Class notes #1 for MTH 481: Spring 2009

Central Limit Theorem Let $X_1, X_2, \ldots$ be independent random variables, all with the same pdf. Thus the X_i all have the same mean and standard deviation: define

$$\mu = \mathrm{E}[X_i], \quad \sigma = \mathrm{STD}[X_i]$$

For each $n \geq 1$ define

$$Y_n = X_1 + \cdots + X_n$$

$$\overline{Y}_n = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} Y_n$$

So Y_n is the sum of the first n random variables, and $\overline{Y}_n$ is the average of the first n random variables.

Using the standard rules (the ones that 'everybody forgets') deduce that

$$\mathrm{E}[Y_n] = n\,\mu, \qquad \mathrm{STD}[Y_n] = \sqrt{n}\,\sigma$$

and also

$$\mathrm{E}[\overline{Y}_n] = \mu, \qquad \mathrm{STD}[\overline{Y}_n] = \frac{\sigma}{\sqrt{n}}$$

The Central Limit Theorem (CLT) says that when n is large, both Y_n and $\overline{Y}_n$ are approximately normal random variables, with the approximation becoming exact as $n \to \infty$. Typically the approximation is pretty good for $n \geq 10$, and really good for $n \geq 30$.

In practice, you use the result as follows. Write

$$Y_n = \sqrt{n}\,\sigma\,Z + n\,\mu$$

$$\overline{Y}_n = \frac{\sigma}{\sqrt{n}}\,Z + \mu$$

and proceed as if Z is a standard normal in both cases.

Example

Suppose $X_1, X_2, \ldots$ are independent uniform random variables on $[0,1]$. Let Y_n and $\overline{Y}_n$ be the sum and average of the first n variables, as above. Compute $P(Y_{25} \leq 15)$ and $P(\overline{Y}_{100} \geq 0.55)$.

Note that $\mu = 0.5$ and $\sigma = 1/2\sqrt{3} = 0.289$. So

$$Y_{25} = 1.445\,Z + 12.5$$

hence

$$P(Y_{25} \leq 15) = P(1.445\,Z + 12.5 \leq 15) = P(Z \leq 1.73) = 0.9582$$

Also

$$\overline{Y}_{100} = 0.0289\,Z + 0.5$$

and therefore

$$P(\overline{Y}_{100} \geq 0.55) = P(0.0289\,Z + 0.5 \geq 0.55) = P(Z \geq 1.73) = 0.0418$$