The value at the mode in multivariate $t$ distributions: a curiosity or not?

Christophe Ley and Anouk Neven

Abstract: It is a well-known fact that multivariate Student $t$ distributions converge to multivariate Gaussian distributions as the number of degrees of freedom $\nu$ tends to infinity, irrespective of the dimension $k \geq 1$. In particular, the Student’s value at the mode (that is, the normalizing constant obtained by evaluating the density at the center) $c_{\nu,k} = \frac{\Gamma\left(\frac{\nu+k}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\pi\nu}{(\pi\nu)^{k/2}}\right)^{1/2}$ converges towards the Gaussian value at the mode $c_k := \frac{1}{(2\pi)^{k/2}}$. In this note, we prove a curious fact: $c_{\nu,k}$ tends monotonically to $c_k$ for each $k$, but the monotonicity changes from increasing in dimension $k = 1$ to decreasing in dimensions $k \geq 3$ whilst being constant in dimension $k = 2$. A brief discussion raises the question whether this a priori curious finding is a curiosity, in fine.

AMS 2000 subject classifications: Primary 62E10; secondary 60E05.

Keywords and phrases: Gamma function, Gaussian distribution, value at the mode, Student $t$ distribution, tail weight.

1. Foreword.

One of the first things we learn about Student $t$ distributions is the fact that, when the degrees of freedom $\nu$ tend to infinity, we retrieve the Gaussian distribution, which has the lightest tails in the Student family of distributions. It is also well-known that the mode of both distributions lies at their center of symmetry, entailing that the value at the mode simply coincides with the corresponding Student $t$ and Gaussian normalizing constants, respectively, and that this value $c_{\nu,k}$ of a $k$-dimensional Student $t$ distribution with $\nu$ degrees of freedom tends to $c_k := \frac{1}{(2\pi)^{k/2}}$, the value at the mode of the $k$-dimensional Gaussian distribution. Now ask yourself the following question: does this convergence of $c_{\nu,k}$ to $c_k$ take place in a monotone way, and would the type of monotonicity (increasing/decreasing) be dimension-dependent? All the statisticians we have asked this question (including ourselves) were expecting monotonicity and nobody could imagine that the dimension $k$ could have an influence on the type of monotonicity, especially because everybody was expecting Gaussian distributions to always have the highest value at the mode, irrespective of the dimension. However, as we shall demonstrate in what follows, this general idea about the absence of effect by the dimension is wrong. Although the Student $t$ distribution is well-studied in the literature, this convergence has nowhere been established to the best of the authors’ knowledge. This, combined with the recurrent astonishment when speaking about the dimension-dependent monotonicity, has led us to writing the present note.

2. Introduction.

Though already introduced by Helmert (1875), Lüroth (1876) and Pearson (1895), the (univariate) $t$ distribution is usually attributed to William Sealy Gosset who, under the pseudonym Student in 1908 (see Student 1908), (re-)defined this probability distribution, whence the commonly used expression...
Student \( t \) distribution. This terminology has been coined by Sir Ronald A. Fisher in Fisher (1925), a paper that has very much contributed to making the \( t \) distribution well-known. This early success has motivated researchers to generalize the \( t \) distribution to higher dimensions; the resulting multivariate \( t \) distribution has been studied, \textit{inter alia}, by Cornish (1954) and Dunnett and Sobel (1954). The success story of the Student \( t \) distribution yet went on, and nowadays it is one of the most commonly used absolutely continuous distributions in statistics and probability. It arises in many situations, including e.g. the Bayesian analysis, estimation, hypotheses testing and modeling of financial data. For a review on the numerous theoretical results and statistical aspects, we refer to Johnson \textit{et al.} (1994) for the one-dimensional and to Johnson and Kotz (1972) for the multi-dimensional setup.

Under its most common form, the \( k \)-dimensional \( t \) distribution admits the density
\[
f_k^\nu(x) := c_{\nu,k} \left(1 + \|x\|^2/\nu\right)^{-\frac{\nu+k}{2}}, \quad x \in \mathbb{R}^k,
\]
with tail weight parameter \( \nu \in \mathbb{R}_0^+ \) and normalizing constant
\[
c_{\nu,k} = \frac{\Gamma\left(\frac{\nu+k}{2}\right)}{(\pi\nu)^{k/2}\Gamma\left(\frac{k}{2}\right)},
\]
where the Gamma function is defined by
\[
\Gamma(z) = \int_{0}^{\infty} \exp(-t)t^{z-1} \, dt.
\]
The kurtosis of the \( t \) distribution is of course regulated by the parameter \( \nu \): the smaller \( \nu \), the heavier the tails. For instance, for \( \nu = 1 \), we retrieve the fat-tailed Cauchy distribution. As already mentioned, of particular interest is the limiting case when \( \nu \) tends to infinity, which yields the multivariate Gaussian distribution with density
\[
(2\pi)^{-k/2} \exp\left(-\frac{1}{2}\|x\|^2\right), \quad x \in \mathbb{R}^k.
\]
The \( t \) model thus embeds the Gaussian distribution into a parametric class of fat-tailed distributions. Indeed, basic calculations show that
\[
\lim_{\nu \to \infty} \left(1 + \|x\|^2/\nu\right)^{-\frac{\nu+k}{2}} = \exp\left(-\frac{1}{2}\|x\|^2\right)
\]
and \( \lim_{\nu \to \infty} c_{\nu,k} = (2\pi)^{-k/2} = c_k \). It is to be noted that \( c_{\nu,k} \) and \( c_k \) respectively correspond to the value at the mode (that is, at the origin) of \( k \)-dimensional Student \( t \) and Gaussian distributions.

In the next section, we shall prove that \( c_{\nu,k} \) converges monotonically towards \( c_k \), in accordance with the general intuition, but, as we shall see, this monotonicity heavily depends on the dimension \( k \): in dimension \( k = 1 \), \( c_{\nu,k} \) increases to \( c_k \), in dimension \( k = 2 \) we have that \( c_{\nu,2} = c_2 \) while for \( k \geq 3 \) \( c_{\nu,k} \) decreases towards \( c_k \). Stated otherwise, the probability mass around the center increases with \( \nu \) in the one-dimensional case whereas it decreases in higher dimensions, an \textit{a priori} unexpected fact in view of the interpretation of \( \nu \) as tail-weight parameter. It is all the more surprising as the variance of each marginal Student \( t \) distribution equals \( \nu/(\nu-2) \) for \( \nu > 2 \) and any dimension \( k \), which is strictly larger than 1, the variance of the Gaussian marginals. An attempt for an explanation hereof is provided, raising the question whether this is a curiosity or not.

3. The (curious?) monotonicity result.

Before establishing our monotonicity results, let us start by introducing some notations that will be useful in the sequel. To avoid repetition, let us mention that the subsequent formulae are all valid for \( x \in \mathbb{R}_0^+ \).

Denoting by \( D_x := D_x^{(1)} \) the first derivative, define
\[
\psi(x) = D_x \log(\Gamma(x)),
\]
the so-called digamma function or first polygamma function. The well-known functional equation
\[ \Gamma(x + 1) = x \Gamma(x) \] (1)
thus allows to obtain, by taking logarithms and differentiating,
\[ \psi(x + 1) = \psi(x) + \frac{1}{x}. \] (2)
Another interesting and useful formula is the series representation of the derivatives of \( \psi \):
\[ D_x^n \psi(x) = \sum_{j=0}^{\infty} \frac{(-1)^{n+1} n!}{(x+j)^{n+1}}, \quad n \geq 1. \] (3)
For an overview and proofs of these results, we refer to Artin (1964).

Now, with these notations in hand, we are ready to state the main result of this paper, namely the announced monotonicity result of the normalizing constants \( c_{\nu,k} \).

**Theorem 1** (Dimension-based monotonicity of the normalizing constants in \( t \) distributions). For \( k \in \mathbb{N}_0 \), define the mapping \( g_k : \mathbb{R}_0^+ \to \mathbb{R}_0^+ \) by \( g_k(\nu) = \frac{\Gamma(\nu + k/2)}{(\pi \nu)^{k/2} \Gamma(k/2)} = c_{\nu,k} \). We have
(i) if \( k = 1 \), \( g_k(\nu) \) is monotonically increasing in \( \nu \);
(ii) if \( k = 2 \), \( g_k(\nu) \) is constant in \( \nu \);
(iii) if \( k \geq 3 \), \( g_k(\nu) \) is monotonically decreasing in \( \nu \).

**Proof.** (i) Basic calculus manipulations show that
\[ D_\nu (\log g_1(\nu)) = \frac{1}{2} \left( \psi \left( \frac{\nu + 1}{2} \right) - \psi \left( \frac{\nu}{2} \right) - \frac{1}{\nu} \right), \]
with \( \psi \) the digamma function. By (3), we know that
\[ D_x \psi(x) = \sum_{j=0}^{\infty} \frac{1}{(x+j)^2} \quad \text{and} \quad D_x^2 \psi(x) = -2 \sum_{j=0}^{\infty} \frac{1}{(x+j)^3}. \]
Thus, \( \psi \) is an increasing and concave function on \( \mathbb{R}_0^+ \). Using concavity together with identity (2), we have in particular
\[ \psi \left( \frac{\nu + 1}{2} \right) \geq \frac{1}{2} \psi \left( \frac{\nu}{2} \right) + \frac{1}{2} \psi \left( \frac{\nu}{2} + 1 \right) = \psi \left( \frac{\nu}{2} \right) + \frac{1}{\nu}. \]
This inequality readily allows us to deduce that \( \log g_1(\nu) \), and hence \( g_1(\nu) \), is monotonically increasing in \( \nu \).

(ii) If \( k = 2 \), the function \( g_2(\nu) \) reduces to \( \frac{1}{2\pi} \) by simply applying (1), in other words it equals its limit, whence the claim.

(iii) Assume first that \( k \geq 3 \) is even, hence that \( k/2 \) is an integer. Using iteratively identity (1), we can write
\[ \Gamma \left( \frac{\nu}{2} + \frac{k}{2} \right) = \left( \frac{\nu}{2} + \frac{k}{2} - 1 \right) \left( \frac{\nu}{2} + \frac{k}{2} - 2 \right) \cdots \left( \frac{\nu}{2} \right) \frac{\Gamma \left( \frac{\nu}{2} \right)}{\frac{k}{2} \text{ factors}}, \]
which implies
\[ g_k(\nu) = \frac{\Gamma(\nu+k/2)}{(\pi \nu)^{k/2} \Gamma(k/2)} = \pi^{-k/2} \left( \frac{1}{2} + \frac{k/2 - 1}{\nu} \right) \left( \frac{1}{2} + \frac{k/2 - 2}{\nu} \right) \cdots \left( \frac{1}{2} + \frac{1}{\nu} \right) \frac{1}{2}. \]
By concavity of the digamma function, we have the inequality
\[ c \leq 0, \]
while the monotone convergence of the values \( g_k(\nu) \) happens to be the product of monotonically decreasing and positive functions in \( \nu \). Thus \( g_k(\nu) \) is itself monotonically decreasing in \( \nu \), which allows to conclude for \( k \) even.

Now assume that \( k \geq 3 \) is odd. We set \( k = 2m + 1 \) with \( m \in \mathbb{N}_0 \). The proof is based on the same idea as the proof for the one-dimensional case. One easily sees that
\[ D_\nu(\log g_k(\nu)) = \frac{1}{2} \left( \psi \left( \frac{\nu + k}{2} \right) - \psi \left( \frac{\nu}{2} \right) - \frac{k}{\nu} \right), \tag{4} \]
with \( \psi \) the digamma function. In the rest of this proof, we establish the monotonicity of \( g_k(\nu) \) by proving by induction on \( k \) (respectively, on \( m \)) that \( D_\nu(\log g_k(\nu)) \leq 0 \) for all \( \nu \in \mathbb{R}_0^+ \).

**Base case:** If \( m = 1 \) (which implies \( k = 3 \)), identity (2) yields \( \psi(\nu/2) = \psi(\nu/2 + 1) - 2/\nu \), hence (4) can be rewritten as
\[ \frac{1}{2} \left( \psi \left( \frac{\nu + 3}{2} \right) - \psi \left( \frac{\nu}{2} + 1 \right) \right) - \frac{1}{\nu} \geq 0. \tag{5} \]
By concavity of the digamma function, we have the inequality
\[ \psi \left( \frac{\nu}{2} + 1 \right) \geq \frac{1}{2} \left( \psi \left( \frac{\nu + 3}{2} \right) + \psi \left( \frac{\nu + 1}{2} \right) \right), \]
and thus (5) can be bounded by
\[ \frac{1}{2} \left( \frac{1}{2} \left( \psi \left( \frac{\nu + 3}{2} \right) - \psi \left( \frac{\nu + 1}{2} \right) \right) - \frac{1}{\nu} \right) = \frac{1}{2} \left( \frac{1}{\nu + 1} - \frac{1}{\nu} \right) = - \frac{1}{2\nu(\nu + 1)} < 0, \]
where we have again used (2). So \( D_\nu(\log g_3(\nu)) \leq 0 \) for all \( \nu \in \mathbb{R}_0^+ \), and the claim holds for the base case.

**Induction case:** Assume that the expression in (4) is negative for \( k = 2m + 1 \) with \( m \in \mathbb{N}_0 \). We now show that the claim is true for \( k' = 2(m + 1) + 1 = k + 2 \). It follows once more from (2), combined with the fact that \( \nu + k \geq \nu \), that
\[ \psi \left( \frac{\nu + k'}{2} \right) - \psi \left( \frac{\nu}{2} \right) - \frac{k'}{\nu} = \psi \left( \frac{\nu + k + 1}{2} \right) - \psi \left( \frac{\nu}{2} \right) - \frac{k}{\nu} - \frac{2}{\nu} \]
\[ = \psi \left( \frac{\nu + k}{2} \right) + \psi \left( \frac{\nu + k}{2} \right) - \psi \left( \frac{\nu}{2} \right) - \frac{k}{\nu} - \frac{2}{\nu} \]
\[ \leq \psi \left( \frac{\nu + k}{2} \right) - \psi \left( \frac{\nu}{2} \right) - \frac{k}{\nu} \]
\[ \leq 0, \]
where the final inequality is due to the induction hypothesis. Thus \( D_\nu(\log g_k(\nu)) \leq 0 \) for all odd \( k \geq 3 \), which concludes the proof. \( \square \)

**Remark 1.** One easily sees that the induction-based proof also holds under slight modifications for \( k \geq 4 \) even, but we prefer to show this shorter proof for the even-\( k \) case.

For the sake of illustration, we provide in Figure 1 the curves of the values at the mode \( g_k(\nu) = c_{\nu,k} \) for \( k = 1, 2, 3 \) and 4. The respective asymptotics of course correspond to the respective limits \( (2\pi)^{-k/2} \). While the monotone convergence of the values \( c_{\nu,k} \) to \( (2\pi)^{-k/2} \) is by no means surprising, the fact that this monotonicity changes from increasing in dimension \( k = 1 \) to decreasing in dimensions \( k \geq 3 \) whilst being constant in dimension \( k = 2 \) seems at first sight, as already mentioned previously, puzzling, as
one would expect the Gaussian to always have the highest peak at the center. That this is not the case is illustrated in Figure 2, where we have plotted, for distinct values of \( k \), several Student \( t \) densities with increasing degrees of freedom \( \nu \). In dimension \( k = 1 \), the Gaussian density has the highest peak, in dimension \( k = 2 \) all peaks have the same value, whereas for \( k \geq 3 \) the Gaussian has the lowest peak. The same conclusion thus also holds true for the probability mass around the center, which increases with \( \nu \) in dimension 1 and decreases for higher dimensions (whilst being nearly unchanged in dimension 2); see Table 1.

Why does this result seem so counter-intuitive? One reason is that, since the kurtosis of Student \( t \) distributions decreases when \( \nu \) increases, one would naturally expect that the probability mass around the center always increases with \( \nu \), all the more so as the variance-covariance of a \( k \)-dimensional Student distribution equals, for \( \nu > 2 \), \( \frac{k}{\nu - 2}I_k \), with \( I_k \) the \( k \)-dimensional identity matrix, hence the marginals’ variance is always larger as that of the Gaussian marginals.

This result is even more astonishing as the moment-based kurtosis ratio between two \( t \) distributions does not alter with the dimension. Indeed, letting, without loss of generality, \( \mathbf{X}_1 \) and \( \mathbf{X}_2 \) be \( k \)-variate random vectors following each a \( t \) distribution with respective parameters \( \nu_1 \) and \( \nu_2 \), straightforward
Fig 2. Plots of the $k$-variate $t$ densities $c_{\nu,k} \left(1 + \|x\|^2/\nu\right)^{-\frac{\nu+k}{2}}$ for dimensions (a) $k = 1$, (b) $k = 2$, (c) $k = 3$, and (d) $k = 4$. Within each sub-figure, we have chosen four values for the tail parameter $\nu$: 1 (blue curves), 2 (red curves), 10 (yellow curves) and $\infty$ (green curves). From dimension 2 onwards, the densities are plotted along the first vector of the canonical basis of $\mathbb{R}^k$. 
calculations show that, for \( m < \min(\nu_1, \nu_2) \),
\[
\frac{E[||X_1||^m]}{E[||X_2||^m]} = \frac{\int_0^\infty r^{m+k-1} c_{\nu_1,k}(1 + r^2/\nu_1)^{-\frac{\nu_1+k}{2}} dr}{\int_0^\infty r^{m+k-1} c_{\nu_2,k}(1 + r^2/\nu_2)^{-\frac{\nu_2+k}{2}} dr} = \frac{\nu_1^{m/2} \Gamma\left(\frac{\nu_1-k}{2}\right)}{\nu_2^{m/2} \Gamma\left(\frac{\nu_2-k}{2}\right)} \frac{\Gamma\left(\frac{\nu_1}{2}\right) \Gamma\left(\frac{\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1-k}{2}\right) \Gamma\left(\frac{\nu_2-k}{2}\right)},
\]
which does not depend on the dimension \( k \). This result can be summarized in the following proposition.

**Proposition 1.** Let \( X_1 \) and \( X_2 \) be \( k \)-variate random vectors following each a \( t \) distribution with respective parameters \( \nu_1 \) and \( \nu_2 \). Then, for all \( m < \min(\nu_1, \nu_2) \), the ratio \( \frac{E[||X_1||^m]}{E[||X_2||^m]} \) does not depend on the dimension \( k \). Consequently, the moment-based kurtosis ratio or fourth standardized moment ratio
\[
\frac{\beta_{\nu_1,k}}{\beta_{\nu_2,k}} := \frac{\frac{E[||X_1||^4]}{E[||X_1||^2]^2}}{\frac{E[||X_2||^4]}{E[||X_2||^2]^2}}
\]
is the same for each dimension \( k \), provided that \( \min(\nu_1, \nu_2) > 4 \).

This proposition seems to add further confusion about the result of Theorem 1. Why is our intuition so misleading? The reason lies most probably in our general understanding of heavy tails in high dimensions. While, in dimension 1, one can clearly observe the tail-weight by looking at the density curves far from the origin, this visualization vanishes more and more with the dimension. This waning difference is however thwarted by the increase in dimension, in the sense that the smaller difference in height between the density curves is integrated over a larger domain, which explains for instance the kurtosis ratio result of Proposition 1. It also explains why, in higher dimensions, the lighter-tailed distributions need not have the highest peaks at the mode, as this difference in peak has weak importance in view of the small domain over which it is integrated compared to the tails.

The dimension thus has an impact on the monotone convergence of the values at the mode \( c_{\nu,k} \) of multivariate Student \( t \) distributions towards the value at the mode \( c_k \) of multivariate Gaussian distributions. Curiosity or not?

**ACKNOWLEDGEMENTS:**

The research of Christophe Ley is supported by a Mandat de Chargé de Recherche du Fonds National de la Recherche Scientifique, Communauté française de Belgique. The research of Anouk Neven is supported by an AFR grant of the Fonds National de la Recherche, Luxembourg (Project Reference 4086487). Both authors thank Davy Paindaveine for interesting discussions on this special topic.

**References**

[1] Artin, E. (1964). *The Gamma function*. New York: Holt, Rinehart and Winston.

[2] Cornish, E. A. (1954). The multivariate \( t \)-distribution associated with a set of normal sample deviates. *Aust. J. Phys.* 7, 531–542.

[3] Dunnett, C. W., and Sobel, M. (1954). A bivariate generalization of Student’s \( t \)-distribution, with tables for certain special cases. *Biometrika* 41, 153–169.

[4] Fisher, R. A. (1925). Applications of “Student’s” distribution. *Metron* 5, 90–104.
[5] Helmert, F. R. (1875). ¨Uber die Bestimmung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler. Z. Math. Phys. 20, 300–303.

[6] Johnson, N. L., and Kotz, S. (1972). Distributions in Statistics: Continuous multivariate distributions. New York: Wiley.

[7] Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). Continuous univariate distributions, vol. 2. New York: Wiley, 2nd edition.

[8] Lüroth, J. (1876). Vergleichung von zwei Werten des wahrscheinlichen Fehlers. Astron. Nachr. 87, 209–220.

[9] Pearson, K. (1895). Contributions to the mathematical theory of evolution, II. Skew variation in homogeneous material. Philos. T. Roy. Soc. Lond. A 186, 343–414.

[10] Student (1908). The probable error of a mean. Biometrika 6, 1–25.