

Student's *t*-distribution

From Wikipedia, the free encyclopedia

In probability and statistics, **Student's *t*-distribution** (or simply the ***t*-distribution**) is any member of a family of continuous probability distributions that arises when estimating the mean of a normally distributed population in situations where the sample size is small and population standard deviation is unknown. It was developed by William Sealy Gosset under the pseudonym *Student*. Whereas a normal distribution describes a full population, *t*-distributions describe samples drawn from a full population; accordingly, the *t*-distribution for each sample size is different, and the larger the sample, the more the distribution resembles a normal distribution.

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. The Student's *t*-distribution also arises in the Bayesian analysis of data from a normal family.

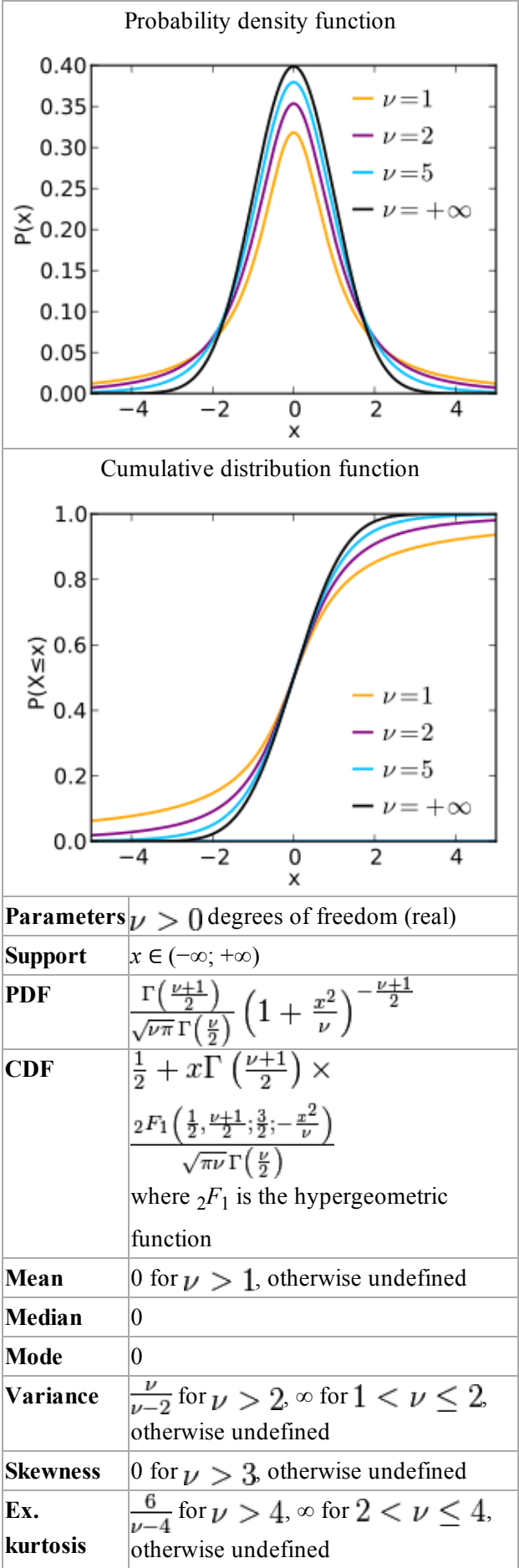
If we take a sample of *n* observations from a normal distribution, then the *t*-distribution with $\nu = n - 1$ degrees of freedom can be defined as the distribution of the location of the true mean, relative to the sample mean and divided by the sample standard deviation, after multiplying by the normalizing term \sqrt{n} . In this way, the *t*-distribution can be used to estimate how likely it is that the true mean lies in any given range.

The *t*-distribution is symmetric and bell-shaped, like the normal distribution, but 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 generalised hyperbolic distribution.

Contents

- 1 History and etymology
- 2 Definition
 - 2.1 Probability density function
 - 2.2 Cumulative distribution function
 - 2.3 Special cases
- 3 How the *t*-distribution arises
 -

Student's *t*



- 3.1 Sampling distribution
 -
 - 3.2 Bayesian inference
 -
- 4 Characterization
 - - 4.1 As the distribution of a test statistic
 - - 4.1.1 Derivation
 - - 4.2 As a maximum entropy distribution
- 5 Properties
 - - 5.1 Moments
 -
 - 5.2 Relation to *F*-distribution
 -
 - 5.3 Monte Carlo sampling
 -
 - 5.4 Integral of Student's probability density function and *p*-value
 -
 - 5.5 Differential equation
 -
- 6 Non-standardized Student's *t*-distribution
 - - 6.1 In terms of scaling parameter σ , or σ^2
 -
 - 6.2 In terms of inverse scaling parameter λ
 -
- 7 Related distributions
 - - 7.1 Noncentral *t*-distribution
 -
 - 7.2 Discrete Student's *t*-distribution
- 8 Further generalizations
 -
- 9 Uses
 - - 9.1 In frequentist statistical inference
 - - 9.1.1 Hypothesis testing
 - 9.1.2 Confidence intervals
 - 9.1.3 Prediction intervals
 - - 9.2 In Bayesian statistics
 - - 9.3 Robust parametric modeling
- 10 Table of selected values
 -
- 11 See also
 -
- 12 Notes
 -
- 13 References
 -

Entropy	<div>$\frac{\nu+1}{2} \left[\psi \left(\frac{1+\nu}{2} \right) - \psi \left(\frac{\nu}{2} \right) \right]$$+ \ln \left[\sqrt{\nu} B \left(\frac{\nu}{2}, \frac{1}{2} \right) \right] \text{ (nats)}$<div><ul style="list-style-type: none">▪ ψ: digamma function,▪ B: beta function</div></div>
MGF	undefined
CF	<div>$\frac{K_{\nu/2}(\sqrt{\nu} t) \cdot (\sqrt{\nu} t)^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2-1}}$ for $\nu > 0$<div><ul style="list-style-type: none">▪ $K_{\nu}(x)$: Modified Bessel function of the second kind^[1]</div></div>

History and etymology

In statistics, the *t*-distribution was first derived as a posterior distribution in 1876 by Helmert^{[2][3][4]} and Lüroth.^{[5][6][7]}

In the English-language literature it takes its name from William Sealy Gosset's 1908 paper in *Biometrika* under the pseudonym "*Student*".^{[8][9]} 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 low as 3. 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 test the quality of raw material.^[10]

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 referred to the value as *t*.^{[11][12]}



Statistician William Sealy Gosset, known as "Student".

Definition

Probability density function

Student's ***t*-distribution** has the probability density function given by

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

where ν is the number of *degrees of freedom* and Γ is the gamma function. This may also be written as

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

where B is the Beta function.

For ν 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 ν odd,

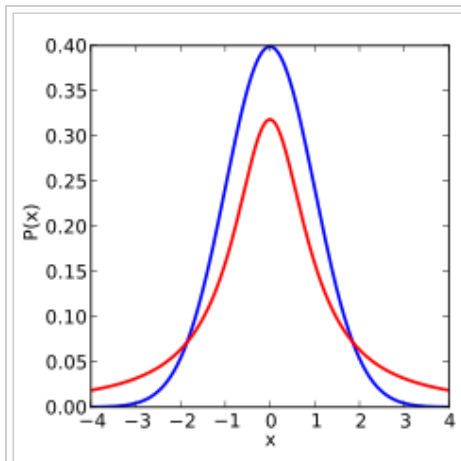
$$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} = \frac{(\nu-1)(\nu-3)\cdots 4\cdot 2}{\pi\sqrt{\nu}(\nu-2)(\nu-4)\cdots 5\cdot 3}.$$

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.

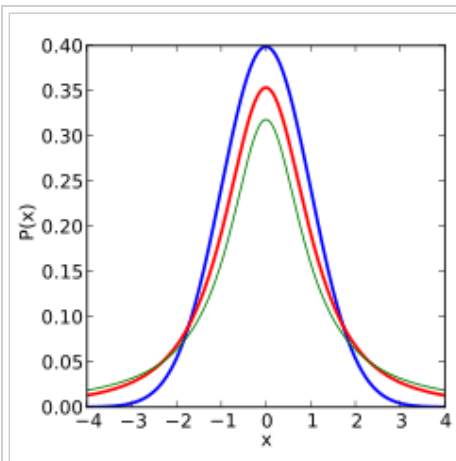
The following images show the density of the t -distribution for increasing values of ν . The normal distribution is shown as a blue line for comparison. Note that the t -distribution (red line) becomes closer to the normal distribution as ν increases.

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

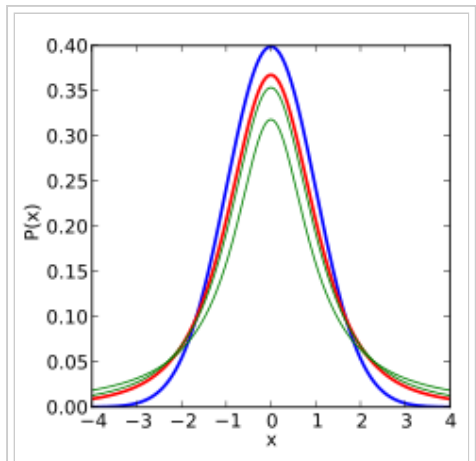
Previous plots shown in green.



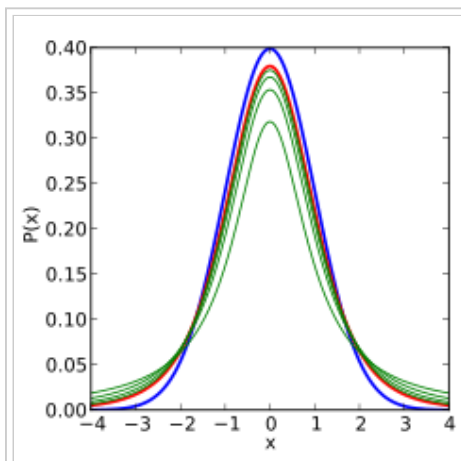
1 degree of freedom



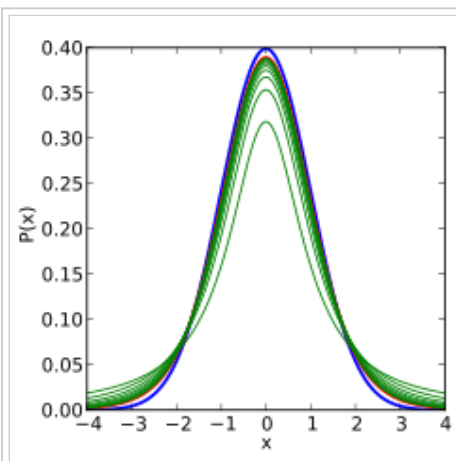
2 degrees of freedom



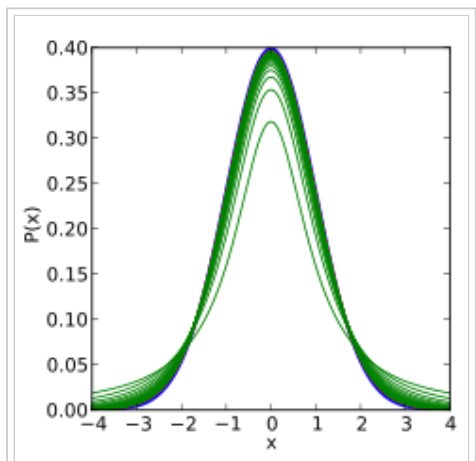
3 degrees of freedom



5 degrees of freedom



10 degrees of freedom



30 degrees of freedom

Cumulative distribution function

The cumulative distribution function can be written in terms of I , the regularized incomplete beta function. For $t > 0$,^[13]

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

with

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

Other values would be obtained by symmetry. An alternative formula, valid for $t^2 < \nu$, is^[13]

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

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

Special cases

Certain values of ν give an especially simple form.

▪ $\nu = 1$

Distribution function:

$$F(t) = \frac{1}{2} + \frac{1}{\pi} \arctan(t).$$

Density function:

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

See Cauchy distribution

▪ $\nu = 2$

Distribution function:

$$F(t) = \frac{1}{2} + \frac{t}{2\sqrt{2+t^2}}.$$

Density function:

$$f(t) = \frac{1}{(2+t^2)^{\frac{3}{2}}}.$$

▪ $\nu = 3$

Density function:

$$f(t) = \frac{6\sqrt{3}}{\pi(3+t^2)^2}.$$

▪ $\nu = \infty$

Density function:

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

See Normal distribution

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 μ . 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 (μ , σ^2) are unknown population parameters, in the sense that the *t*-value has then a probability distribution that depends on neither μ nor σ^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 marginalised out.^[14]

$$p(\mu \mid D, I) = \int p(\mu, \sigma^2 \mid D, I) d\sigma^2$$

$$= \int p(\mu \mid D, \sigma^2, I) p(\sigma^2 \mid D, I) d\sigma^2$$

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 μ given the data and σ^2 with the marginal distribution of σ^2 given the data.

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

$$p(\mu \mid D, \sigma^2, I) \sim N(\bar{x}, \sigma^2/n)$$

$$p(\sigma^2 \mid D, I) \sim \text{Scale-inv-}\chi^2(\nu, s^2)$$

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 marginalisation integral thus becomes

$$p(\mu|D, I) \propto \int_0^\infty \frac{1}{\sqrt{\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}n(\mu - \bar{x})^2\right) \cdot \sigma^{-\nu-2} \exp(-\nu s^2/2\sigma^2) d\sigma^2$$

$$\propto \int_0^\infty \sigma^{-\nu-3} \exp\left(-\frac{1}{2\sigma^2}(n(\mu - \bar{x})^2 + \nu s^2)\right) d\sigma^2$$

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

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

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

$$p(\mu | D, I) \propto A^{-\frac{\nu+1}{2}} \propto \left(1 + \frac{n(\mu - \bar{x})^2}{\nu s^2}\right)^{-\frac{\nu+1}{2}}$$

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 standardised 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 μ and σ^2 ; but it will be apparent that any priors which 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 $P(\mu|D, I)$, although the scaling parameter corresponding to 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 ν degrees of freedom can be defined as the distribution of the random variable T with [13][15]

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

where

- Z is a standard normal with expected value 0 and variance 1;
- V has a chi-squared distribution with ν degrees of freedom;
- Z and V are independent.

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

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

This random variable has a noncentral t -distribution with noncentrality parameter μ . 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 μ and variance σ^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).^[16] 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 μ and variance σ^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 σ is replaced by the random variable S_n , has a Student's t -distribution as defined above. Notice that the unknown population variance σ^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 ν equal to $n - 1$, and Fisher proved it in 1925.^[11]

The distribution of the test statistic, T , depends on ν , but not μ or σ ; the lack of dependence on μ and σ 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 $E(\ln(\nu + X^2))$ is fixed.^[17]

Properties

Moments

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

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

Moments of order ν or higher do not exist.^[18]

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

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

For a t -distribution with ν degrees of freedom, the expected value is 0, 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 $\nu > 4$.

Relation to F -distribution

- $Y \sim F(\nu_1 = 1, \nu_2 = \nu)$ has an F -distribution if $Y = X^2$ and $X \sim t(\nu)$ has a Student's t -distribution.

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.^[19] It has the merit that it applies equally well to all real positive degrees of freedom, ν , while many other candidate methods fail if ν is close to zero.^[19]

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

The function $A(t|\nu)$ 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 $A(t|\nu)$ 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 ν degrees of freedom, $A(t|\nu)$ 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:

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

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

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

Differential equation

The pdf of the t -distribution is a solution to the following differential equation:

$$\left\{ \begin{array}{l} (\nu + x^2) f'(x) + (\nu + 1) x f(x) = 0, \\ f(1) = \frac{\nu^{\nu/2} (\nu+1)^{-\frac{\nu}{2}-\frac{1}{2}}}{B(\frac{\nu}{2}, \frac{1}{2})} \end{array} \right\}$$

Non-standardized Student's t -distribution

In terms of scaling parameter σ , or σ^2

Student's t distribution can be generalized to a three parameter location-scale family, introducing a location parameter μ and a scale parameter σ , through the relation

$$X = \mu + \sigma T$$

or

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

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

The resulting **non-standardized Student's t -distribution** has a density defined by^[20]

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

Here, σ 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. σ simply sets the overall scaling of the distribution. In the Bayesian derivation of the marginal distribution of an unknown normal mean μ above, σ 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 σ^2 , the square of this scale parameter:

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

Other properties of this version of the distribution are:^[20]

$$E(X) = \mu \quad \text{for } \nu > 1,$$

$$\text{var}(X) = \sigma^2 \frac{\nu}{\nu - 2} \quad \text{for } \nu > 2,$$

$$\text{mode}(X) = \mu.$$

This distribution results from compounding a Gaussian distribution (normal distribution) with mean μ and unknown variance, with an inverse gamma distribution placed over the variance with parameters $a = \nu/2$ and $b = \nu\sigma^2/2$. 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 ν and σ^2 . The scaled-inverse-chi-squared distribution is exactly the same distribution as the inverse gamma distribution, but with a different parameterization, i.e. $\nu = 2a$, $\sigma^2 = b/a$.

In terms of inverse scaling parameter λ

An alternative parameterization in terms of an inverse scaling parameter λ (analogous to the way precision is the reciprocal of variance), defined by the relation $\lambda = \frac{1}{\sigma^2}$. Then the density is defined by^[21]

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

Other properties of this version of the distribution are:^[21]

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

This distribution results from compounding a Gaussian distribution with mean μ 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

Noncentral *t*-distribution

The noncentral *t*-distribution is a different way of generalizing 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).

Discrete Student's *t*-distribution

The **discrete Student's *t*-distribution** is defined by its probability mass function at r being proportional to^[22]

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

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.^[23]

Further generalizations

We can generate Student-t samples by taking the ratio of variables from the normal distribution and the square-root of chi-squared distribution. If we use instead of the normal distribution e.g. the Irwin-Hall, 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 way it is e.g. also more flexible than some other symmetric generalizations of the Gaussian distribution.

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 ρ , 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

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

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 μ . 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 σ^{-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.^[24] It is frequently the case that real data have heavier tails than the normal distribution allows for. The classical approach was to identify outliers 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.

Lange et al. explored the use of the t -distribution for robust modeling of heavy tailed data in a variety of contexts. A Bayesian account can be found in Gelman et al. 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.

Table of selected values

Most statistical textbooks list t distribution tables. Nowadays, the better way to a fully precise critical t value or a cumulative probability is the statistical function implemented in spreadsheets (Office Excel, OpenOffice Calc, etc.), or an interactive calculating web page. The relevant spreadsheet functions are TDIST and TINV, while online calculating pages save troubles like positions of parameters or names of functions. For example, a MediaWiki page supported by R extension can easily give the interactive result of critical values or cumulative probability, even for noncentral t -distribution.

The following table lists a few selected values for t -distributions with v degrees of freedom for a range of *one-sided* or *two-sided* critical regions. For an example of how to read this table, take the fourth row, which begins with 4; that means v , the number of degrees of freedom, is 4 (and if we are dealing, as above, with n values with a fixed sum, $n = 5$). Take the fifth entry, in the column headed 95% for *one-sided* (90% for *two-sided*). The value of that entry is "2.132". Then the probability that T is less than 2.132 is 95% or $\Pr(-\infty < T < 2.132) = 0.95$; this also means that $\Pr(-2.132 < T < 2.132) = 0.9$.

This can be calculated by the symmetry of the distribution,

$$\Pr(T < -2.132) = 1 - \Pr(T > -2.132) = 1 - 0.95 = 0.05,$$

and so

$$\Pr(-2.132 < T < 2.132) = 1 - 2(0.05) = 0.9.$$

Note that the last row also gives critical points: a t -distribution with infinitely many degrees of freedom is a normal distribution. (See Related distributions above).

The first column is the number of degrees of freedom.

<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	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
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
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

The number at the beginning of each row in the table above is v which has been defined above as $n - 1$. The percentage along the top is $100\%(1 - \alpha)$. The numbers in the main body of the table are $t_{\alpha, v}$. If a quantity T is distributed as a Student's t distribution with v degrees of freedom, then there is a probability $1 - \alpha$ that T will be less than $t_{\alpha, v}$. (Calculated as for a one-tailed or one-sided test, as opposed to a two-tailed test.)

For example, given a sample with a sample variance 2 and sample mean of 10, taken from a sample set of 11 (10 degrees of freedom), using the formula

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

We can determine that at 90% confidence, we have a true mean lying below

$$10 + 1.37218 \frac{\sqrt{2}}{\sqrt{11}} = 10.58510.$$

(In other words, on average, 90% of the times that an upper threshold is calculated by this method, this upper threshold exceeds the true mean.) And, still at 90% confidence, we have a true mean lying over

$$10 - 1.37218 \frac{\sqrt{2}}{\sqrt{11}} = 9.41490.$$

(In other words, on average, 90% of the times that a lower threshold is calculated by this method, this lower threshold lies below the true mean.) So that at 80% confidence (calculated from $1 - 2 \times (1 - 90\%) = 80\%$), we have a true mean lying within the interval

$$\left(10 - 1.37218 \frac{\sqrt{2}}{\sqrt{11}}, 10 + 1.37218 \frac{\sqrt{2}}{\sqrt{11}} \right) = (9.41490, 10.58510).$$

(In other words, on average, 80% of the times that upper and lower thresholds are calculated by this method, the true mean is both below the upper threshold and above the lower threshold. This is not the same thing 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.)

For information on the inverse cumulative distribution function see *quantile function*.

See also

- Chi-squared distribution
- *F*-distribution
- Gamma distribution
- Hotelling's *T*-squared distribution
- Multivariate Student distribution
- *t*-statistic
- Tau-distribution, for internally studentized residuals
- Wilks' lambda distribution
- Wishart distribution

Notes

1. Hurst, Simon. *The Characteristic Function of the Student-t Distribution* (<http://wwwmaths.anu.edu.au/research.reports/srr/95/044/>), Financial Mathematics Research Report No. FMRR006-95, Statistics Research Report No. SRR044-95 Archived (<https://web.archive.org/web/20150430193238/http://wwwmaths.anu.edu.au/research.reports/srr/95/044/>) April 30, 2015 at the Wayback Machine
2. Helmert, F. R. (1875). "Über die Bestimmung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler". *Z. Math. Phys.*, 20, 300–3.
3. Helmert, F. R. (1876a). "Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und über einige damit in Zusammenhang stehende Fragen". *Z. Math. Phys.*, 21, 192–218.
4. Helmert, F. R. (1876b). "Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers directer Beobachtungen gleicher Genauigkeit", *Astron. Nachr.*, 88, 113–32.
5. Lüroth, J (1876). "Vergleichung von zwei Werten des wahrscheinlichen Fehlers". *Astron. Nachr.* **87** (14): 209–20. Bibcode:1876AN.....87..209L. doi:10.1002/asna.18760871402.

6. Pfanzagl, J.; Sheynin, O. (1996). "A forerunner of the t -distribution (Studies in the history of probability and statistics XLIV)". *Biometrika* **83** (4): 891–898. doi:10.1093/biomet/83.4.891. MR 1766040.
7. Sheynin, O. (1995). "Helmert's work in the theory of errors". *Arch. Hist. Ex. Sci.* **49**: 73–104. doi:10.1007/BF00374700.
8. "Student" [William Sealy Gosset] (March 1908). "The probable error of a mean" (PDF). *Biometrika* **6** (1): 1–25. doi:10.1093/biomet/6.1.1.
9. "Student" (William Sealy Gosset), original Biometrika paper as a scan (http://www.atmos.washington.edu/~robwood/teaching/451/student_in_biometrika_vol6_no1.pdf)
10. Mortimer, Robert G. (2005) *Mathematics for Physical Chemistry*, Academic Press. 3 edition. ISBN 0-12-508347-5 (page 326)
11. Fisher, R. A. (1925). "Applications of "Student's" distribution" (PDF). *Metron* **5**: 90–104.
12. Walpole, Ronald; Myers, Raymond; Myers, Sharon; Ye, Keying. (2002) *Probability and Statistics for Engineers and Scientists*. Pearson Education, 7th edition, pg. 237 ISBN 81-7758-404-9
13. Johnson, N.L., Kotz, S., Balakrishnan, N. (1995) *Continuous Univariate Distributions, Volume 2*, 2nd Edition. Wiley, ISBN 0-471-58494-0 (Chapter 28)
14. A. Gelman *et al* (1995), *Bayesian Data Analysis*, Chapman & Hall. ISBN 0-412-03991-5. p. 68
15. Hogg & Craig (1978, Sections 4.4 and 4.8.)
16. Cochran, W. G. (April 1934). "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance". *Mathematical Proceedings of the Cambridge Philosophical Society* **30** (2): 178–191. Bibcode:1934PCPS...30..178C. doi:10.1017/S0305004100016595.
17. Park, Sung Y.; Bera, Anil K. (2009). "Maximum entropy autoregressive conditional heteroskedasticity model" (PDF). *Journal of Econometrics* (Elsevier): 219–230. Retrieved 2011-06-02.
18. See, for example, page 56 of Casella and Berger, *Statistical Inference*, 1990 Duxbury.
19. Bailey, R. W. (1994). "Polar Generation of Random Variates with the t -Distribution". *Mathematics of Computation* **62** (206): 779–781. doi:10.2307/2153537.
20. Jackman, Simon (2009). *Bayesian Analysis for the Social Sciences*. Wiley. p. 507.
21. Bishop, C.M. (2006). *Pattern recognition and machine learning*. Springer.
22. Ord, J.K. (1972) *Families of Frequency Distributions*, Griffin. ISBN 0-85264-137-0 (Table 5.1)
23. Ord, J.K. (1972) *Families of Frequency Distributions*, Griffin. ISBN 0-85264-137-0 (Chapter 5)
24. Lange, Kenneth L.; Little, Roderick J.A.; Taylor, Jeremy M.G. (1989). "Robust statistical modeling using the t -distribution". *JASA* **84** (408): 881–896. doi:10.1080/01621459.1989.10478852. JSTOR 2290063.

References

- Senn, S.; Richardson, W. (1994). "The first t -test". *Statistics in Medicine* **13** (8): 785–803. doi:10.1002/sim.4780130802. PMID 8047737.
- Hogg, R. V.; Craig, A. T. (1978). *Introduction to Mathematical Statistics*. New York: Macmillan.
- Venables, W. N.; Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth ed.). Springer.
- Gelman, Andrew; John B. Carlin; Hal S. Stern; Donald B. Rubin (2003). *Bayesian Data Analysis (Second Edition)*. CRC/Chapman & Hall. ISBN 1-58488-388-X.

External links

- Hazewinkel, Michiel, ed. (2001), "Student distribution", *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Earliest Known Uses of Some of the Words of Mathematics (S) (<http://jeff560.tripod.com/s.html>) (*Remarks on the history of the term "Student's distribution"*)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Student%27s_t-distribution&oldid=703866486"

Categories: Continuous distributions | Special functions | Normal distribution
 | Probability distributions with non-finite variance | Infinitely divisible probability distributions
 | Probability distributions | Location-scale family probability distributions

- This page was last modified on 8 February 2016, at 03:19.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.