal distributions. Thus, the lemma is completely proved. \(\Box\)

**Proof of Proposition 1.1.** Take indeed any \(p\) and \(q\) such that \(0 < p < q \leq \infty\). By Lemma 2.1 and the symmetry of the SDp,

\[ d_{TV}(p, q) = \int_0^{x_{p,q}} (f_q - f_p) + \int_{x_{p,q}}^{\infty} (f_p - f_q) = 2 \int_0^{x_{p,q}} (f_q - f_p) = 2(F_q(x_{p,q}) - F_p(x_{p,q})) = 2d_{Ko}(p, q), \]

which proves part (i) of Proposition 1.1. Part (ii) of the proposition now follows by the second equality in (1.4) and the stochastic monotonicity result of [3], which implies that \(F_p(x)\) is strictly increasing in \(p \in (0, \infty)\) for each \(x \in (0, \infty)\). \(\Box\)

**Proof of Theorem 1.2.** This proof is based on a number of more or less technical statements in [5]. Indeed, the relations in (1.6) immediately follow by Theorem 1.3 and Lemma 2.6 in [5]. It remains to verify (1.8). First here, use l’Hospital’s rule to find that for all real \(x\)

\[ \lim_{a \to 0} \frac{f_{1/a}(x) - f_{\infty}(x)}{a} = \lim_{a \to 0} \frac{\partial f_{1/a}(x)}{\partial a} = \lambda(x) := \frac{x^4 - 2x^2 - 1}{4} \varphi(x). \]  

(2.4)

Next, introduce

\[ c_a := f_{1/a}(0) \quad \text{and} \quad g_a(x) = f_{1/a}(x)/c_a \]  

(2.5)

for all real \(a \geq 0\), assuming the convention \(1/0 := \infty\), so that \(f_{1/a}(x) = c_ag_a(x)\). Then for all real \(a \geq 0\) and all real \(x\)

\[ |f_{1/a}(x) - f_{\infty}(x)| \leq |c_a - c_0|g_a(x) + c_0|g_a(x) - g_0(x)| \leq |c_a - c_0| + |g_a(x) - g_0(x)|, \]  

(2.6)
since \( g_a(x) \leq 1 \) and \( c_0 = 1/\sqrt{2\pi} < 1 \). By (2.4) and (2.5), the ratio \( \frac{|c_a-c_0|}{a} \) is continuous in \( a > 0 \) and converges to a finite limit \( (\varphi(0)/4) \) as \( a \downarrow 0 \), and hence is bounded in \( a \in (0,1] \). Now note that

\[
\left| \frac{\partial g_a(x)}{\partial a} \right| = (1 + ax^2)^{-1/(1+3a)/2} |(Dg)(a,x)| \leq |(Dg)(a,x)|,
\]

where

\[
(Dg)(a,x) := \frac{1 + ax^2}{2a^2} \ln (1 + ax) - a(1 + a)x^2.
\]

Using the Taylor expansion \( \ln(1+u) = u - \theta u^3/2 \) for \( u > 0 \) and some \( \theta = \theta(u) \in (0,1) \), one sees that \( (Dg)(a,x) \) is a polynomial in \( a, x, \theta \) and hence bounded in \( (a,x) \in (0,1] \times [0, \tilde{x}_0] \) – note that, in accordance with the definition of \( \tilde{x}_a \) in Theorem 1.3 in [5],

\[
\tilde{x}_0 = \sqrt{1 + \sqrt{2}} \in (0,\infty);
\]

hence, \( \left| \frac{\partial g_a(x)}{\partial a} \right| \) is bounded in \( (a,x) \in (0,1] \times [0, \tilde{x}_0] \) and, by the mean value theorem, so is \( \left| \frac{g_c(x) - g_d(x)}{a} \right| \). Recalling also (2.6) and that the ratio \( \frac{|c_a-c_0|}{a} \) is bounded in \( a \in (0,1] \), one concludes that the ratio \( \frac{|f_{1/a}(x) - f_{\infty}(x)}{a} \) is bounded in \( (a,x) \in (0,1] \times [0, \tilde{x}_0] \). So, by (2.4) and dominated convergence,

\[
p_d K_0(p) \geq p \left[ F_\infty(\tilde{x}_0) - F_p(\tilde{x}_0) \right] = - \int_0^{\tilde{x}_0} \frac{f_{1/a}(x) - f_{\infty}(x)}{a} \, dx \rightarrow - \int_0^{\tilde{x}_0} \lambda(x) \, dx = \frac{(\tilde{x}_0^3 + \tilde{x}_0) \varphi(\tilde{x}_0)}{4} = C,
\]

where \( \lambda(x) \) is defined in (2.4). This, together with (1.6), implies (1.8). The proof of Theorem 1.2 is now complete. □

References

[1] S. Bourguin and C. Tudor, Malliavin calculus and self normalized sums. Seminaire de Probabilites XLIV. Lect. Notes Math. Springer (2013) 323.
[2] T. Cacoullos, V. Papathanasiou, and S.A. Utev, Variational inequalities with examples and an application to the central limit theorem. Ann. Probab. 22 (1994) 1607–1618.
[3] I. Pinelis, Monotone tail and moment ratio properties of Student’s family of distributions. Version 2. http://arxiv.org/abs/1101.3286 (2012).
[4] I. Pinelis, On the Berry–Esseen bound for the Student statistic. Version 2. http://arxiv.org/abs/1101.3286 (2012).
[5] I. Pinelis, Exact bounds on the closeness between the Student and standard normal distributions. Version 2. http://arxiv.org/abs/1101.3286 (2013).
[6] I. Pinelis and R. Molzon, Berry-Esséen bounds for general nonlinear statistics, with applications to Pearson’s and non-central Student’s and Hotelling’s. http://arxiv.org/abs/math/0701806 (2009).
[7] R.S. Pinkham and M.B. Wilk, Tail areas of the \( t \)-distribution from a Mills’-ratio-like expansion. Ann. Math. Statist. 34 (1963) 335–337.
[8] H. Scheffé, A useful convergence theorem for probability distributions. Ann. Math. Statist. 18 (1947) 434–438.
[9] R. Shimizu, Error bounds for asymptotic expansion of the scale mixtures of the normal distribution. Ann. Inst. Statist. Math. 39 (1987) 611–622.