Chi-Square Distribution

If X_i are k independent, normally distributed random variables with mean 0 and variance 1, then the random variable

$$Q = \sum_{i=1}^{k} X_i^2$$

is distributed according to the chi-square distribution. This is usually written

$$Q \sim \chi_k^2.$$

The chi-square distribution has one parameter: k - a positive integer that specifies the number of degrees of freedom (i.e. the number of X_i)

Probability density function

A probability density function of the chi-square distribution is

$$f(x;k) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)}\, x^{k/2-1} e^{-x/2} & \text{for } x > 0, \\ 0 & \text{for } x \le 0, \end{cases}$$

where Γ denotes the Gamma function, which takes particular values at the half-integers.

In mathematics, the **Gamma function** (represented by the capitalized Greek letter **Γ**) is an extension of the factorial function to real and complex numbers. For a complex number z with positive real part it is defined by

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, \mathrm{d}t$$

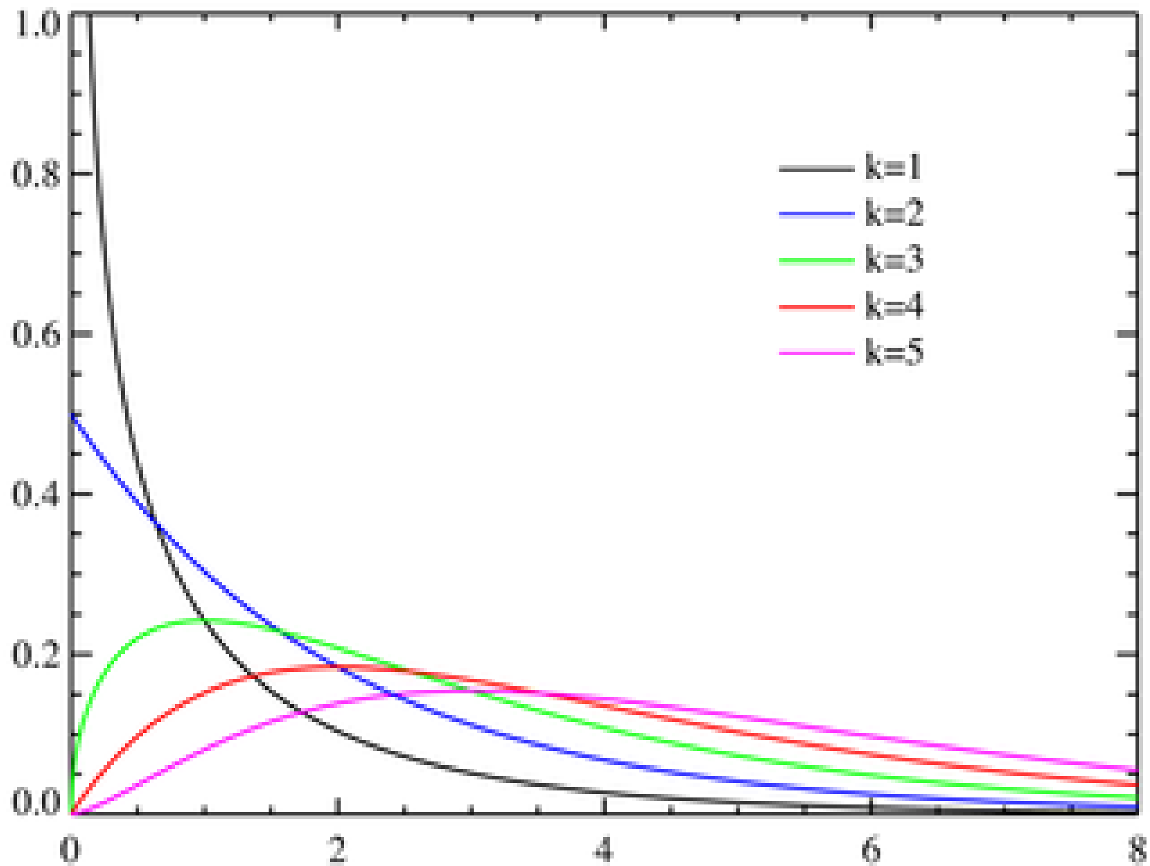
which can be extended to the rest of the complex plane, excepting the non-positive integers.

If z is a positive integer, then

$$\Gamma(z) = (z-1)!$$

chi-square

Probability density function



Degrees of Freedom

Estimates of parameters can be based upon different amounts of information. The number of independent pieces of information that go into the estimate of a parameter is called the degrees of freedom (df).

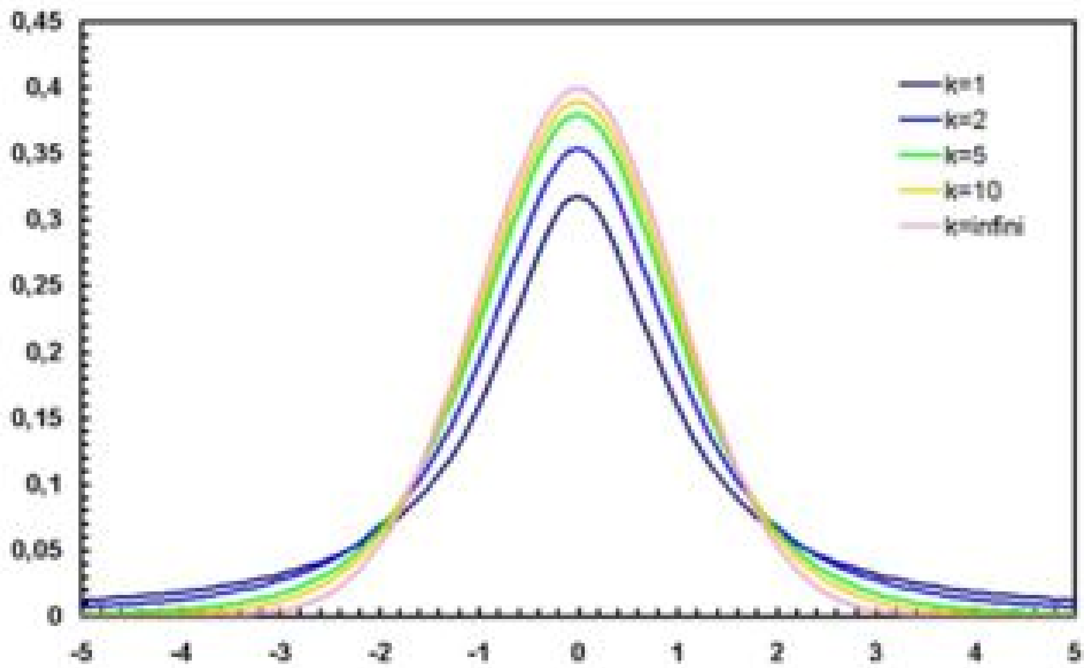
In general, the degrees of freedom of an estimate is equal to the number of independent scores that go into the estimate minus the number of parameters estimated as intermediate steps in the estimation of the parameter itself.

For example, if the variance, σ^2, is to be estimated from a random sample of N independent scores, then the degrees of freedom is equal to the number of independent scores (N) minus the number of parameters estimated as intermediate steps (one, μ estimated by sample mean) and is therefore equal to N-1.

Student's t-distribution

Student's *t*

Probability density function



Parameters	$\nu > 0$ degrees of freedom (real)
Support	$x \in (-\infty; +\infty)$
Probability density function (pdf)	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-(\frac{\nu+1}{2})}$
Mean	0 for $\nu > 1$, otherwise undefined
Median	0
Mode	0
Variance	$\frac{\nu}{\nu-2}$ for $\nu > 2$, otherwise undefined
Moment-generating function (mgf)	(Not defined)

The **Student's *t*-distribution** (or also ***t*-distribution**).

The derivation of the *t*-distribution was first published in 1908 by William Sealy Gosset, while he worked at a Guinness Brewery in Dublin. He was not allowed to publish under his own name, so the paper was written under the pseudonym *Student*. The *t*-test and the associated theory became well-known through the work of R.A. Fisher, who called the distribution "Student's distribution".

Student's distribution arises when (as in nearly all practical statistical work) the population standard deviation is unknown and has to be estimated from the data. Textbook problems treating the standard deviation as if it were known are 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.

Why use the Student's t*-distribution*

Confidence intervals and hypothesis tests rely on Student's *t*-distribution to cope with uncertainty resulting from estimating the standard deviation from a sample, whereas if the population standard deviation were known, a normal distribution would be used.

How Student's t*-distribution comes about*

Suppose $X_1, \ldots, X_n$ are independent random variables that are normally distributed with expected value μ and variance σ^2. Let

$$\overline{X}_n = (X_1 + \cdots + X_n)/n$$

be the sample mean, and

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

be the sample variance. It is readily shown that the quantity

$$Z = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

is normally distributed with mean 0 and variance 1, since the sample mean $\overline{X}_n$ is normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$.

Gosset studied a related quantity,

$$T = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}},$$

which differs from *Z* in that the exact standard deviation σ is replaced by the random variable S_n. Technically, $(n-1)S_n^2/\sigma^2$ has a χ_{n-1}^2 distribution by Cochran's theorem. Gosset's work showed that *T* has the probability density function

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

with ν equal to *n* − 1 and where Γ is the Gamma function.

This may also be written as

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

where *B* is the Beta function.

The distribution of *T* is now called the ***t*-distribution**. The parameter ν is called the number of **degrees of freedom**. The distribution 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.

The moments of the t-distribution are

$$E(T^k) = \begin{cases} 0 & \text{k odd}, \quad 0 < k < \nu \\ \frac{\Gamma(\frac{k+1}{2})\Gamma(\frac{\nu-k}{2})\nu^{k/2}}{\sqrt{\pi}\Gamma(\frac{\nu}{2})} & \text{k even}, \quad 0 < k < \nu \\ \text{NaN} & \text{k odd}, \quad 0 < \nu \le k \\ \infty & \text{k even}, \quad 0 < \nu \le k \end{cases}$$

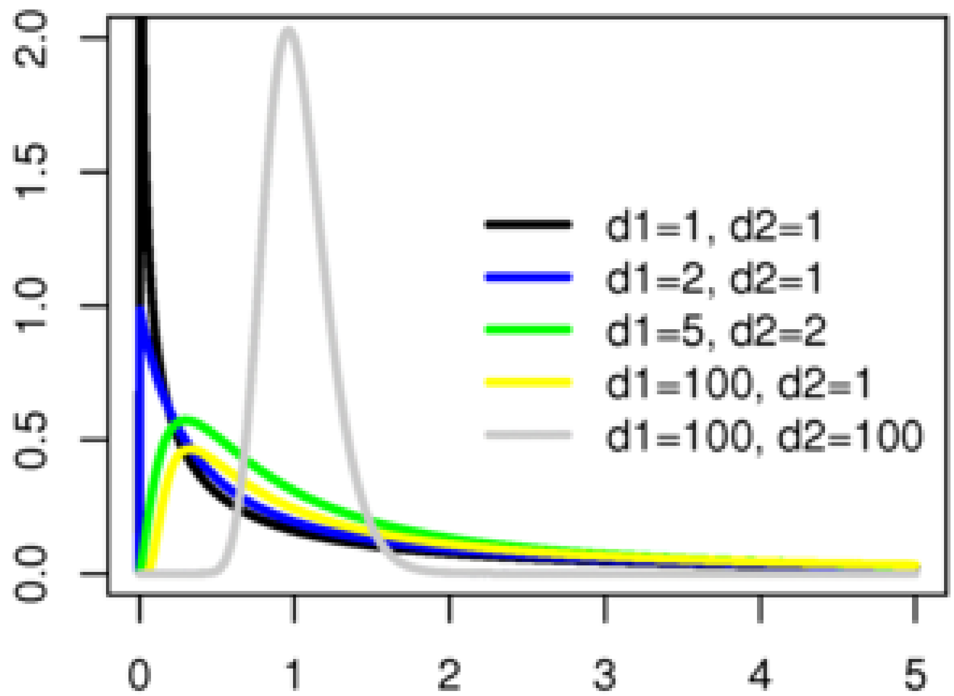
It should be noted that the term for 0 < *k* < ν, *k* even, may be simplified using the properties of the Gamma function to

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

F-distribution

Fisher-Snedecor

Probability density function



Parameters	$d_1 > 0,\ d_2 > 0$ deg. of freedom
Support	$x \in [0; +\infty)$
Probability density function (pdf)	$\frac{\sqrt{\frac{(d_1\,x)^{d_1}\,\,d_2^{d_2}}{(d_1\,x+d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\!\left(\frac{d_1}{2},\frac{d_2}{2}\right)}$
Mean	$\frac{d_2}{d_2-2}$ for $d_2 > 2$
Variance	$\frac{2\,d_2^2\,(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$ for $d_2 > 4$
Moment-generating function (mgf)	*see text for raw moments*

In probability theory and statistics, the ***F*-distribution** is a continuous probability distribution. It is also known as **Snedecor's *F* distribution** or the **Fisher-Snedecor distribution** (after R.A. Fisher and George W. Snedecor).

A random variate of the *F*-distribution arises as the ratio of two chi-squared variates:

$$\frac{U_1/d_1}{U_2/d_2}$$

where

- U_1 and U_2 have chi-square distributions with d_1 and d_2 degrees of freedom respectively, and
- U_1 and U_2 are independent (see Cochran's theorem for an application).

The *F*-distribution arises frequently as the null distribution of a test statistic, especially in likelihood-ratio tests, perhaps most notably in the analysis of variance; see F-test.

The probability density function of an $F(d_1, d_2)$ distributed random variable is given by

$$g(x) = \frac{1}{\mathrm{B}(d_1/2, d_2/2)} \left(\frac{d_1\,x}{d_1\,x + d_2}\right)^{d_1/2} \left(1 - \frac{d_1\,x}{d_1\,x + d_2}\right)^{d_2/2} x^{-1}$$

for real $x \geq 0$, where d_1 and d_2 are positive integers, and B is the beta function.

- One interesting property is that if $X \sim F(\nu_1, \nu_2),\ \frac{1}{X} \sim F(\nu_2, \nu_1)$