t-distribution ($X \sim t(\nu)$)

A *continuous probability distribution* that generalizes the *standard normal distribution* when the *sample* size is small. With growing sample size, the t-distribution becomes asymptotically standard normal. A *random variable* $X \sim t(\nu)$ is t-distributed with ν (nu) *degrees of freedom*. Sometimes the letter T is used to denote t-distributed variables, sometimes the number of degrees of freedom is attached as a subscript: $T_\nu \sim t(\nu)$. The t-distribution is used for various *estimates* and *hypothesis tests*, like location tests and correlation tests; see *t-test for location* and *t-test for regression coefficients*, respectively. The parameters and functions of the t-distribution are shown in figure TDD, sample plots of its probability functions in figure TDC.

$X \sim t(\nu)$	
PDF	$f(x) = \dfrac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu} \cdot \Gamma(\frac{\nu}{2})} \cdot \left(1 + \dfrac{x^2}{2}\right)^{-\frac{\nu+1}{2}}$
CDF	$F(x) = \dfrac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu} \cdot \Gamma(\frac{\nu}{2})} \cdot \int_{-\infty}^{x} \left(1 + \dfrac{w^2}{2}\right)^{-\frac{\nu+1}{2}} dw$
	$F(x) = 1 - \frac{1}{2} I_{\frac{\nu}{x^2+\nu}}(\frac{\nu}{2}, \frac{1}{2})$
Statistic	$x \in \mathbf{R}$: t statistic
Parameters	$\nu \in \mathbf{R}^+$: degrees of freedom
μ	0
σ^2	$\begin{cases} \dfrac{\nu}{\nu-2} & for\ \nu > 2 \\ \infty & for\ 1 < \nu \le 2 \end{cases}$
Skewness (γ_1)	0
Approximations	$N(0, 1)$ for $\nu \ge 30$

Figure **TDD:** t-distribution; $\Gamma(x)$ is the complete gamma function, $I_x(a, b)$ is the regularized incomplete beta function

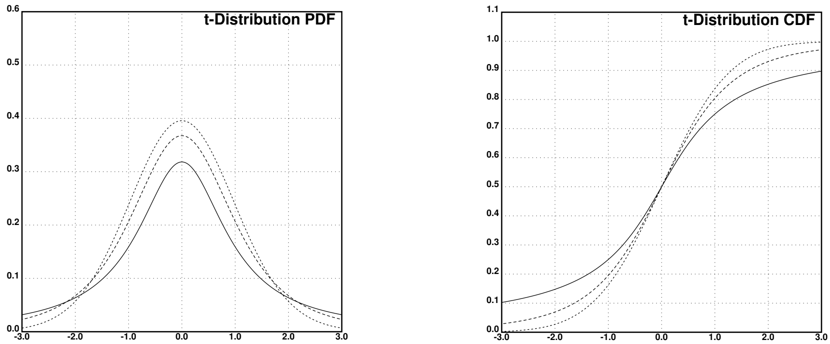

Figure **TDC:** t-distribution probability functions; left PDF with $\nu = 1$ (solid), $\nu = 3$ (dashed), and $\nu = 30$ (dotted); right: CDF with same parameters

An *interval estimator* predicting the *mean* of a *population* given a small sample size of typically less than 30 *specimens* can be created on the basis of the t-distribution. This approach assumes that the population is approximately normally distributed. The estimator is created as follows. The *sampling distribution of the mean* $\bar{X}$ of the given sample is t-distributed with $n-1$ degrees of freedom:

$$\frac{\bar{X}}{\sqrt{s^2/n}} \sim t(n-1)$$

where n is the sample size and s^2 is the sample variance. So the population mean μ falls within the *confidence interval*

$$P(\bar{x} - t \cdot \sqrt{s^2/n} \ \leq \ \bar{X} \ \leq \ \bar{x} + t \cdot \sqrt{s^2/n}) \ = \ c$$

where $\bar{x}$ is the sample mean and $t = F_{t(n-1)}(c + \frac{1-c}{2})$ is the *quantile* of the t-distribution given a probability of c plus half the probability associated with a *critical region* (because the confidence interval is *two-tailed*).

Example: There is a sample of 10 marbles with an average diameter of $\bar{x} = 15mm$ and a variance of $s^2 = 2.5$. Then,

$$\frac{\bar{X}}{\sqrt{2.5/10}} \ = \ \frac{\bar{X}}{0.5} \ \sim \ t(9)$$

and, given a *level of confidence* of $c = 0.9$:

$$F_{t(9)}(0.95) \ \approx \ 1.83$$

and therefore,

$$P(15 - 1.83 \cdot 0.5 \ \leq \ \bar{X} \ \leq \ 15 + 1.83 \cdot 0.5)$$
$$= \ P(14.085 \ \leq \ \bar{X} \ \leq \ 15.915) \ \approx \ 0.9$$

I.e. the probability for the population mean μ to fall within the interval $[14.085, 15.915]$ would be $p = 0.9$.