Originally distributed by Dr. Bahman Shafii

Sampling Distribution and the Central Limit Theorem

In many applied problems, it is reasonable to assume that the observable random variables in a random sample, $x_1, x_2, \dots, x_n$ are independent with a common normal density function. In such situations, the following theorem establishes the sampling distribution of the statistic $\bar{x}$.

Theorem:
Let $x_1, x_2, \dots, x_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is normally distributed with mean μ and variance ${\sigma^2}/{n}$ or short-hand is $\bar{x} \sim N\left(\mu, {\sigma^2}/{n}\right)$

Notice that since under this theorem $\bar{x}$ is distributed with mean μ and variance $\sigma_{\bar{x}}^2 = {\sigma^2}/{n}$, it follows that:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \sqrt{n}\frac{\bar{x} - \mu}{\sigma}$$

has a standard normal distribution.

Also observe that the above theorem requires $x_i, i = 1, 2, \dots, n$, to be normally distributed. What can we then say about the sampling distribution of $\bar{x}$ if the $x_i's$ are <u>not</u> normally distributed? The following theorem, called the Central Limit Theorem (CLT) establishes the essence of the answer to this question:

<u>Theorem (CLT):</u>
Let $x_1, x_2, \dots, x_n$ be independent and identically distributed random variables with $E(x_i) = \mu$ and $V(x_i) = \sigma^2$. Then the distribution of U_n given by:

$$U_n = \sqrt{n}\frac{\bar{x} - \mu}{\sigma}$$

converges to a <u>standard normal</u> distribution function as $n \to \infty$. Alternatively stated,

$$P(a \le U_n \le b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du \text{ as } n \to \infty$$

That is, probability statements about U_n can be approximated by the corresponding probabilities for the standard normal random variable if n is large.[1]

It is very important for you to notice that the CLT can be applied for a random sample $x_1, x_2, \dots, x_n$ from <u>any</u> distribution, so long as $E(x_i) = \mu$ and $V(x_i) = \sigma^2$ are both finite and the sample size is large.

The significance of the CLT is twofold. First, it explains why some measurements tend to posses approximately a normal distribution. That is, CLT provides an explanation of rather common occurrences of normally distributed random variables in nature. Second, and more important, is the contribution of CLT to <u>statistical inference</u>. Many estimators and decision tools that are used to make inferences about population parameters are sums or averages of the sample measurements. When the sample size, n, is sufficiently large, we would expect such estimators or decision makers to possess a normally probability distribution in repeated sampling according to the CLT.

---

[1]Usually, a value of $n > 30$ will ensure that the distribution of U_n can be closely approximated by a normal distribution.