A Simple Analytic Approximation for Entropy of Student-\( t\) Distribution and its Relation with Normal Distribution

LAMBERT PIERRAT*\(^1\) and SAMUEL KOTZ\(^2\)

\(^1\) LJ - Consulting & LJK-LAB, UMR CNRS 5224, Grenoble Alps University, 21 Allée Jean Wiener, 38040 Saint Martin d’Hères Cedex, FRANCE

\(^2\) George Washington University, Washington DC, USA (Deceased)

(Received on June 24, 2014, revised on December 15, 2014)

Abstract: In the expression corresponding to the Shannon’s entropy of the Student-\( t\) distribution the gamma and digamma integral functions appear. We propose a simple analytical approximation for its entropy function for all degrees of freedom which assures the continuity between normal and Cauchy distributions. A possible application is to define a normal distribution “equivalent” of the Student-\( t\), usable for any degrees of freedom (integral or fractional) larger than 7.

Keywords: Student-\( t\) distribution, entropy, normal distribution, Cauchy distribution, standard deviation, analytical approximation.

1. Introduction

The Student-\( t\) distribution \([1]\) is of substantial popularity and importance in classical statistical inference as well as in applications of statistical methodology in science, engineering and especially metrology \([2]\). A classical case is the statistical estimation of a sample mean drawn from a population having a normal distribution. Because this estimator is Student-\( t\) distributed, it is often easier to approximate its confidence intervals by using an “equivalent” normal distribution. But such a simplified approach can be used only if the number degree of freedom is sufficiently large (currently: \(\nu > 30\)). The proposed approach based on entropy permits to define analytically an approximate equivalence for a smaller number degree of freedom (practically: \(\nu > 7\)).

2. An Exact Expression for the Entropy

Using the expression of Student-\( t\) probability density \([3]\):

\[
f_{\nu}(x) = \frac{1}{\sqrt{\nu \cdot B(\nu/2, 1/2)}} \left( \frac{x^2}{\nu} \right)^{\frac{\nu-1}{2}}
\]

One can calculate its information entropy introduced by Shannon \([4]\):

\[
H_{\nu} = -\int_{-\infty}^{\infty} f_{\nu}(x) \cdot \ln[f_{\nu}(x)] \cdot dx
\]

In this case, taking into account the number degrees of freedom (\(\nu \geq 1\)), the entropy is obtained analytically after integration:

\[
H_{\nu} = \left( \frac{\nu+1}{2} \right) \left[ \Psi \left( \frac{\nu+1}{2} \right) - \Psi \left( \frac{\nu}{2} \right) \right] + \ln \left( \sqrt{\nu \cdot B(\nu/2, (\nu/2))} \right)
\]

(3)

The “integral functions” \([5]\) appearing in the expression for \(H_{\nu}\) in (3) are the following:

a) The complete Beta function given by:
The Digamma function defined as the ratio:
\[ \Psi(z) = \frac{\Gamma'(z)}{\Gamma(z)} \]  

(5)

The parameter \( \nu \) is continuously varies in the range \( (\nu = 1) \), corresponding to the Cauchy distribution, up to \( (\nu = \infty) \), i.e. the case of the normal distribution.

3. An Approximate Analytic Expression for the Entropy

Using the exact expression (3) it is quite difficult to assess clearly the effect of the number degrees of freedom \( (\nu) \) on the entropy. Using asymptotic developments of the two preceding “integral functions”, direct but somewhat tedious calculations yield the following approximate expression:

\[
H'_n \approx \left( \frac{1}{\nu} \right) + L \left[ \frac{\sqrt{2\pi} \left( 7 \nu - 61 \right)}{7 \nu - 61} \right] + \ldots
\]

(6)

\[
\ldots + \left( \frac{\nu + 1}{2} \right) \left[ L \left( \frac{\nu + 3}{\nu + 2} \right) + \left( \frac{1}{(\nu + 2) (\nu + 3)} \right) \right] + \left( \frac{\sqrt{2\nu + 5}}{3 \cdot (\nu + 2) (\nu + 3)} \right)^2
\]

The relative precision of approximation (6) is at least \( 0.5 \times 10^{-5} \) for \( (\nu \geq 1) \).

Utilizing (6) - after some algebra manipulations, truncations and simplifications - one can arrive at a simpler continuous approximation:

\[
H'_n \approx L \left( \frac{9\nu + 2}{1 \nu 8} \right)
\]  

(7)

This one has a relative precision of at least \( 0.5 \times 10^{-2} \) for \( (\nu \geq 1) \).

As \( (\nu \to \infty) \) approximation (7) converges to true exact value of the entropy corresponding to the standardized normal distribution \[6\]:

\[
H'_n (\nu \to \infty) = H_n = L \frac{2\pi}{e}
\]  

(8)

For \( (\nu = 1) \) we obtain an approximation which is extremely close to the exact value of the entropy of the Cauchy distribution \[7\]:

\[
H'_n (\nu = 1) = H_n = L (4\pi) = 2.5 \ldots
\]  

(9)

From (7) the result is \( (H_n \approx 2.530\ldots) \), that corresponds to a relative error inferior to \( 4 \times 10^{-4} \). Thus the precision for both extreme approximations \( (\nu = 1) \) and \( (\nu = \infty) \) and hence for any value of \( (\nu) \) as given by (7) is at least 0.6 %. This is evidently sufficient for most practical applications taking into account the simplicity of the formula (7) and the continuity from \( (\nu = 1) \) to \( (\nu = \infty) \) for all intermediate values of \( (\nu) \), both integral and fractional.

Note that the Student-t distribution is currently used for estimating the variance of a normal mean issued from a sample of \( (n \geq 2) \) independent observations (in this case, \( n = \nu - 1 \)). Because the observation number is an integer, the printed tables of the Student-t quantiles \[8\] are currently arranged in an ascending order \( (\nu = 1,2,3,\ldots,100,\infty) \). But in the context of metrology \[2\], if the considered variance is the sum of two or more different variance components, its distribution can be approximated by a Student-t having a non integer degree of freedom calculated from the so-called Welch-Satterthwaite formula \[2, 9\].
4. “Equivalent” Normal Distribution

One of applications of the Shannon entropy in statistical work, popularized by [10] is to utilize the entropy for the purpose of approximating the variance of a distribution by the variance of an “equivalent” normal distribution (in the entropy sense).

Let’s consider the Student-$t$ distribution whose standard deviation is:

$$\sigma_t = \sqrt{\frac{\nu}{\nu - 2}}$$

(10)

Note that this one is commensurable only for ($\nu > 2$). The variance of the “equivalent” normal distribution can explicitly be determined using the corresponding entropy expressions.

Indeed the entropy of the standardized normal distribution is:

$$H_n = L \left( \sigma_n \cdot\sqrt{2\pi} \cdot e \right)$$

(11)

Equalling (7) and (11) leads to the very simple relationship:

$$\sigma_n \approx e^{\frac{2}{\nu - 8}}$$

(12)

It is of interest to compare ($\sigma_t$) given in (10) with the “normally equivalent” ($\sigma_n$) given in (12) as a function of ($\nu$). This is presented in Table 1.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_t$</td>
<td>-</td>
<td>-</td>
<td>1.732</td>
<td>1.291</td>
<td>1.183</td>
<td>1.118</td>
<td>1.054</td>
<td>1.035</td>
<td>1.021</td>
<td>1.000</td>
</tr>
<tr>
<td>$\sigma_n$</td>
<td>3.038</td>
<td>1.743</td>
<td>1.448</td>
<td>1.249</td>
<td>1.172</td>
<td>1.118</td>
<td>1.057</td>
<td>1.038</td>
<td>1.022</td>
<td>1.000</td>
</tr>
<tr>
<td>$\varepsilon$ (%)</td>
<td>-</td>
<td>-</td>
<td>-16</td>
<td>-3.3</td>
<td>-0.93</td>
<td>±0.0</td>
<td>+0.3</td>
<td>+0.3</td>
<td>+0.1</td>
<td>±0.0</td>
</tr>
</tbody>
</table>

Note that for ($\nu > 7$) the values of ($\sigma_t$) and ($\sigma_n$) practically coincide (the relative error is less than 1%) knowing that the Student-$t$ distribution converges to the normal distribution as ($\nu \to \infty$). Recall, however, that a practical rule of statistics is to approximate the Student-$t$-distribution by a normal distribution for the values of ($\nu > 30$). This numerical convergence of variances seems to be faster than that of the distributions ones.

Because the variance of the Student-$t$ distribution for ($\nu = 1 \& 2$) is incommensurable, one could compare the 97.5 % quantiles of the two distributions which correspond to the superior bound of a bilateral confidence interval of significance 95 % (equivalent to an exceeding risk of 5 %). The corresponding values of ($q_{0.975}$) and ($q_{0.025}$) are presented in Table 2.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_t$</td>
<td>-</td>
<td>-</td>
<td>1.732</td>
<td>1.291</td>
<td>1.183</td>
<td>1.118</td>
<td>1.054</td>
<td>1.035</td>
<td>1.021</td>
<td>1.000</td>
</tr>
<tr>
<td>$\sigma_n$</td>
<td>3.038</td>
<td>1.743</td>
<td>1.448</td>
<td>1.249</td>
<td>1.172</td>
<td>1.118</td>
<td>1.057</td>
<td>1.038</td>
<td>1.022</td>
<td>1.000</td>
</tr>
<tr>
<td>$\varepsilon$ (%)</td>
<td>-</td>
<td>-</td>
<td>-16</td>
<td>-3.3</td>
<td>-0.93</td>
<td>±0.0</td>
<td>+0.3</td>
<td>+0.3</td>
<td>+0.1</td>
<td>±0.0</td>
</tr>
</tbody>
</table>

Observe that for ($\nu < 5$), in particular for ($\nu \leq 2$) the two quantiles are quite different due to the heavy tails of the Student-$t$ distribution for these degrees of freedom (as compared to the standard normal distribution) especially for the limit case of the Cauchy distribution (i.e., $\nu = 1$). But according to the proximity of their standard deviations, a sufficient close value of their quantiles is obtained for ($\nu > 7$).
5. Conclusion

An application of the proposed approach based on equality of entropies is the possibility to define an “equivalent” normal distribution of a Student-\(t\) distribution. Such a very simple and approximate “equivalence” is practically usable for any degree of freedom (integral or fractional), starting from a low value (\(\nu > 7\)).

Table 2: Values of Student-\(t\)-Quantile (\(q_{t,0.975}\)) and Its Equivalent Normal (\(q_{n,0.975}\))

<table>
<thead>
<tr>
<th>(\nu)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>(q_{t,0.975})</td>
<td>12.71</td>
<td>4.303</td>
<td>3.182</td>
<td>2.571</td>
<td>2.365</td>
<td>2.228</td>
<td>2.086</td>
</tr>
<tr>
<td>(q_{n,0.975})</td>
<td>5.955</td>
<td>3.416</td>
<td>2.838</td>
<td>2.448</td>
<td>2.297</td>
<td>2.191</td>
<td>2.072</td>
</tr>
<tr>
<td>(q_n / q_t)</td>
<td>0.469</td>
<td>0.794</td>
<td>0.892</td>
<td>0.952</td>
<td>0.971</td>
<td>0.983</td>
<td>0.993</td>
</tr>
</tbody>
</table>

Acknowledgement: Working with entropy initiated in [10], was done jointly with Professor Samuel Kotz who passed away in March 2010, after the preparation of the paper. Author would like to thank the anonymous referees and the Editor who helped improve this short communication.

References


Lambert Pierrat is an electrical engineer graduated from National Polytechnic Institute of Grenoble, France. He is currently a consulting engineer with LJ-Consulting and a free invited researcher in the Proba-Stat Department of Laboratoire Jean Kuntzmann, CNRS, Grenoble Alps University, France. His current research interests concern physical and statistical modeling of component reliability.