Answer on Question #57116 – Math – Statistics and Probability

Write the t distribution and its properties with useful examples.

Solution

The "*t*-distribution", or "central Student's *t* ", distribution $f_T(T_\nu; \nu)$ resulting for the measurable quantity

$$T_\nu = \frac{Z}{\sqrt{\frac{Y_\nu}{\nu}}}$$

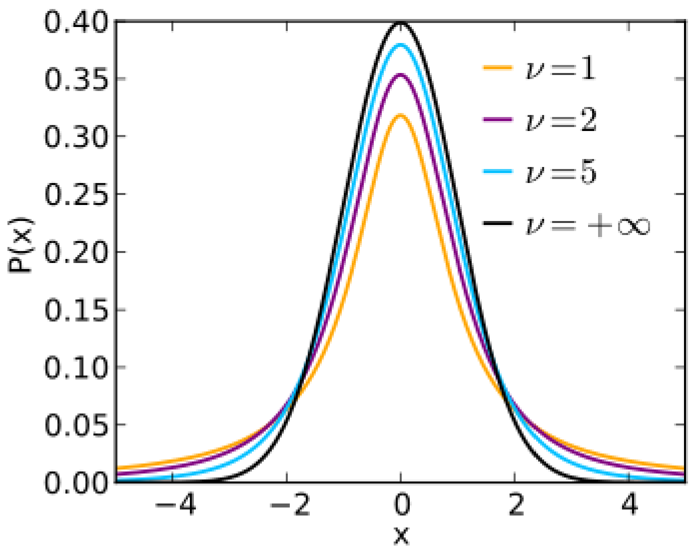
if Z satisfies the standard normal distribution and Y_ν satisfies the chi-squared distribution with ν degrees of freedom, Z and Y_ν are independent.

Probability density function is

$$p(x) = p(x; n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\cdot\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-(n+1)/2},$$

where $x \in (-\infty; +\infty)$,

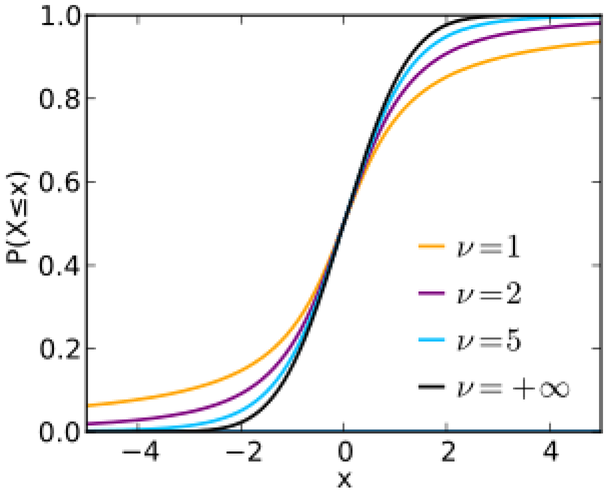
n is the number of degrees of freedom and Γ is the gamma function.



The curve never touches the x axis. It is symmetric, bell-shaped and centered at 0 just like the standard normal density, but is more spread out (higher variance), a lower and wider. The shape of probability density function is dependent on the sample size ν.

Cumulative distribution function is

$$F_T(u; \nu) = P(T_\nu \le u) = \int_{-\infty}^{u} p(x)dx.$$



The parameter of distribution is the random sample size ν.

Expected value is $m = 0$ for $\nu > 1$, otherwise undefined.

Median and mode are zero.

Variance is

$\sigma^2 = \frac{\nu}{\nu-2}$. for $\nu > 2$.

∞ for $1 < \nu \leq 2$

undefined otherwise.

The variance is greater than 1.

The odd raw moments of the t-distribution are zero. Student's t-distribution is a continuous probability distribution that arises when estimating the mean of a normally distributed population and when the sample size is small and population standard deviation is unknown.

The standardized arithmetic mean of a sample as a variable

$$T_s = \frac{\bar{x} - m}{\sigma_\nu}\sqrt{\nu} = \frac{\bar{x} - m}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{\nu}(x_i - \bar{x})^2}}\sqrt{\nu}$$

satisfies the $f_T(T_s; \nu - 1)$ distribution.

Student's *t*-distribution is also applied in the Students'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, in the Bayesian analysis of data from a normal family.

In Excel 2010/2013 T.DIST (x, df, TRUE) is the cumulative distribution function for the *t* -distribution with df degrees of freedom and T.DIST(x, df, FALSE) is the probability density function for the *t* –distribution. In the previous versions of Excel the cumulative distribution function is given by 1-TDIST(*x*,df,1) when $x \geq 0$ and by TDIST(-*x*, df, 1) when $x < 0$.

In R package dt(x,df, ncp,log=FALSE), pt(q, df, ncp, lower.tail=TRUE, log.p=FALSE) give the probability density and cumulative distribution functions respectively.

As the degrees of freedom increases $\nu \to \infty$, the $t-$ distribution converge to the standard normal. The approximation is quite close for $\nu \geq 30$. If $\nu = \infty$ then *t*- distribution can be substituted by the standard normal distribution.

If $\nu = 1$, then *t*-distribution coincides with Cauchy distribution. There exist different generalizations of *t*-distribution (for example, noncentral *t*-distribution, non-standardized Student's *t*-distribution).

Properties of these types can differ. For example, the central *t* distribution is symmetric, while the noncentral *t* is skewed.