# Key Probability Distributions in Econometrics

The normal or Gaussian distribution is a symmetrical bell curve. It is found everywhere, and the Central Limit Theorem tells us why: because whenever a large number of independently distributed random variables are added together the sum tends to the normal distribution, even if the distributions of the individual random variables are far from normal.
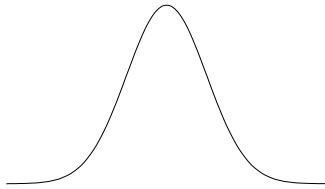
The formula for the normal pdf (density function) is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

where μ denotes the mean of the distribution and σ its standard deviation.

The probability of x falling into any given range can be found by integrating the above pdf from the lower to the upper limit of the range. A couple of results to commit to memory are $P(\mu - 2\sigma < x < \mu + 2\sigma) \approx 0.95$ and $P(\mu - 3\sigma < x < \mu + 3\sigma) \approx 0.997$. A compact notation for saying that x is distributed normally with mean μ and variance σ^2 is $x \sim N(\mu, \sigma^2)$.

The χ^2 distribution represents the distribution of a sum of squares of normal random variables. χ^2 is indexed by a degrees of freedom (df) term which corresponds to the number of normal variables whose squares compose the sum. It is bounded by zero at the low end, and skewed to the right.

In case you were wondering, the pdf for the chi-square distribution with m degrees of freedom (χ^2_m) is
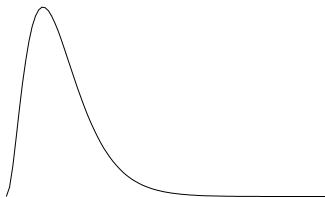
$$f(x) = \frac{e^{-\frac{x}{2}} x^{(\frac{m}{2}-1)}}{2^{\frac{m}{2}} \Gamma(\frac{m}{2})} \qquad x > 0$$

where $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$, $a > 0$ (Euler's gamma function).

This distribution naturally arises in econometrics when we are considering variances, since the numerator of the variance is the sum of squared deviations about the mean.

The t distribution represents the distribution of the ratio of a normal variate to the square root of an independent χ^2. It is indexed by a df term, equal to the df of the χ^2 in the denominator. In an econometric context the df is the sample size minus the number of parameters being estimated. At low degrees of freedom the t distribution is thicker in the tails than the normal distribution but when df is large it becomes practically indistinguishable from the normal.

The F distribution arises as the ratio of two independent chi-squares. If $U \sim \chi^2_m$ and $V \sim \chi^2_n$ then $F_{m,n} = U/m \div V/n$ follows the F distribution, with pdf

$$f(x) = \frac{\Gamma(\frac{m+n}{2})(\frac{m}{n})^{\frac{m}{2}} x^{\frac{m}{2}-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})(1 + \frac{mx}{n})^{\frac{m+n}{2}}}$$

This distribution has a similar shape—a skewed bell—to the χ^2. It is indexed by two df terms, one for the numerator and one for the denominator.

Note the connections among these distributions, with the normal being the "parent" of all the others: Aggregation of independent random variables $\longrightarrow$ normal. Sum of squares of normal variables $\longrightarrow \chi^2$. Ratio of normal to (square root of) chi-square $\longrightarrow t$; ratio of two chi-squares $\longrightarrow F$.