

Na Kyeong Lee

John Meakin

John Singleton

16 November 2010



STUDENT'S T DISTRIBUTION LECTURE NOTES

I. Functional Form

The density function or pdf of Student's t distribution is:

$$f_X(x) = \frac{\Gamma[(k+1)/2]}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{(1 + \frac{x^2}{k})^{(k+1)/2}}$$

where Γ is the gamma function defined by $\Gamma(t) = \int_{-\infty}^{\infty} x^{t-1} e^{-x} dx$. Note that the only parameter is k , which must be > 0 for the function to be defined. The t's cumulative density function is given by:

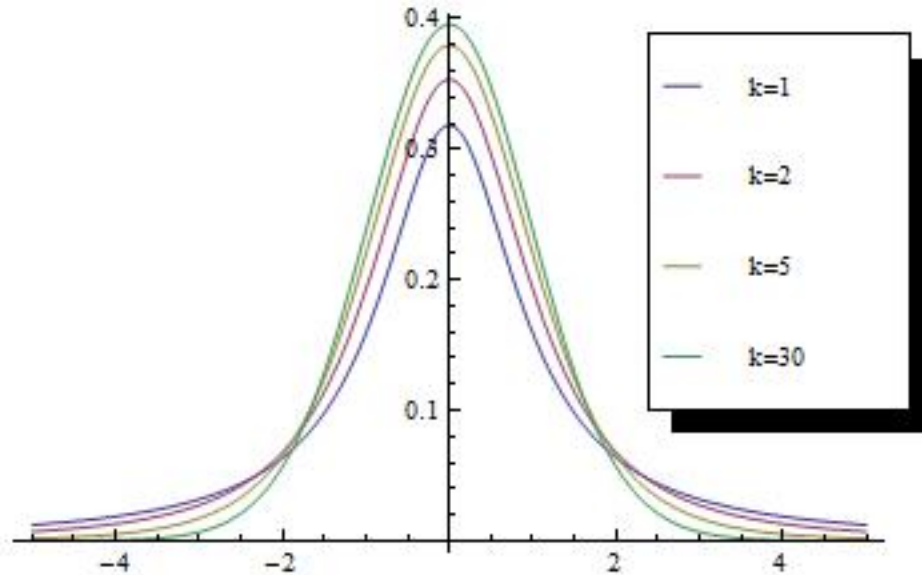
$$F_X(x) = \frac{1}{2} + \frac{x\Gamma(\frac{1}{2}(k+1)) {}_2F_1(\frac{1}{2}(k+1), \frac{(k+1)}{2}; \frac{3}{2}; -\frac{k^2}{k})}{\sqrt{\pi k}\Gamma(\frac{1}{2}k)}$$

where ${}_2F_1(a, b; c; z)$ is a hypergeometric function.¹ Both of these equations are rather messy, making it extremely difficult to intuit the shape or properties of the distribution. So, let's plot the pdf and cdf while varying k to get a feel for it and then analyze the distribution mathematically.

¹A generalized hypergeometric function can be expressed a multitude of ways. As shown by Euler, written as a function of gamma functions, ${}_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a} dt$. For additional information, see Weisstein's "Hypergeometric Function."

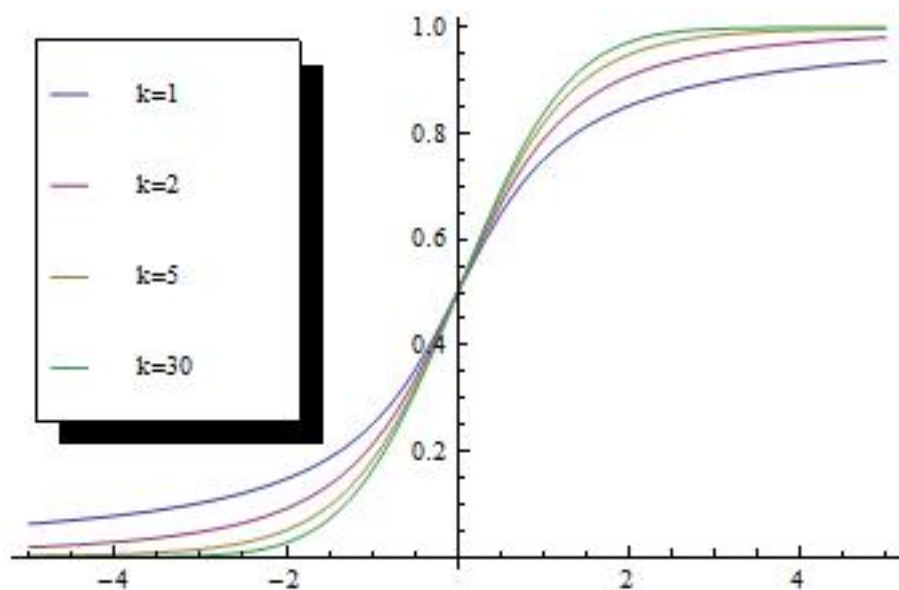
II. Visual

The picture below displays the density function or pdf of Student's t distribution for different values of k .



We can see that the function is defined for all real values of x ; it is continuous. Note that the distribution is centered at 0 and is symmetric for all values of k . If we were guessing based on the picture alone, we might guess that the mean, mode, median and skewness are all 0. As k increases from 1 to 30, the distribution's height becomes taller while the tails become flatter. In other words, the area under the tails shrinks and the area in the central part of the distribution increases. For the smaller values of k , the t distribution appears to be a "squatter" standard normal.

The following plot displays the cdf of the t for different values of k .



As k increases, the cdf becomes steeper around its inflection point while the tails become flatter. This reflects the observation made above that the area under the central part of the density function increases as k increases. For example, the cdf shows that $x = 1$ is associated with increasingly larger increases in probability from $x = 0$ as k increases.

III. Moments

We surmised by visual inspection that the mean and skewness of the t distribution are both 0. Can we show this mathematically? And, moreover, what are its other moments, like the variance and kurtosis?

We might begin finding the t distribution's properties by forming the moment generating function and evaluating its first derivative at 0.

$$m(t) = E(e^{tx}) = \int e^{tx} \frac{\Gamma[(k+1)/2]}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{(1 + \frac{x^2}{k})^{(k+1)/2}} dx$$

However, the above integral is not defined, so $m(t)$ does not exist. Instead, we could proceed by evaluating $E[X]$ to find the mean and $E[(x - E[X])^r]$ to find the r th central moment of the t . Unfortunately, these integrals are very difficult to evaluate directly, so their derivation is not presented here.

The mean of the t distribution is 0 where $k > 1$, the variance is $k/(k - 2)$ where $k > 2$, its skewness is 0 for $k > 3$, and the kurtosis is $\frac{6}{k-4}$ for $k > 4$. Higher moments can be determined by using the following function:

$$E[X^r] = \left\{ \begin{array}{ll} 0 & r \text{ odd, } 0 < r < k \\ \frac{\Gamma(\frac{r+1}{2})\Gamma(\frac{k-r}{2})k^{r/2}}{\pi\Gamma(\frac{k}{2})} & r \text{ even, } 0 < r < k \\ \text{undefined} & r \text{ odd, } 0 < k \leq r \\ \infty & r \text{ even, } 0 < k \leq r \end{array} \right\}$$

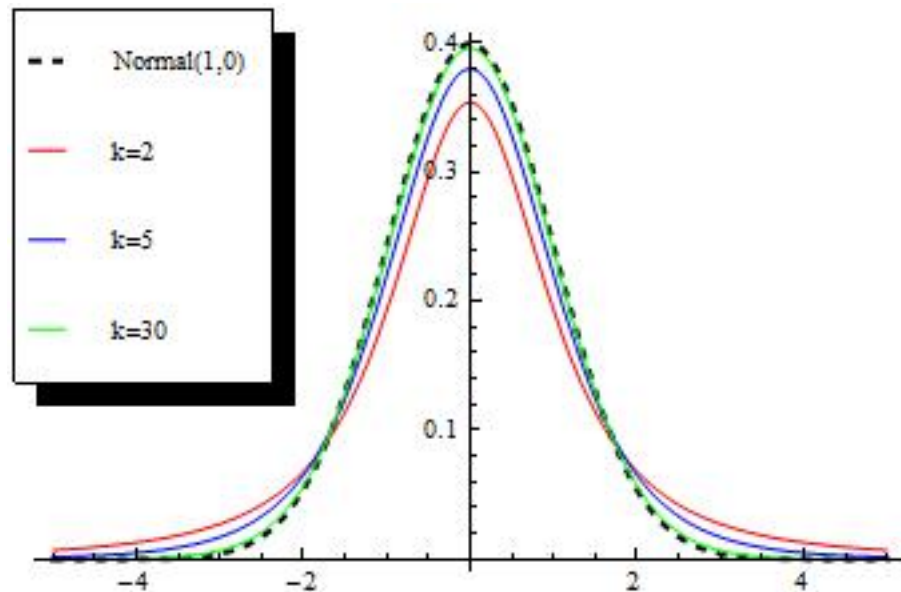
These moments verify that the t distribution is symmetric and centered at 0 for all values of k . Additionally, they are suggestive about the t 's relationship to the standard normal distribution.

IV. Relationship to Other Distributions

a. Standard Normal

As we've seen, the pdf of the t distribution is symmetric, "bell-shaped," and centered at 0. As a result, its general shape is very similar to that of the standard normal distribution for large values of k . In fact, we can show graphically and mathematically that as k increases,

the t distribution approaches the standard normal. The image below plots the standard normal distribution and the t distribution with increasing values of k .



Note that when $k = 30$, shown by the green line, the two distributions are nearly identical. For $k < 30$, the tails of the t distribution are fatter and its height lower than the standard normal's, indicating extreme values are more likely. A numerical example demonstrates this relationship: The probability that $x \leq 2$ in the standard normal, $\Phi(2) = .9772$. For the t distribution, when $k = 2$, the probability that $x \leq 2 = .9082$. With 30 degrees of freedom, however, the probability increases to .9727.

By evaluating the limits of the t distribution's variance and kurtosis as k goes to infinity, we can conclude that the t and standard normal have the same first four moments. However, this is not sufficient to prove that the standard normal is a special case of the t.

Note, though, that:

$$\lim_{k \rightarrow \infty} \frac{\Gamma(k + \frac{1}{2})}{\sqrt{2n\pi}\Gamma(k)} = \frac{1}{\sqrt{2\pi}}$$

and

$$\lim_{k \rightarrow \infty} \left(1 + \frac{x^2}{k}\right)^{-(\frac{k+1}{2})} = e^{-x^2/2}$$

Using the above two limit results, as $k \rightarrow \infty$, we could show that the pdf of the t distribution converges to the pdf $\phi(x)$ of the standard normal distribution for every value of x ($-\infty < x < \infty$). As a result, the standard normal is a special case of Student's t distribution.

b. Cauchy

Similarly, the restricted Cauchy distribution where $\alpha = 0$ and $\beta = 1$ is a special case of Student's t where $k = 1$. We prove this result below.

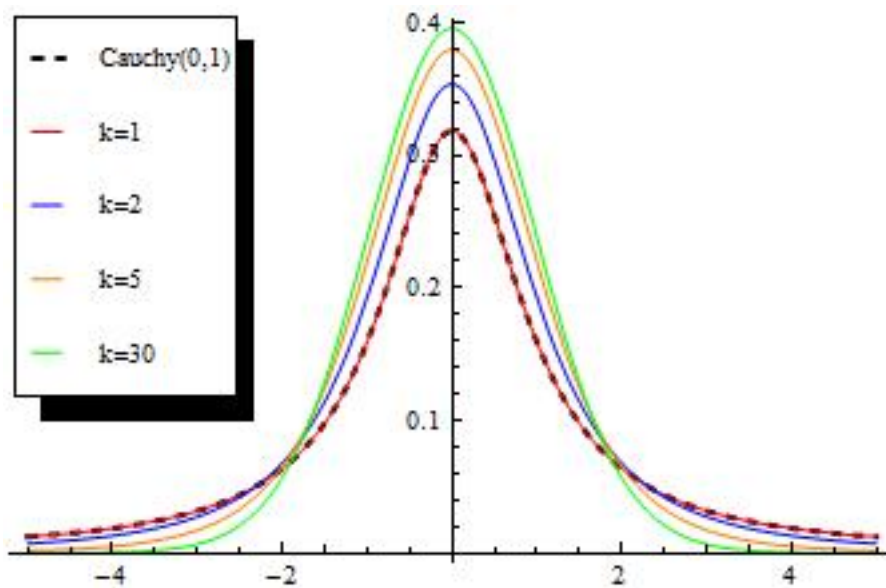
The pdf of the Cauchy when $\alpha = 0$ and $\beta = 1$ is given by:

$$f(x) = \frac{1}{\pi(1 + x^2)}$$

When $k = 1$, the pdf of the t distribution simplifies to:

$$f(x) = \frac{1}{\Gamma(\frac{1}{2})\sqrt{\pi}(1 + x^2)}$$

Since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, this is exactly the pdf of the Cauchy where $\alpha = 0$ and $\beta = 1$. This result can also be shown visually. The image below plots this restricted Cauchy distribution along with the t distribution with decreasing values of k .



We can see that when $k = 1$, shown by the red line, the two distributions overlap completely.

c. Noncentral Student's t

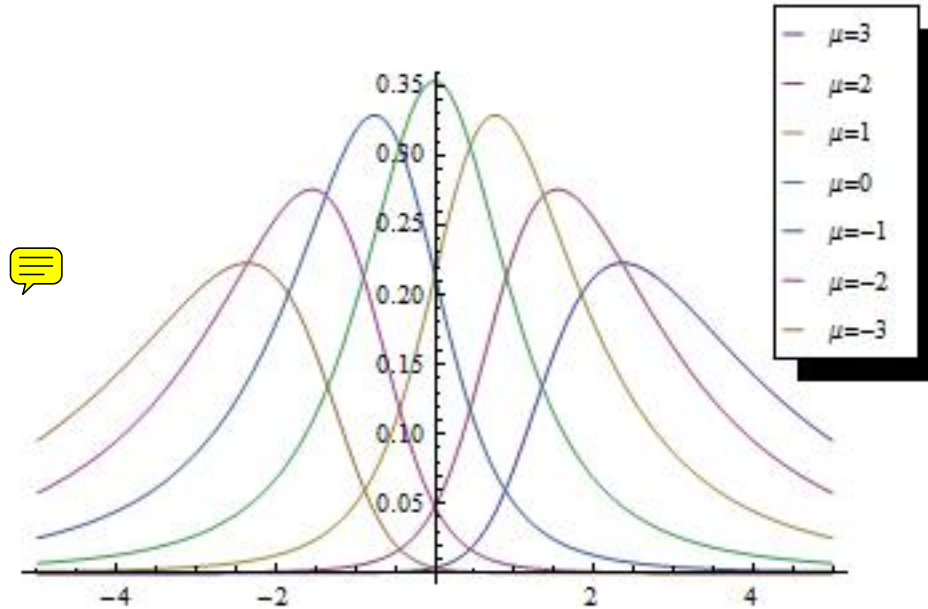
The Noncentral Student's t distribution's relationship to the t distribution is implied by its name: it possesses an additional parameter, μ , called the noncentrality parameter, that translates the distribution along the x-axis such that the mean is not necessarily fixed at 0. The Noncentral t distribution is considered a t distribution because its derivation is analogous to that of the central t; it is a generalization of the t.² Its pdf is given by:

$$f_X(x) = \frac{k^{k/2} k!}{2^k e^{\lambda^2/2} (k+x^2)^{k/2} \Gamma(\frac{1}{2}k)} \left\{ \frac{\sqrt{2}\lambda x {}_1F_1(\frac{1}{2}k+1; \frac{3}{2}; \frac{\lambda^2 x^2}{2(k+x^2)})}{(k+x^2)\Gamma[\frac{1}{2}(k+1)]} + \frac{{}_1F_1(\frac{1}{2}(k+1); \frac{1}{2}; \frac{\lambda^2 x^2}{2(k+x^2)})}{\sqrt{k+x^2}\Gamma(\frac{1}{2}k+1)} \right\}$$

²The random variable with a noncentral t distribution is defined as $\frac{Z+\mu}{\sqrt{U/k}}$ where, as before, Z is distributed standard normal and U is distributed chi-squared with degrees of freedom k .

where $\Gamma(z)$ is the gamma function and ${}_1F_1(a; b; z)$ is the confluent hypergeometric function.³

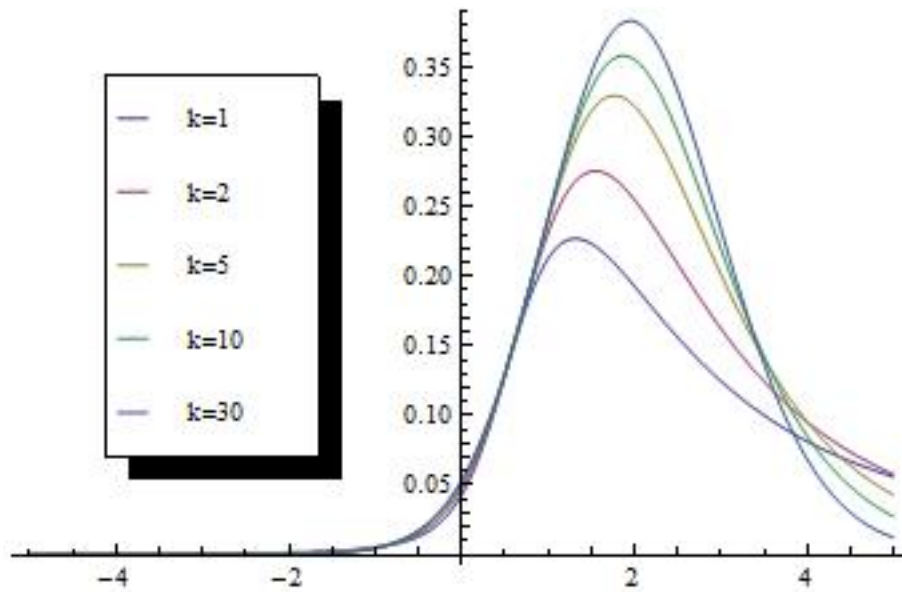
To investigate its relationship to the t further, we plot it. In the picture below, k is fixed at 2 and we allow μ to vary.



The picture confirms that the noncentrality parameter shifts the distribution along the x-axis. For example, when $\mu = 2$, the distribution is centered on a point slightly less than 2. Moreover, for nonzero values of μ , the distribution is slightly skewed and asymmetrical, but reduces to the central t when $\mu = 0$. Note that when $\mu = 3$, the right tail of the distribution is visibly fatter than the left tail. The distribution appears to become more skewed as μ increases in absolute value, k held constant. How does the distribution change when μ is fixed and we vary k ?

The following picture holds $\mu = 2$ while varying the degrees of freedom.

³The 1 preceding the F indicates that the first argument is just a, whereas when 2 preceded F the first argument was (a,b). See Paoletta (2007, 193) for additional information.



As k increases, the right tails becomes less and less fat. In fact, the distribution becomes increasingly symmetrical. The picture suggests that as k goes to infinity, the skewness approaches 0 regardless of μ .

V. Application

The following is an example review question where the data-generating process could be modelled by a t-distribution:

a. Imagine you are the owner of an ice cream shop. When a customer pays for an ice cream treat, you enter the price on the cash register, the amount of cash the customer paid you, and then the cash register tells you how much change to return to them. The cash register keeps a record of each day's revenue. At the end of each day, you must count the money in the cash register. Let X be a random variable defined as the difference between the cash register's record for revenue that day and the amount of money you count in the register at

the end of the day. If no counting errors are made tendering change to customers (or they cancel out), the difference is 0. Nonetheless, frequently the difference is nonzero. Sometimes it is negative and sometimes it is positive; no discernible pattern exists and double digit values are quite rare. However differences greater than 2 dollars occur more than 10 percent of the time. Construct a density function for X .



b. Now imagine you hire two employees, George and Margaret, who only work the cash register. George works the register 60 percent of the hours the store is open and Margaret works the register the other 40 percent. Additionally, Margaret is a little spacey; she is much more prone to making mistakes than George. Moreover, on sunny days, there are frequent ice cream rushes and the shop is very busy. When this is the case, Margaret tends to make even more mistakes. Let $S = 1$ if it's a sunny day and 0 otherwise. With this additional information, construct a density function for X and determine its variance. List the parameters of your density function.

In the above example, the t distribution is a good candidate for modelling the distribution of the random variable described for a few reasons. In the first place, all differences of real values are theoretically possible, so the density function is continuous.⁴ Secondly, the expected value of the data-generating process is 0. Thirdly, positive and negative differences are equally likely, so a symmetrical distribution, like the t , is required. Moreover, large differences like double digit values are rare. Finally, the t distribution better describes the data generating process than the standard normal since values greater than 2 have an associated probability greater than $\frac{1}{10}$. This piece of information indicates that extreme



⁴This is not exactly true. Money is generally denominated only to the second decimal place.

values are more likely than with the standard normal.⁵



VI. Origin

On page 250, Mood, Graybill and Boes (1974) state:

Theorem 10: If Z has a standard normal distribution, if U has a chi-squared distribution with k degrees of freedom, and if Z and U are independent, then $\frac{Z}{\sqrt{U/k}}$ has a Student's t distribution with k degrees of freedom.

This theorem hints at the origin of the distribution. Indeed, it was "Student" that proved Theorem 10 and derived the pdf of the t distribution in a paper published in *Biometrika* in 1908. His name, however, was not actually Student, but William Sealy Gosset. Here is a picture of him:



Per the terms of his employment with Guinness, Gosset was prohibited from publishing under his own name. This policy had been adopted in response to a previous Guinness employee divulging brewery secrets. So Gosset picked "Student" as his pseudonym. His buddy, R.A.

⁵Certainly, other density functions exist that share the characteristics listed that make the t a good candidate in this case, including the normal and generalized hyperbolic distributions. Note, however, that the t distribution is a special case of both of these, whereas the standard normal is, as we have shown, a special case of the t .



Fisher, popularized this distribution, referring to it as "Student's distribution," and the rest is history.

Additionally, the theorem attests to the t distribution's origin as the sampling distribution of a test statistic. It also provides a simpler procedure for determining its mean and variance by defining $X = \frac{Z}{\sqrt{U/k}}$ so that X has a Student's t distribution and evaluating $E[X]$ and $var(X) = E[X^2]$.⁶

VII. Derivation

The following proves Theorem 10 and derives the density function of the t distribution:

Let Z and U be independent random variables, where Z has a standard normal distribution and U is distributed chi-squared with k degrees of freedom. Since U and Z are independent, their joint pdf $f_{U,Z}(u, z)$ is equal to the product $f_U(u)f_Z(z)$, where $f_U(u)$ is the pdf of the χ^2 distribution with k degrees of freedom, and $f_Z(z)$ is the pdf of the standard normal distribution:

$$f_{U,Z}(u, z) = \frac{1}{\Gamma(k/2)} (1/2)^{k/2} u^{k/2-1} e^{-(1/2)u} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

In order to derive the desired density function, we use the transformation technique, as described by Mood, Graybill and Boes (1974, 198-206) and applying Theorem 13 on page 205.⁷

This allows us to obtain the pdf of $\frac{Z}{\sqrt{U/k}}$ from $f_{U,Z}(u, z)$ by performing a transformation.

⁶Since Z and U are independent and $E[Z] = 0$, $E[X] = 0$ where $k > 1$. In addition, it can be shown that X^2 has an F distribution, so $var(X) = E[X^2] = k/(k-2)$ where $k > 2$.

⁷The authors derive the t distribution's density function on pages 249-250.

Define X by $\frac{Z}{\sqrt{U/k}}$ and let $W = U$, so by rearranging:

$$Z = X\sqrt{W/k}$$

The Jacobian of the one-to-one transformation from X and W to U and Z is given by:

$$J = \begin{bmatrix} \frac{dZ}{dX} & \frac{dZ}{dW} \\ \frac{dU}{dX} & \frac{dU}{dW} \end{bmatrix}$$

so $|J| = \frac{w}{k}^{\frac{1}{2}}$, which is nonzero. Thus, the joint pdf $f_{X,W}(x, w)$ is as follows, for $-\infty < x < \infty$ and $w > 0$:

$$f_{X,W}(x, w) = |J|f_{U,Z}(u, z)$$

By evaluating this expression and determining the marginal density of X , we will obtain our desired density function.

$$\begin{aligned} f_{X,W}(x, w) &= f(w)f\left(x\left(\frac{w}{k}\right)^{1/2}\right)\left(\frac{w}{k}\right)^{1/2} \\ &= cw^{(k+1)/2-1}\exp\left[-\frac{1}{2}\left(1 + \frac{x^2}{k}\right)w\right] \end{aligned}$$

where

$$c = [2^{(k+1)/2}(k\pi)^{1/2}\Gamma\left(\frac{k}{2}\right)]^{-1}$$

Thus, the marginal density function $f_X(x)$ of X is obtained from $f_{X,W}(x, w)$ by using the relation:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,W}(x, w)dw \\ &= c \int_0^{\infty} w^{(k+1)/2-1} \exp[-wh(x)]dw \end{aligned}$$

where $h(x) = [1 + x^2/k]/2$. Evaluating this,

$$f_X(x) = c \frac{\Gamma((k+1)/2)}{h(x)^{(k+1)/2}} = \frac{\Gamma(\frac{k+1}{2})}{(k\pi)^{1/2}\Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-(k+1)/2}$$

for $-\infty < x < \infty$.

This is the density function of the t distribution given above. Therefore, we have shown that X has a t distribution with k degrees of freedom.

VIII. Common Usage

In applied statistics, the primary application of Student's t distribution is for determining confidence intervals and hypothesis testing. Since in applied settings random variables are frequently presumed to have normal distributions and in practice the population variance is never known, the sample variance is substituted to form a test statistic of interest. This resulting statistic can be shown to have a t distribution and the hypothesis test is known as the t-test. In addition, when a variable is distributed normal but the sample size is relatively small, the t distribution produces a more accurate confidence interval. The t distribution, therefore, has significant practical applications.

IX. Bibliography

- Bertsekas, Dimitri P., John N. Tsitsiklis. *Introduction to Probability* 2nd ed. Belmont, MA: Athena Scientific, 2008.
- Burnett, Brian, Sam Butler, Greg Martin, and Jeff Merrell. "Discussion Notes for the t-distribution." On file. (n.d.)
- DeGroot, Morris H., Mark J. Schervish. *Probability and Statistics*. 3rd ed. Reading, MA: Addison-Wesley, 2002.
- Mood, Alexander M., Franklin A. Graybill, and Duane C. Boes. *Introduction to the Theory of Statistics*. New York: McGraw-Hill, 1974.
- Paoletta, Marc S. *Intermediate Probability*. Wiley-InterScience, 2007.
- Student. "The Probable Error of a Mean." *Biometrika* 6 (1908): 1-25.
- Weisstein, Eric W. "Hypergeometric Function." MathWorld-A Wolfram Web Resource. <http://mathworld.wolfram.com/HypergeometricFunction.html>.
- Weisstein, Eric W. "Noncentral Student's t-Distribution." MathWorld-A Wolfram Web Resource. <http://mathworld.wolfram.com/NoncentralStudentst-Distribution.html>.
- Weisstein, Eric W. "Student's t-Distribution." MathWorld-A Wolfram Web Resource. <http://mathworld.wolfram.com/Studentst-Distribution.html>.
- Ziliak, Stephen T. "Guinnessometrics: The Economic Foundation of "Student's" t" *Journal of Economic Perspectives* 22, no. 4 (2008): 199-216.