



## Contents

---

[move to sidebar](#)[hide](#)

- (Top)
- 1History and etymology
- 2How Student's distribution arises from sampling
- 3Definition
  - 3.1Probability density function
  - 3.2Cumulative distribution function
  - 3.3Special cases
- 4How the *t*-distribution arises
  - 4.1Sampling distribution
  - 4.2Bayesian inference
- 5Characterization
  - 5.1As the distribution of a test statistic
    - 5.1.1Derivation
  - 5.2As a maximum entropy distribution
- 6Properties
  - 6.1Moments
  - 6.2Monte Carlo sampling
  - 6.3Integral of Student's probability density function and *p*-value
- 7Generalized Student's *t*-distribution
  - 7.1In terms of scaling parameter  $\hat{\sigma}$  or  $\hat{\sigma}^2$
  - 7.2In terms of inverse scaling parameter  $\lambda$
- 8Related distributions
- 9Bayesian inference: prior distribution for the degrees of the freedom

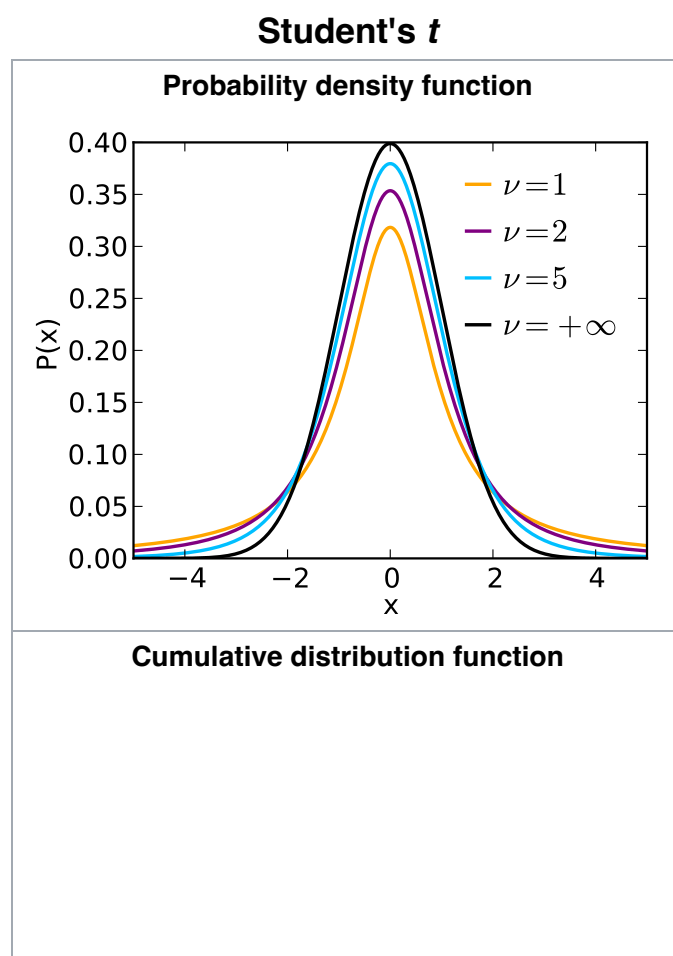
- 10Uses
  - 10.1In frequentist statistical inference
    - 10.1.1Hypothesis testing
    - 10.1.2Confidence intervals
    - 10.1.3Prediction intervals
  - 10.2In Bayesian statistics
  - 10.3Robust parametric modeling
- 11Student's *t*-process
- 12Table of selected values
- 13See also
- 14Notes
- 15References
- 16External links

# Student's *t*-distribution

In probability and statistics, **Student's *t*-distribution** (or simply the ***t*-distribution**) is any member of a family of continuous probability distributions that arise when estimating the mean of a normally distributed population in situations where the sample size is small and the population's standard deviation is unknown. It was developed by English statistician William Sealy Gosset under the pseudonym "Student".

The *t*-distribution plays a role in a number of widely used statistical analyses, including Student's *t*-test for assessing the statistical significance of the difference between two sample means, the construction of confidence intervals for the difference between two population means, and in linear regression analysis. Student's *t*-distribution also arises in the Bayesian analysis of data from a normal family.

If we take a sample of  $n$  observations from a normal distribution, then the *t*-distribution with  $\nu = n - 1$  degrees of freedom can be defined as the distribution of the location of the sample mean relative to the true mean, divided by the sample



standard deviation, after multiplying by the standardizing term  $\sqrt{n}$ . In this way, the  $t$ -distribution can be used to construct a confidence interval for the true mean.

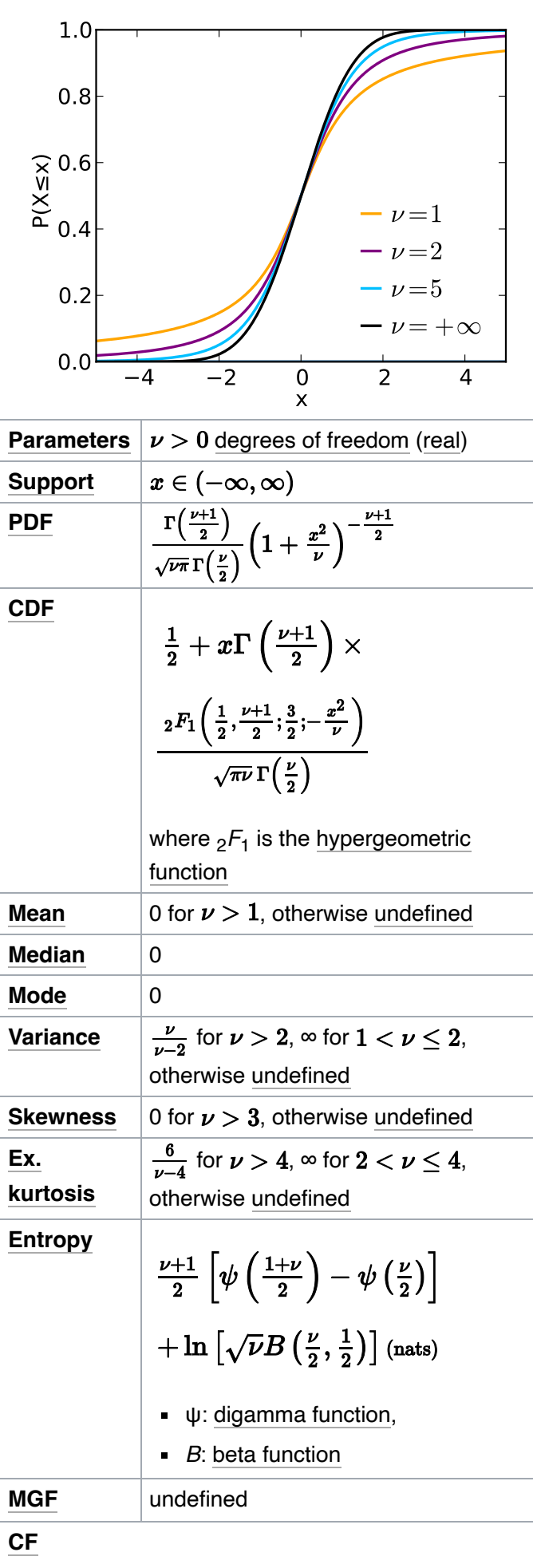
The  $t$ -distribution is symmetric and bell-shaped, like the normal distribution. However, the  $t$ -distribution has heavier tails, meaning that it is more prone to producing values that fall far from its mean. This makes it useful for understanding the statistical behavior of certain types of ratios of random quantities, in which variation in the denominator is amplified and may produce outlying values when the denominator of the ratio falls close to zero. The Student's  $t$ -distribution is a special case of the generalized hyperbolic distribution.

## History and etymology

In statistics, the  $t$ -distribution was first derived as a posterior distribution in 1876 by Helmert<sup>[2][3][4]</sup> and Lüroth.<sup>[5][6][7]</sup> The  $t$ -distribution also appeared in a more general form as Pearson Type IV distribution in Karl Pearson's 1895 paper.<sup>[8]</sup>

In the English-language literature, the distribution takes its name from William Sealy Gosset's 1908 paper in *Biometrika* under the pseudonym "Student".<sup>[9]</sup> One version of the origin of the pseudonym is that Gosset's employer preferred staff to use pen names when publishing scientific papers instead of their real name, so he used the name "Student" to hide his identity. Another version is that Guinness did not want their competitors to know that they were using the  $t$ -test to determine the quality of raw material.<sup>[10][11]</sup>

Gosset worked at the Guinness Brewery in Dublin, Ireland, and was interested in the problems of small samples – for example, the chemical properties of barley where sample sizes might be as few as 3. Gosset's paper refers to the distribution as the "frequency distribution of standard deviations of samples drawn from a normal population". It became well known through the work of Ronald



Fisher, who called the distribution "Student's distribution" and represented the test value with the letter  $t$ .<sup>[12][13]</sup>

## How Student's distribution arises from sampling

Let  $X_1, \dots, X_n$  be independently and identically drawn from the distribution  $\mathcal{N}(\mu, \sigma^2)$ , i.e. this is a sample of size  $n$  from a normally distributed population with expected mean value  $\mu$  and variance  $\sigma^2$ .

Let

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

be the sample mean and let

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

be the (Bessel-corrected) sample variance. Then the random variable

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

has a standard normal distribution (i.e. normal with expected mean 0 and variance 1), and the random variable

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

i.e. where  $S$  has been substituted for  $\sigma$ , has a Student's  $t$ -distribution with  $n - 1$  degrees of freedom. Since  $S$  has replaced  $\sigma$ , the only unobservable quantity in this expression is  $\mu$ , so this can be used to derive confidence intervals for  $\mu$ . The numerator and the denominator in the preceding expression are statistically independent random variables despite being based on the same sample  $X_1, \dots, X_n$ . This can be seen by observing that  $\text{cov}(\bar{X}, X_i - \bar{X}) = 0$ , and recalling that  $\bar{X}$  and  $X_i - \bar{X}$  are both linear combinations of the same set of i.i.d. normally distributed random variables.

## Definition

### Probability density function

Student's  **$t$ -distribution** has the probability density function (PDF) given by

$$\frac{K_{\nu/2}(\sqrt{\nu}|t|) \cdot (\sqrt{\nu}|t|)^{\nu/2}}{\Gamma(\nu/2) 2^{\nu/2-1}} \text{ for } \nu > 0$$

- $K_{\nu}(x)$ : modified Bessel function of the second kind<sup>[1]</sup>



Statistician William Sealy Gosset, known as "Student"

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2},$$

where  $\nu$  is the number of degrees of freedom and  $\Gamma$  is the gamma function. This may also be written as

$$f(t) = \frac{1}{\sqrt{\nu} B(\frac{1}{2}, \frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2},$$

where B is the Beta function. In particular for integer valued degrees of freedom  $\nu$  we have:

For  $\nu > 1$  even,

$$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} = \frac{(\nu-1)(\nu-3)\cdots 5\cdot 3}{2\sqrt{\nu}(\nu-2)(\nu-4)\cdots 4\cdot 2}.$$

For  $\nu > 1$  odd,

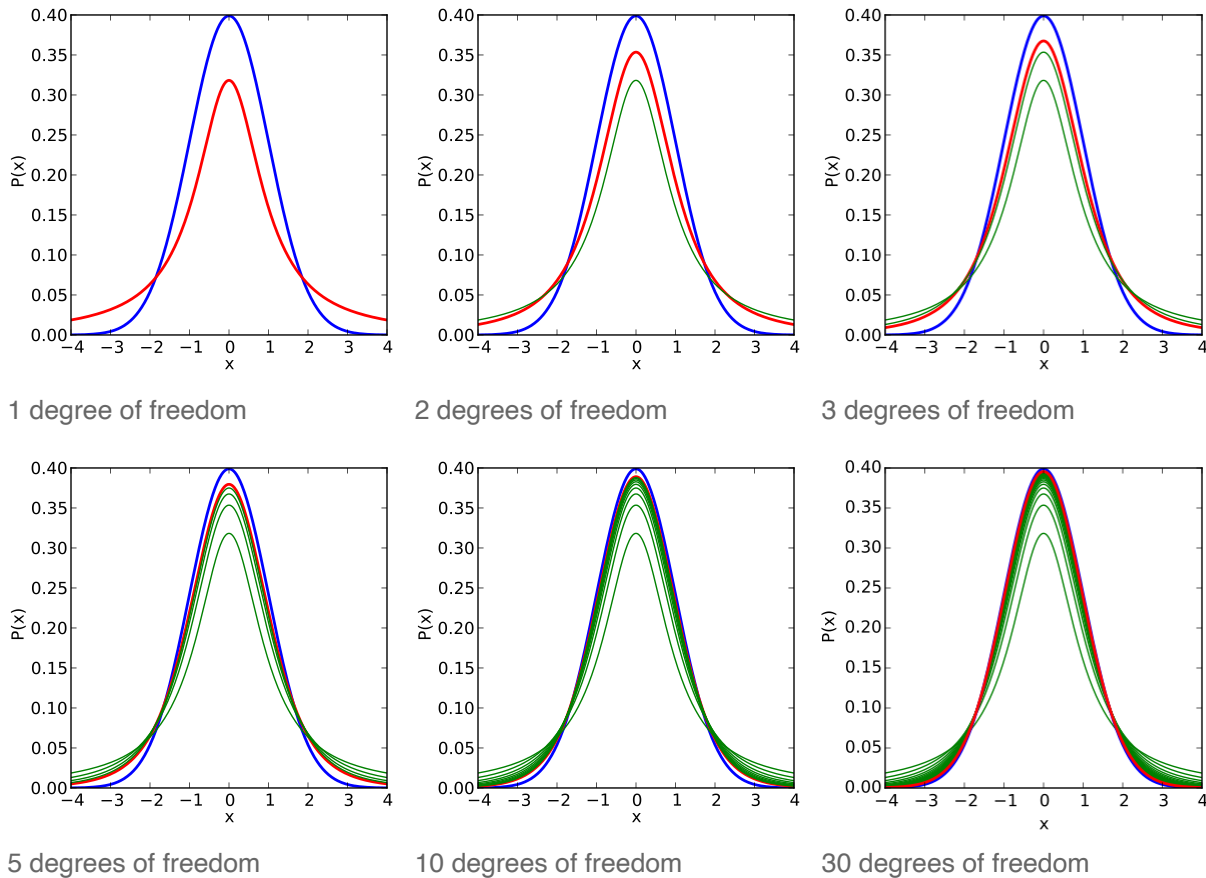
$$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} = \frac{(\nu-1)(\nu-3)\cdots 4\cdot 2}{\pi\sqrt{\nu}(\nu-2)(\nu-4)\cdots 5\cdot 3}.$$

The probability density function is symmetric, and its overall shape resembles the bell shape of a normally distributed variable with mean 0 and variance 1, except that it is a bit lower and wider. As the number of degrees of freedom grows, the  $t$ -distribution approaches the normal distribution with mean 0 and variance 1. For this reason  $\nu$  is also known as the normality parameter.<sup>[14]</sup>

The following images show the density of the  $t$ -distribution for increasing values of  $\nu$ . The normal distribution is shown as a blue line for comparison. Note that the  $t$ -distribution (red line) becomes closer to the normal distribution as  $\nu$  increases.

Density of the  $t$ -distribution (red) for 1, 2, 3, 5, 10, and 30 degrees of freedom compared to the standard normal distribution (blue).

Previous plots shown in green.



## Cumulative distribution function

The cumulative distribution function (CDF) can be written in terms of  $I$ , the regularized incomplete beta function. For  $t > 0$ ,<sup>[15]</sup>

$$F(t) = \int_{-\infty}^t f(u) du = 1 - \frac{1}{2} I_{x(t)} \left( \frac{\nu}{2}, \frac{1}{2} \right),$$

where

$$x(t) = \frac{\nu}{t^2 + \nu}.$$

Other values would be obtained by symmetry. An alternative formula, valid for  $t^2 < \nu$ , is<sup>[15]</sup>

$$\int_{-\infty}^t f(u) du = \frac{1}{2} + t \frac{\Gamma\left(\frac{1}{2}(\nu + 1)\right)}{\sqrt{\pi\nu} \Gamma\left(\frac{\nu}{2}\right)} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}(\nu + 1); \frac{3}{2}; -\frac{t^2}{\nu}\right),$$

where  ${}_2F_1$  is a particular case of the hypergeometric function.

Special cases

Certain values of  $\nu$  give a simple form for Student's t-distribution.

$\nu$	PDF	CDF	notes
1	$\frac{1}{\pi(1+t^2)}$	$\frac{1}{2} + \frac{1}{\pi} \arctan(t)$	See <a href="#">Cauchy distribution</a>
2	$\frac{1}{2\sqrt{2}\left(1+\frac{t^2}{2}\right)^{3/2}}$	$\frac{1}{2} + \frac{t}{2\sqrt{2}\sqrt{1+\frac{t^2}{2}}}$	
3	$\frac{2}{\pi\sqrt{3}\left(1+\frac{t^2}{3}\right)^2}$	$\frac{1}{2} + \frac{1}{\pi} \left[ \frac{1}{\sqrt{3}} \frac{t}{1+\frac{t^2}{3}} + \arctan\left(\frac{t}{\sqrt{3}}\right) \right]$	
4	$\frac{3}{8\left(1+\frac{t^2}{4}\right)^{5/2}}$	$\frac{1}{2} + \frac{3}{8} \frac{t}{\sqrt{1+\frac{t^2}{4}}} \left[ 1 - \frac{1}{12} \frac{t^2}{1+\frac{t^2}{4}} \right]$	
5	$\frac{8}{3\pi\sqrt{5}\left(1+\frac{t^2}{5}\right)^3}$	$\frac{1}{2} + \frac{1}{\pi} \left[ \frac{t}{\sqrt{5}\left(1+\frac{t^2}{5}\right)} \left( 1 + \frac{2}{3\left(1+\frac{t^2}{5}\right)} \right) + \arctan\left(\frac{t}{\sqrt{5}}\right) \right]$	
$\infty$	$\frac{1}{\sqrt{2\pi}} e^{-t^2/2}$	$\frac{1}{2} \left[ 1 + \operatorname{erf}\left(\frac{t}{\sqrt{2}}\right) \right]$	See <a href="#">Normal distribution</a> , <a href="#">Error function</a>

How the *t*-distribution arises

Sampling distribution

Let  $x_1, \dots, x_n$  be the numbers observed in a sample from a continuously distributed population with expected value  $\mu$ . The sample mean and [sample variance](#) are given by:

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n},$$
$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

The resulting *t*-value is

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}.$$

The  $t$ -distribution with  $n - 1$  degrees of freedom is the sampling distribution of the  $t$ -value when the samples consist of independent identically distributed observations from a normally distributed population. Thus for inference purposes  $t$  is a useful "pivotal quantity" in the case when the mean and variance  $(\mu, \sigma^2)$  are unknown population parameters, in the sense that the  $t$ -value has then a probability distribution that depends on neither  $\mu$  nor  $\sigma^2$ .

## Bayesian inference

In Bayesian statistics, a (scaled, shifted)  $t$ -distribution arises as the marginal distribution of the unknown mean of a normal distribution, when the dependence on an unknown variance has been marginalized out:<sup>[16]</sup>

$$\begin{aligned} p(\mu \mid D, I) &= \int p(\mu, \sigma^2 \mid D, I) d\sigma^2 \\ &= \int p(\mu \mid D, \sigma^2, I) p(\sigma^2 \mid D, I) d\sigma^2, \end{aligned}$$

where  $D$  stands for the data  $\{x_i\}$ , and  $I$  represents any other information that may have been used to create the model. The distribution is thus the compounding of the conditional distribution of  $\mu$  given the data and  $\sigma^2$  with the marginal distribution of  $\sigma^2$  given the data.

With  $n$  data points, if uninformative, or flat, the location prior  $p(\mu \mid \sigma^2, I) = \text{const}$  can be taken for  $\mu$ , and the scale prior  $p(\sigma^2 \mid I) \propto 1/\sigma^2$  can be taken for  $\sigma^2$ , then Bayes' theorem gives

$$\begin{aligned} p(\mu \mid D, \sigma^2, I) &\sim N(\bar{x}, \sigma^2/n), \\ p(\sigma^2 \mid D, I) &\sim \text{Scale-inv-}\chi^2(\nu, s^2), \end{aligned}$$

a normal distribution and a scaled inverse chi-squared distribution respectively, where  $\nu = n - 1$  and

$$s^2 = \sum \frac{(x_i - \bar{x})^2}{n - 1}.$$

The marginalization integral thus becomes

$$\begin{aligned} p(\mu \mid D, I) &\propto \int_0^\infty \frac{1}{\sqrt{\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} n(\mu - \bar{x})^2\right) \cdot \sigma^{-\nu-2} \exp(-\nu s^2/2\sigma^2) d\sigma^2 \\ &\propto \int_0^\infty \sigma^{-\nu-3} \exp\left(-\frac{1}{2\sigma^2} (n(\mu - \bar{x})^2 + \nu s^2)\right) d\sigma^2. \end{aligned}$$

This can be evaluated by substituting  $z = A/2\sigma^2$ , where  $A = n(\mu - \bar{x})^2 + \nu s^2$ , giving

$$dz = -\frac{A}{2\sigma^4} d\sigma^2,$$

so



$$p(\mu \mid D, I) \propto A^{-(\nu+1)/2} \int_0^\infty z^{(\nu-1)/2} \exp(-z) dz.$$

But the  $z$  integral is now a standard Gamma integral, which evaluates to a constant, leaving

$$\begin{aligned} p(\mu \mid D, I) &\propto A^{-(\nu+1)/2} \\ &\propto \left(1 + \frac{n(\mu - \bar{x})^2}{\nu s^2}\right)^{-(\nu+1)/2}. \end{aligned}$$

This is a form of the  $t$ -distribution with an explicit scaling and shifting that will be explored in more detail in a further section below. It can be related to the standardized  $t$ -distribution by the substitution

$$t = \frac{\mu - \bar{x}}{s/\sqrt{n}}.$$

The derivation above has been presented for the case of uninformative priors for  $\mu$  and  $\sigma^2$ ; but it will be apparent that any priors that lead to a normal distribution being compounded with a scaled inverse chi-squared distribution will lead to a  $t$ -distribution with scaling and shifting for  $P(\mu \mid D, I)$ , although the scaling parameter corresponding to  $\frac{s^2}{n}$  above will then be influenced both by the prior information and the data, rather than just by the data as above.

## Characterization

---

### As the distribution of a test statistic

Student's  $t$ -distribution with  $\nu$  degrees of freedom can be defined as the distribution of the random variable  $T$  with<sup>[15][17]</sup>

$$T = \frac{Z}{\sqrt{V/\nu}} = Z \sqrt{\frac{\nu}{V}},$$

where

- $Z$  is a standard normal with expected value 0 and variance 1;
- $V$  has a chi-squared distribution ( **$\chi^2$ -distribution**) with  $\nu$  degrees of freedom;
- $Z$  and  $V$  are independent;

A different distribution is defined as that of the random variable defined, for a given constant  $\mu$ , by

$$(Z + \mu) \sqrt{\frac{\nu}{V}}.$$

This random variable has a noncentral  $t$ -distribution with noncentrality parameter  $\mu$ . This distribution is important in studies of the power of Student's  $t$ -test.

## Derivation

Suppose  $X_1, \dots, X_n$  are independent realizations of the normally-distributed, random variable  $X$ , which has an expected value  $\mu$  and variance  $\sigma^2$ . Let

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$

be the sample mean, and

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

be an unbiased estimate of the variance from the sample. It can be shown that the random variable

$$V = (n-1) \frac{S_n^2}{\sigma^2}$$

has a chi-squared distribution with  $\nu = n - 1$  degrees of freedom (by Cochran's theorem).<sup>[18]</sup> It is readily shown that the quantity

$$Z = (\bar{X}_n - \mu) \frac{\sqrt{n}}{\sigma}$$

is normally distributed with mean 0 and variance 1, since the sample mean  $\bar{X}_n$  is normally distributed with mean  $\mu$  and variance  $\sigma^2/n$ . Moreover, it is possible to show that these two random variables (the normally distributed one  $Z$  and the chi-squared-distributed one  $V$ ) are independent. Consequently the pivotal quantity

$$T \equiv \frac{Z}{\sqrt{V/\nu}} = (\bar{X}_n - \mu) \frac{\sqrt{n}}{S_n},$$

which differs from  $Z$  in that the exact standard deviation  $\sigma$  is replaced by the random variable  $S_n$ , has a Student's  $t$ -distribution as defined above. Notice that the unknown population variance  $\sigma^2$  does not appear in  $T$ , since it was in both the numerator and the denominator, so it canceled. Gosset intuitively obtained the probability density function stated above, with  $\nu$  equal to  $n - 1$ , and Fisher proved it in 1925.<sup>[12]</sup>

The distribution of the test statistic  $T$  depends on  $\nu$ , but not  $\mu$  or  $\sigma$ ; the lack of dependence on  $\mu$  and  $\sigma$  is what makes the  $t$ -distribution important in both theory and practice.

## As a maximum entropy distribution

Student's  $t$ -distribution is the maximum entropy probability distribution for a random variate  $X$  for which  $\mathbf{E}(\ln(\nu + X^2))$  is fixed.<sup>[19]</sup>

## Properties

---

### Moments

For  $\nu > 1$ , the raw moments of the  $t$ -distribution are

$$\mathbf{E}(T^k) = \begin{cases} 0 & k \text{ odd, } 0 < k < \nu \\ \frac{1}{\sqrt{\pi}\Gamma(\frac{\nu}{2})} \left[ \Gamma\left(\frac{k+1}{2}\right) \Gamma\left(\frac{\nu-k}{2}\right) \nu^{\frac{k}{2}} \right] & k \text{ even, } 0 < k < \nu. \end{cases}$$

Moments of order  $\nu$  or higher do not exist.<sup>[20]</sup>

The term for  $0 < k < \nu$ ,  $k$  even, may be simplified using the properties of the gamma function to

$$\mathbf{E}(T^k) = \nu^{\frac{k}{2}} \prod_{i=1}^{k/2} \frac{2i-1}{\nu-2i} \quad k \text{ even, } 0 < k < \nu.$$

For a  $t$ -distribution with  $\nu$  degrees of freedom, the expected value is 0 if  $\nu > 1$ , and its variance is  $\frac{\nu}{\nu-2}$  if  $\nu > 2$ . The skewness is 0 if  $\nu > 3$  and the excess kurtosis is  $\frac{6}{\nu-4}$  if  $\nu > 4$ .

### Monte Carlo sampling

There are various approaches to constructing random samples from the Student's  $t$ -distribution. The matter depends on whether the samples are required on a stand-alone basis, or are to be constructed by application of a quantile function to uniform samples; e.g., in the multi-dimensional applications basis of copula-dependency. In the case of stand-alone sampling, an extension of the Box–Muller method and its polar form is easily deployed.<sup>[21]</sup> It has the merit that it applies equally well to all real positive degrees of freedom,  $\nu$ , while many other candidate methods fail if  $\nu$  is close to zero.<sup>[21]</sup>

### Integral of Student's probability density function and $p$ -value

The function  $A(t \mid \nu)$  is the integral of Student's probability density function,  $f(t)$  between  $-t$  and  $t$ , for  $t \geq 0$ . It thus gives the probability that a value of  $t$  less than that calculated from observed data would occur by chance. Therefore, the function  $A(t \mid \nu)$  can be used when testing whether the difference between the means of two sets of data is statistically significant, by calculating the corresponding value of  $t$  and the probability of its occurrence if the two sets of data were drawn from the same population. This is used in a variety of situations, particularly in  $t$ -tests. For the statistic  $t$ , with  $\nu$  degrees of freedom,  $A(t \mid \nu)$  is the probability that  $t$  would be less than the observed value if the two means were the same (provided that the smaller mean is subtracted from the larger, so that  $t \geq 0$ ). It can be easily calculated from the cumulative distribution function  $F_\nu(t)$  of the  $t$ -distribution:

$$A(t \mid \nu) = F_\nu(t) - F_\nu(-t) = 1 - I_{\frac{\nu}{\nu+t^2}} \left( \frac{\nu}{2}, \frac{1}{2} \right),$$

where  $I_x$  is the regularized incomplete beta function ( $a, b$ ).

For statistical hypothesis testing this function is used to construct the p-value.

## Generalized Student's *t*-distribution

---

### In terms of scaling parameter $\hat{\sigma}$ or $\hat{\sigma}^2$

Student's *t* distribution can be generalized to a three parameter location-scale family, introducing a location parameter  $\hat{\mu}$  and a scale parameter  $\hat{\sigma}$ , through the relation

$$X = \hat{\mu} + \hat{\sigma}T$$

or

$$T = \frac{X - \hat{\mu}}{\hat{\sigma}}$$

This means that  $\frac{x - \hat{\mu}}{\hat{\sigma}}$  has a classic Student's *t* distribution with  $\nu$  degrees of freedom.

The resulting **non-standardized Student's *t*-distribution** has a density defined by:<sup>[22]</sup>

$$p(x \mid \nu, \hat{\mu}, \hat{\sigma}) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\hat{\sigma}} \left( 1 + \frac{1}{\nu} \left( \frac{x - \hat{\mu}}{\hat{\sigma}} \right)^2 \right)^{-(\nu+1)/2}$$

Here,  $\hat{\sigma}$  does *not* correspond to a standard deviation: it is not the standard deviation of the scaled *t* distribution, which may not even exist; nor is it the standard deviation of the underlying normal distribution, which is unknown.  $\hat{\sigma}$  simply sets the overall scaling of the distribution. In the Bayesian derivation of the marginal distribution of an unknown normal mean  $\hat{\mu}$  above,  $\hat{\sigma}$  as used here corresponds to the quantity  $s/\sqrt{n}$ , where

$$s^2 = \sum \frac{(x_i - \bar{x})^2}{n-1}.$$

Equivalently, the distribution can be written in terms of  $\hat{\sigma}^2$ , the square of this scale parameter:

$$p(x \mid \nu, \hat{\mu}, \hat{\sigma}^2) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\hat{\sigma}^2} \left( 1 + \frac{1}{\nu} \frac{(x - \hat{\mu})^2}{\hat{\sigma}^2} \right)^{-(\nu+1)/2}$$

Other properties of this version of the distribution are:<sup>[22]</sup>

$$\begin{aligned} \mathbf{E}(X) &= \hat{\mu} & \text{for } \nu > 1 \\ \mathbf{var}(X) &= \hat{\sigma}^2 \frac{\nu}{\nu - 2} & \text{for } \nu > 2 \\ \mathbf{mode}(X) &= \hat{\mu} \end{aligned}$$

This distribution results from compounding a Gaussian distribution (normal distribution) with mean  $\mu$  and unknown variance, with an inverse gamma distribution placed over the variance with parameters  $a = \nu/2$  and  $b = \nu\hat{\sigma}^2/2$ . In other words, the random variable  $X$  is assumed to have a Gaussian distribution with an unknown variance distributed as inverse gamma, and then the variance is marginalized out (integrated out). The reason for the usefulness of this characterization is that the inverse gamma distribution is the conjugate prior distribution of the variance of a Gaussian distribution. As a result, the non-standardized Student's  $t$ -distribution arises naturally in many Bayesian inference problems. See below.

Equivalently, this distribution results from compounding a Gaussian distribution with a scaled-inverse-chi-squared distribution with parameters  $\nu$  and  $\hat{\sigma}^2$ . The scaled-inverse-chi-squared distribution is exactly the same distribution as the inverse gamma distribution, but with a different parameterization, i.e.  $\nu = 2a$ ,  $\hat{\sigma}^2 = \frac{b}{a}$ .

This version of the t-distribution can be useful in financial modeling. For example, Platen and Sidorowicz found that among the family of generalized hyperbolic distributions, this form of the t-distribution with about 4 degrees of freedom was the best fit for the (log) return of many worldwide stock indices.<sup>[23]</sup>

## In terms of inverse scaling parameter $\lambda$

An alternative parameterization in terms of an inverse scaling parameter  $\lambda$  (analogous to the way precision is the reciprocal of variance), defined by the relation  $\lambda = \frac{1}{\hat{\sigma}^2}$ . The density is then given by:<sup>[24]</sup>

$$p(x \mid \nu, \hat{\mu}, \lambda) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left( \frac{\lambda}{\pi\nu} \right)^{1/2} \left( 1 + \frac{\lambda(x - \hat{\mu})^2}{\nu} \right)^{-(\nu+1)/2}.$$

Other properties of this version of the distribution are:<sup>[24]</sup>

$$\begin{aligned} \mathbf{E}(X) &= \hat{\mu} & \text{for } \nu > 1 \\ \mathbf{var}(X) &= \frac{1}{\lambda} \frac{\nu}{\nu - 2} & \text{for } \nu > 2 \\ \mathbf{mode}(X) &= \hat{\mu} \end{aligned}$$

This distribution results from compounding a Gaussian distribution with mean  $\hat{\mu}$  and unknown precision (the reciprocal of the variance), with a gamma distribution placed over the precision with parameters  $a = \nu/2$  and  $b = \nu/(2\lambda)$ . In other words, the random variable  $X$  is assumed to have a

normal distribution with an unknown precision distributed as gamma, and then this is marginalized over the gamma distribution.

## Related distributions

- If  $X$  has a Student's  $t$ -distribution with degree of freedom  $\nu$  then  $X^2$  has an  $F$ -distribution:  
 $X^2 \sim F(\nu_1 = 1, \nu_2 = \nu)$
- The noncentral  $t$ -distribution generalizes the  $t$ -distribution to include a location parameter. Unlike the nonstandardized  $t$ -distributions, the noncentral distributions are not symmetric (the median is not the same as the mode).
- The **discrete Student's  $t$ -distribution** is defined by its probability mass function at  $r$  being proportional to:<sup>[25]</sup>

$$\prod_{j=1}^k \frac{1}{(r+j+a)^2 + b^2} \quad r = \dots, -1, 0, 1, \dots$$

Here  $a$ ,  $b$ , and  $k$  are parameters. This distribution arises from the construction of a system of discrete distributions similar to that of the Pearson distributions for continuous distributions.<sup>[26]</sup>

- One can generate Student- $t$  samples by taking the ratio of variables from the normal distribution and the square-root of  $\chi^2$ -**distribution**. If we use instead of the normal distribution, e.g., the Irwin–Hall distribution, we obtain over-all a symmetric 4-parameter distribution, which includes the normal, the uniform, the triangular, the Student- $t$  and the Cauchy distribution. This is also more flexible than some other symmetric generalizations of the normal distribution.
- $t$ -distribution is an instance of ratio distributions.

## Bayesian inference: prior distribution for the degrees of the freedom

Suppose that  $\mathbf{x} = (x_1, \dots, x_N)$  represents  $N$  number of independently and identically distributed samples drawn from the Student  $t$ -distribution

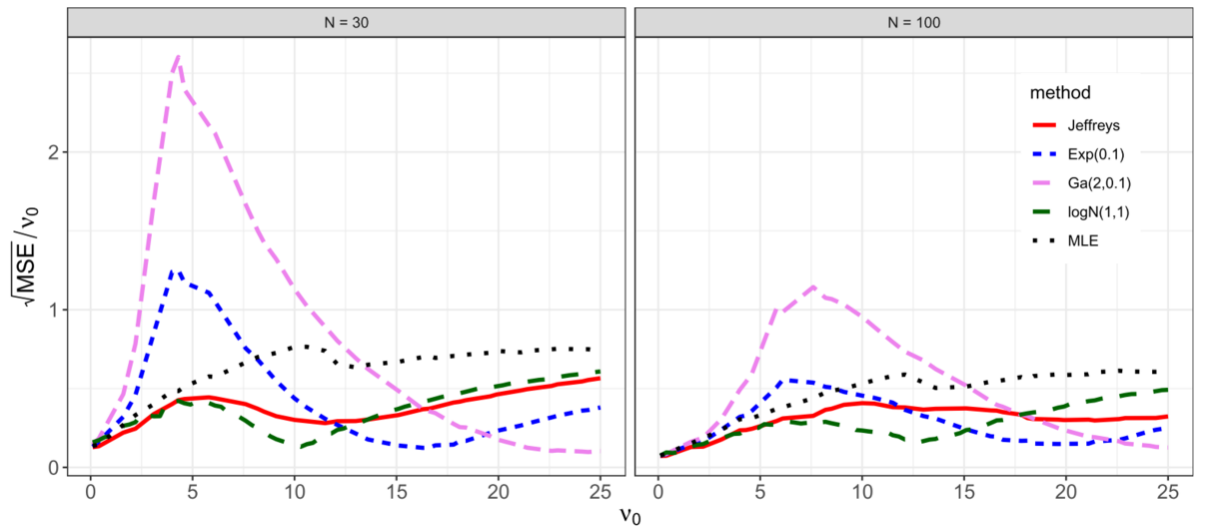
$$t_\nu(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad \mathbf{x} \in \mathbb{R}.$$

With a choice a prior for the degrees of freedom  $\nu$ , denoted as  $\pi(\nu)$ , Bayesian inference seeks to evaluate the posterior distribution

$$\pi(\nu|\mathbf{x}) = \frac{\prod t_\nu(x_i) \cdot \pi(\nu)}{\int \prod t_\nu(x_i) \cdot \pi(\nu) d\nu}, \quad \nu \in \mathbb{R}^+.$$

Some popular choices of the priors are:

- Jeffreys prior <sup>[28]</sup>



Mean squared error comparison between Bayes estimators based on the four priors and maximum likelihood estimator for the degrees of the freedom. Data is simulated from the student t distribution with the degrees of freedom  $\nu_0$  varying from 0 to 25 with the sample size  $N = 30$  (left) and  $N = 100$  (right). Lower value for MSE implies better accuracy.<sup>[27]</sup>

$$\pi_J(\nu) \propto \left( \frac{\nu}{\nu+3} \right)^{1/2} \left( \psi' \left( \frac{\nu}{2} \right) - \psi' \left( \frac{\nu+1}{2} \right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right)^{1/2}, \quad \nu \in \mathbb{R}^+, \quad \text{where} \quad \psi'(x)$$

represents trigamma function.

- Exponential prior <sup>[29]</sup>

$$\pi_E(\nu) = Ga(\nu|1, 0.1) = \exp(\nu|0.1) = \frac{1}{10} e^{-\nu/10}, \quad \nu \in \mathbb{R}^+$$

- Gamma prior <sup>[30]</sup>

$$\pi_G(\nu) = Ga(\nu|2, 0.1) = \frac{\nu}{100} e^{-\nu/10}, \quad \nu \in \mathbb{R}^+$$

- Log-normal prior <sup>[31]</sup>

$$\pi_L(\nu) = \log N(\nu|1, 1) = \frac{1}{\nu\sqrt{2\pi}} \exp \left[ -\frac{(\log \nu - 1)^2}{2} \right], \quad \nu \in \mathbb{R}^+$$

The right panels show the result of the numerical experiments. The Bayes estimator based on the Jeffreys prior  $\pi_J(\nu)$  results in relatively lower Mean Squared Error (MSE) than the Maximum Likelihood Estimator (MLE) over the values  $\nu_0 \in (0, 25)$ . It is important to note that no Bayes estimator dominates other estimators over the interval  $(0, 25)$ . In other words, each Bayes estimator has its own region where the estimator is non-inferior to others.

## Uses

---

### In frequentist statistical inference

Student's  $t$ -distribution arises in a variety of statistical estimation problems where the goal is to estimate an unknown parameter, such as a mean value, in a setting where the data are observed with additive errors. If (as in nearly all practical statistical work) the population standard deviation of these errors is unknown and has to be estimated from the data, the  $t$ -distribution is often used to account for the extra uncertainty that results from this estimation. In most such problems, if the standard deviation of the errors were known, a normal distribution would be used instead of the  $t$ -distribution.

Confidence intervals and hypothesis tests are two statistical procedures in which the quantiles of the sampling distribution of a particular statistic (e.g. the standard score) are required. In any situation where this statistic is a linear function of the data, divided by the usual estimate of the standard deviation, the resulting quantity can be rescaled and centered to follow Student's  $t$ -distribution. Statistical analyses involving means, weighted means, and regression coefficients all lead to statistics having this form.

Quite often, textbook problems will treat the population standard deviation as if it were known and thereby avoid the need to use the Student's  $t$ -distribution. These problems are generally of two kinds: (1) those in which the sample size is so large that one may treat a data-based estimate of the variance as if it were certain, and (2) those that illustrate mathematical reasoning, in which the problem of estimating the standard deviation is temporarily ignored because that is not the point that the author or instructor is then explaining.

### Hypothesis testing

A number of statistics can be shown to have  $t$ -distributions for samples of moderate size under null hypotheses that are of interest, so that the  $t$ -distribution forms the basis for significance tests. For example, the distribution of Spearman's rank correlation coefficient  $\rho$ , in the null case (zero correlation) is well approximated by the  $t$  distribution for sample sizes above about 20.

### Confidence intervals

Suppose the number  $A$  is so chosen that

$$\Pr(-A < T < A) = 0.9,$$

when  $T$  has a  $t$ -distribution with  $n - 1$  degrees of freedom. By symmetry, this is the same as saying that  $A$  satisfies

$$\Pr(T < A) = 0.95,$$



so  $A$  is the "95th percentile" of this probability distribution, or  $A = t_{(0.05, n-1)}$ . Then

$$\Pr \left( -A < \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}} < A \right) = 0.9,$$

and this is equivalent to

$$\Pr \left( \bar{X}_n - A \frac{S_n}{\sqrt{n}} < \mu < \bar{X}_n + A \frac{S_n}{\sqrt{n}} \right) = 0.9.$$

Therefore, the interval whose endpoints are

$$\bar{X}_n \pm A \frac{S_n}{\sqrt{n}}$$

is a 90% confidence interval for  $\mu$ . Therefore, if we find the mean of a set of observations that we can reasonably expect to have a normal distribution, we can use the  $t$ -distribution to examine whether the confidence limits on that mean include some theoretically predicted value – such as the value predicted on a null hypothesis.

It is this result that is used in the Student's  $t$ -tests: since the difference between the means of samples from two normal distributions is itself distributed normally, the  $t$ -distribution can be used to examine whether that difference can reasonably be supposed to be zero.

If the data are normally distributed, the one-sided  $(1 - \alpha)$ -upper confidence limit (UCL) of the mean, can be calculated using the following equation:

$$\text{UCL}_{1-\alpha} = \bar{X}_n + t_{\alpha, n-1} \frac{S_n}{\sqrt{n}}.$$

The resulting UCL will be the greatest average value that will occur for a given confidence interval and population size. In other words,  $\bar{X}_n$  being the mean of the set of observations, the probability that the mean of the distribution is inferior to  $\text{UCL}_{1-\alpha}$  is equal to the confidence level  $1 - \alpha$ .

## Prediction intervals

The  $t$ -distribution can be used to construct a prediction interval for an unobserved sample from a normal distribution with unknown mean and variance.

## In Bayesian statistics

The Student's  $t$ -distribution, especially in its three-parameter (location-scale) version, arises frequently in Bayesian statistics as a result of its connection with the normal distribution. Whenever the variance of a normally distributed random variable is unknown and a conjugate prior placed over it that follows an inverse gamma distribution, the resulting marginal distribution of the variable will follow a Student's  $t$ -distribution. Equivalent constructions with the same results involve a conjugate scaled-inverse-chi-squared distribution over the variance, or a conjugate gamma distribution over the precision. If an improper prior proportional to  $\sigma^{-2}$  is placed over the variance, the  $t$ -distribution also

arises. This is the case regardless of whether the mean of the normally distributed variable is known, is unknown distributed according to a conjugate normally distributed prior, or is unknown distributed according to an improper constant prior.

Related situations that also produce a  $t$ -distribution are:

- The marginal posterior distribution of the unknown mean of a normally distributed variable, with unknown prior mean and variance following the above model.
- The prior predictive distribution and posterior predictive distribution of a new normally distributed data point when a series of independent identically distributed normally distributed data points have been observed, with prior mean and variance as in the above model.

## Robust parametric modeling

The  $t$ -distribution is often used as an alternative to the normal distribution as a model for data, which often has heavier tails than the normal distribution allows for; see e.g. Lange et al.<sup>[32]</sup> The classical approach was to identify outliers (e.g., using Grubbs's test) and exclude or downweight them in some way. However, it is not always easy to identify outliers (especially in high dimensions), and the  $t$ -distribution is a natural choice of model for such data and provides a parametric approach to robust statistics.

A Bayesian account can be found in Gelman et al.<sup>[33]</sup> The degrees of freedom parameter controls the kurtosis of the distribution and is correlated with the scale parameter. The likelihood can have multiple local maxima and, as such, it is often necessary to fix the degrees of freedom at a fairly low value and estimate the other parameters taking this as given. Some authors report that values between 3 and 9 are often good choices. Venables and Ripley suggest that a value of 5 is often a good choice.

## Student's $t$ -process

For practical regression and prediction needs, Student's  $t$ -processes were introduced, that are generalisations of the Student  $t$ -distributions for functions. A Student's  $t$ -process is constructed from the Student  $t$ -distributions like a Gaussian process is constructed from the Gaussian distributions. For a Gaussian process, all sets of values have a multidimensional Gaussian distribution. Analogously,  $X(t)$  is a Student  $t$ -process on an interval  $I = [a, b]$  if the correspondent values of the process  $X(t_1), \dots, X(t_n)$  ( $t_i \in I$ ) have a joint multivariate Student  $t$ -distribution.<sup>[34]</sup> These processes are used for regression, prediction, Bayesian optimization and related problems. For multivariate regression and multi-output prediction, the multivariate Student  $t$ -processes are introduced and used.<sup>[35]</sup>

## Table of selected values

The following table lists values for  $t$ -distributions with  $\nu$  degrees of freedom for a range of one-sided or two-sided critical regions. The first column is  $\nu$ , the percentages along the top are confidence levels, and the numbers in the body of the table are the  $t_{\alpha, n-1}$  factors described in the section on confidence intervals.

The last row with infinite  $\nu$  gives critical points for a normal distribution since a  $t$ -distribution with infinitely many degrees of freedom is a normal distribution. (See Related distributions above).

<i>One-sided</i>	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
<i>Two-sided</i>	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416

<b>100</b>	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
<b>120</b>	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
<b>∞</b>	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
<b>One-sided</b>	<b>75%</b>	<b>80%</b>	<b>85%</b>	<b>90%</b>	<b>95%</b>	<b>97.5%</b>	<b>99%</b>	<b>99.5%</b>	<b>99.75%</b>	<b>99.9%</b>	<b>99.95%</b>
<b>Two-sided</b>	<b>50%</b>	<b>60%</b>	<b>70%</b>	<b>80%</b>	<b>90%</b>	<b>95%</b>	<b>98%</b>	<b>99%</b>	<b>99.5%</b>	<b>99.8%</b>	<b>99.9%</b>

## Calculating the confidence interval

Let's say we have a sample with size 11, sample mean 10, and sample variance 2. For 90% confidence with 10 degrees of freedom, the one-sided  $t$ -value from the table is 1.372. Then with confidence interval calculated from

$$\bar{X}_n \pm t_{\alpha, \nu} \frac{S_n}{\sqrt{n}},$$

we determine that with 90% confidence we have a true mean lying below

$$10 + 1.372 \frac{\sqrt{2}}{\sqrt{11}} = 10.585.$$

In other words, 90% of the times that an upper threshold is calculated by this method from particular samples, this upper threshold exceeds the true mean.

And with 90% confidence we have a true mean lying above

$$10 - 1.372 \frac{\sqrt{2}}{\sqrt{11}} = 9.414.$$

In other words, 90% of the times that a lower threshold is calculated by this method from particular samples, this lower threshold lies below the true mean.

So that at 80% confidence (calculated from  $100\% - 2 \times (1 - 90\%) = 80\%$ ), we have a true mean lying within the interval

$$\left( 10 - 1.372 \frac{\sqrt{2}}{\sqrt{11}}, 10 + 1.372 \frac{\sqrt{2}}{\sqrt{11}} \right) = (9.414, 10.585).$$

Saying that 80% of the times that upper and lower thresholds are calculated by this method from a given sample, the true mean is both below the upper threshold and above the lower threshold is not the same as saying that there is an 80% probability that the true mean lies between a particular pair of upper and lower thresholds that have been calculated by this method; see confidence interval and prosecutor's fallacy.

Nowadays, statistical software, such as the R programming language, and functions available in many spreadsheet programs compute values of the *t*-distribution and its inverse without tables.

## See also



- F-distribution
- Folded-*t* and half-*t* distributions
- Hotelling's *T*-squared distribution
- Multivariate Student distribution
- Standard normal table (Z-distribution table)
- *t*-statistic
- Tau distribution, for internally studentized residuals
- Wilks' lambda distribution
- Wishart distribution
- Modified half-normal distribution<sup>[36]</sup> with the pdf on  $(0, \infty)$  is given as

$$f(x) = \frac{2\beta^{\frac{\alpha}{2}} x^{\alpha-1} \exp(-\beta x^2 + \gamma x)}{\Psi\left(\frac{\alpha}{2}, \frac{\gamma}{\sqrt{\beta}}\right)}, \text{ where } \Psi(\alpha, z) = {}_1\Psi_1\left(\begin{matrix} (\alpha, \frac{1}{2}) \\ (1, 0) \end{matrix}; z\right) \text{ denotes the Fox-}$$

Wright Psi function.

## Notes

1. Hurst, Simon. "The Characteristic Function of the Student *t* Distribution" (<https://web.archive.org/web/20100218072259/http://wwwmaths.anu.edu.au/research.reports/srr/95/044/>). *Financial Mathematics Research Report No. FMRR006-95, Statistics Research Report No. SRR044-95*. Archived from the original (<http://wwwmaths.anu.edu.au/research.reports/srr/95/044/>) on February 18, 2010.
2. Helmert FR (1875). "Über die Berechnung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler". *Z. Math. U. Physik*. **20**: 300–3.
3. Helmert FR (1876). "Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und über einige damit in Zusammenhang stehende Fragen". *Z. Math. Phys.* **21**: 192–218.
4. Helmert FR (1876). "Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers directer Beobachtungen gleicher Genauigkeit" (<https://zenodo.org/record/1424695>) [The accuracy of Peters' formula for calculating the probable observation error of direct observations of the same accuracy]. *Astron. Nachr.* (in German). **88** (8–9): 113–132. Bibcode:1876AN.....88..113H (<https://ui.adsabs.harvard.edu/abs/1876AN.....88..113H>). doi:10.1002/asna.18760880802 (<https://doi.org/10.1002%2Fasna.18760880802>).
5. Lüroth J (1876). "Vergleichung von zwei Werten des wahrscheinlichen Fehlers" (<https://zenodo.org/record/1424693>). *Astron. Nachr.* **87** (14): 209–20. Bibcode:1876AN.....87..209L (<https://ui.adsabs.harvard.edu/abs/1876AN.....87..209L>). doi:10.1002/asna.18760871402 (<https://doi.org/10.1002%2Fasna.18760871402>).
6. Pfanzagl J, Sheynin O (1996). "Studies in the history of probability and statistics. XLIV. A forerunner of the *t*-distribution". *Biometrika*. **83** (4): 891–898. doi:10.1093/biomet/83.4.891 (<https://doi.org/10.1093%2Fbiomet%2F83.4.891>). MR 1766040 (<https://mathscinet.ams.org/mathscinet-getitem?mr=1766040>).

7. Sheynin O (1995). "Helmert's work in the theory of errors". *Arch. Hist. Exact Sci.* **49** (1): 73–104. doi:[10.1007/BF00374700](https://doi.org/10.1007/BF00374700) (<https://doi.org/10.1007%2FBF00374700>). S2CID [121241599](https://api.semanticscholar.org/CorpusID:121241599) (<https://api.semanticscholar.org/CorpusID:121241599>).
8. Pearson, K. (1895-01-01). "Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material" (<https://doi.org/10.1098%2Frsa.1895.0010>). *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. **186**: 343–414 (374). Bibcode:1895RSPTA.186..343P (<https://ui.adsabs.harvard.edu/abs/1895RSPTA.186..343P>). doi:10.1098/rsta.1895.0010 (<https://doi.org/10.1098%2Frsa.1895.0010>). ISSN [1364-503X](https://www.worldcat.org/issn/1364-503X) (<https://www.worldcat.org/issn/1364-503X>).
9. "Student" [William Sealy Gosset] (1908). "The probable error of a mean" (<http://www.york.ac.uk/depts/maths/histstat/student.pdf>) (PDF). *Biometrika*. **6** (1): 1–25. doi:10.1093/biomet/6.1.1 (<https://doi.org/10.1093%2Fbiomet%2F6.1.1>). hdl:10338.dmlcz/143545 (<https://hdl.handle.net/10338.dmlcz%2F143545>). JSTOR [2331554](https://www.jstor.org/stable/2331554) (<https://www.jstor.org/stable/2331554>).
10. Wendl MC (2016). "Pseudonymous fame". *Science*. **351** (6280): 1406. Bibcode:2016Sci...351.1406W (<https://ui.adsabs.harvard.edu/abs/2016Sci...351.1406W>). doi:10.1126/science.351.6280.1406 (<https://doi.org/10.1126%2Fscience.351.6280.1406>). PMID [27013722](https://pubmed.ncbi.nlm.nih.gov/27013722) (<https://pubmed.ncbi.nlm.nih.gov/27013722>).
11. Mortimer RG (2005). *Mathematics for physical chemistry* ([https://archive.org/details/mathematicsforph00mort\\_321](https://archive.org/details/mathematicsforph00mort_321)) (3rd ed.). Burlington, MA: Elsevier. pp. 326 ([https://archive.org/details/mathematicsforph00mort\\_321/page/n339](https://archive.org/details/mathematicsforph00mort_321/page/n339)). ISBN [9780080492889](https://www.worldcat.org/oclc/156200058). OCLC [156200058](https://www.worldcat.org/oclc/156200058) (<https://www.worldcat.org/oclc/156200058>).
12. Fisher RA (1925). "Applications of 'Student's' distribution" ([https://web.archive.org/web/20160305130235/http://www.bothis.ro/user/content/4ef6e90670749a86-student\\_distribution\\_1925.pdf](https://web.archive.org/web/20160305130235/http://www.bothis.ro/user/content/4ef6e90670749a86-student_distribution_1925.pdf)) (PDF). *Metron*. **5**: 90–104. Archived from the original ([http://www.bothis.ro/user/content/4ef6e90670749a86-student\\_distribution\\_1925.pdf](http://www.bothis.ro/user/content/4ef6e90670749a86-student_distribution_1925.pdf)) (PDF) on 5 March 2016.
13. Walpole RE, Myers R, Myers S, et al. (2006). *Probability & Statistics for Engineers & Scientists* (7th ed.). New Delhi: Pearson. p. 237. ISBN [9788177584042](https://www.worldcat.org/oclc/818811849). OCLC [818811849](https://www.worldcat.org/oclc/818811849) (<https://www.worldcat.org/oclc/818811849>).
14. Kruschke JK (2015). *Doing Bayesian Data Analysis* (2nd ed.). Academic Press. ISBN [9780124058880](https://www.worldcat.org/oclc/959632184). OCLC [959632184](https://www.worldcat.org/oclc/959632184) (<https://www.worldcat.org/oclc/959632184>).
15. Johnson NL, Kotz S, Balakrishnan N (1995). "Chapter 28". *Continuous Univariate Distributions*. Vol. 2 (2nd ed.). Wiley. ISBN [9780471584940](https://www.worldcat.org/oclc/959632184).
16. Gelman AB, Carlin JS, Rubin DB, et al. (1997). *Bayesian Data Analysis* (2nd ed.). Boca Raton: Chapman & Hall. p. 68. ISBN [9780412039911](https://www.worldcat.org/oclc/959632184).
17. Hogg RV, Craig AT (1978). *Introduction to Mathematical Statistics* (4th ed.). New York: Macmillan. ASIN [B010WFO0SA](https://www.amazon.com/dp/B010WFO0SA) (<https://www.amazon.com/dp/B010WFO0SA>). Sections 4.4 and 4.8
18. Cochran WG (1934). "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance". *Math. Proc. Camb. Philos. Soc.* **30** (2): 178–191. Bibcode:1934PCPS...30..178C (<https://ui.adsabs.harvard.edu/abs/1934PCPS...30..178C>). doi:10.1017/S0305004100016595 (<https://doi.org/10.1017%2FS0305004100016595>). S2CID [122547084](https://api.semanticscholar.org/CorpusID:122547084) (<https://api.semanticscholar.org/CorpusID:122547084>).
19. Park SY, Bera AK (2009). "Maximum entropy autoregressive conditional heteroskedasticity model". *J. Econom.* **150** (2): 219–230. doi:10.1016/j.jeconom.2008.12.014 (<https://doi.org/10.1016%2Fj.jeconom.2008.12.014>).
20. Casella G, Berger RL (1990). *Statistical Inference*. Duxbury Resource Center. p. 56. ISBN [9780534119584](https://www.worldcat.org/oclc/959632184).