

# Central Limit Theorem

The Central Limit Theorem (CLT) is an asymptotic result of summing probability distributions. Thus, it is very useful to model the sums from random variables (e.g. sums of losses, sales, yields). It also explains why so many distributions sometimes look like the Normal. We won't look at the derivation, just see some examples and its use.

## The CLT result

The sum  $S$  of  $n$  independent random variables  $x(i)$  (where  $n$  is large), all of which have the same distribution will asymptotically take approaches a [Normal](#) distribution with known mean and standard deviation:

$$\sum_{i=1}^n x(i) \approx \text{Normal}\left(n\mu, \sigma\sqrt{n}\right) \quad (1)$$

where  $m$ ,  $s$  are the mean and standard deviation of the distribution from which the  $n$  samples are drawn.

## Examples

Imagine that the distribution of the weight of random nails produced by some company has a mean of 27.4g and standard deviation 1.3g. What will be the weight of a box of 100 nails? The answer is the following Normal distribution:

=Normal(100\*27.4, SQRT(100)\*1.3) grams

=Normal(2740, 13)

This CLT result turns out to be very important in risk analysis. Many of distributions are the sum of a number of identical random variables, and so as that sum gets larger, the distribution tends to look like a Normal distribution. For example:

[Gamma\(0,b,a\)](#) is the sum of a independent [Exponential\(1/b\)](#) distributions, so as  $a$  gets larger, [the Gamma distribution](#) looks progressively more like a Normal distribution. An [Exponential](#) distribution has mean and variance of  $b$ , so we have:

Gamma(0,b,a) Normal(ab, ab) as  $n$

Other examples are discussed in the section on [approximating one distribution by another](#).

## How large does $n$ have to be for the sum to be distributed Normally?

Distribution of individual	Sufficient $n$
Uniform	12 (try it: an old way of generating Normal distributions)

<b>Symmetric triangular</b>	6 (because $U(a,b)+U(a,b) = T(2a,a+b,2b)$ )
<b>Normal</b>	1 !
<b>Skewed</b>	30+ (30 lots of Poisson(2) = Poisson(60) )
<b>Exponential</b>	50+ (check with Gamma(0,b,a) = sum of a Expon(1/b)'s )

## Other related results

### *The average of a large number of independent, identical distributions*

Dividing both sides of Equation (1) by  $n$ , the average  $\bar{x}$  of  $n$  variables drawn independently from the same distribution is given by:

$$\bar{x} = \frac{\text{Normal}(n\mu, \sqrt{n}\sigma)}{n} = \text{Normal}\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad (2)$$

Note: the result of Equation (2) is correct because both the mean and standard deviation of the Normal distribution are in the same units as the variable itself. However, be warned that for most distributions one cannot simply divide by  $n$  the distribution parameters of a variable  $X$  to get the distribution of  $X/n$ . It works for the normal distribution because both parameters are in the same units as  $x$ .

### *The product of a large number of independent, identical distributions*

CLT can also be applied where a large number of identical random variables are being *multiplied* together, for the following reason:

Let  $\Pi$  be the product of a large number of random variables  $x(i)$ ;  $i = 1$  to  $n$ , i.e.:

$$\Pi = \prod_{i=1}^n x(i)$$

Taking logs of both sides, we get:

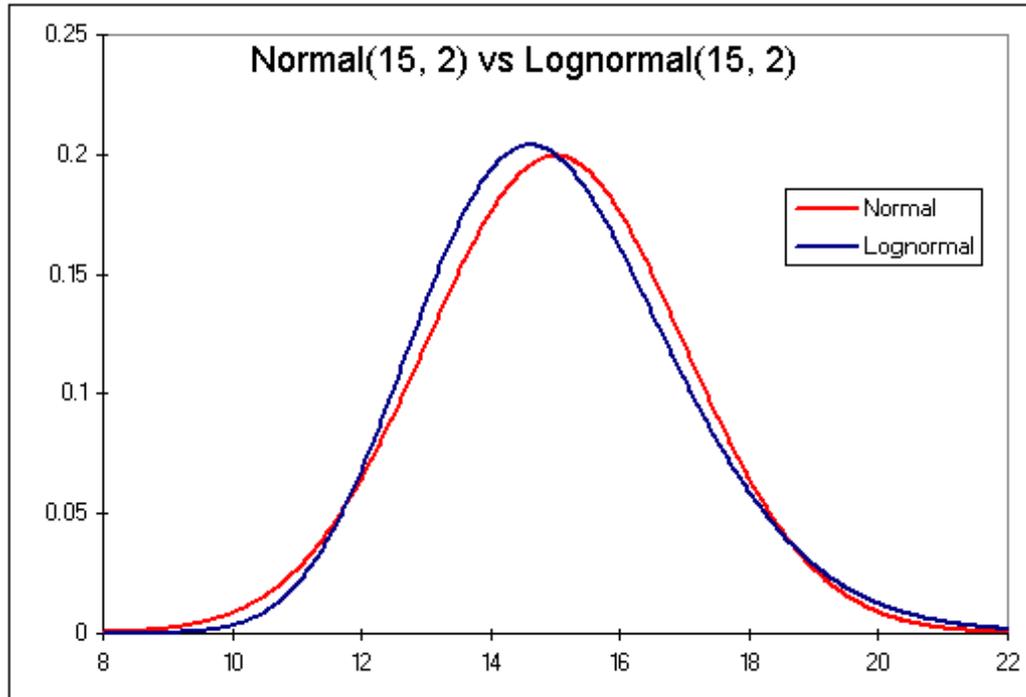
$$\ln[\Pi] = \sum_{i=1}^n \ln[x(i)]$$

The right hand side is the sum of a large number of random variables, and will therefore tend to a Normal distribution. Thus, from the definition of a [Lognormal distribution](#), must be lognormally distributed.

A neat result from this is that if all  $X_i$  are Lognormally distributed, their product will also be Lognormally distributed.

## Is CLT why the Normal distribution is so popular?

Many stochastic variables are neatly described as the sum or product, or a mixture, of a number of random variables. A very loose form of CLT says that if you add up a large number  $n$  of different random variables, and if none of those variables dominate the resultant distribution spread, the sum will eventually look Normal as  $n$  gets bigger. The same applies to multiplying (positive) different random variables and the Lognormal distribution. In fact, a Lognormal distribution will also look very similar to a Normal distribution if its mean is much larger than its standard deviation (see graph below), so perhaps it should not be too surprising that so many variables in nature seem to be somewhere between Lognormally and Normally distributed.



Once you have reviewed the material in this topic, you might like to test how much you have learned by taking the self-test quiz:

Take

A quiz on The Central Limit Theorem: [quiz](#)

### Some Excel functions useful with the Central Limit Theorem

Use	Function	Explanation
Normal probability	=NORM.DIST(x,m,s, cumulative)	The Normal density for x (cumulative = FALSE), or cumulative probability <x (cumulative = TRUE)
Lognormal probability	=LOGNORM.DIST(x,m,s)	The cumulative probability for x where $\ln(x) = \text{Normal}(m,s)$
Lognormal inverse probability	=LOGNORM.INV(P,m,s)	The value x such that $P(\text{variable } x) = P$ where $\ln(x) = \text{Normal}(m,s)$
Normal inverse probability	=NORM.INV(P,m,s)	The value x such that $P(\text{variable } x) = P$
Unit Normal inverse probability	=NORM.S.INV(z)	The value z such that $P(\text{variable } z) = P$ for a Normal(0,1) distribution (i.e. a z-test limit)

