A bit of a review on continuous probability distributions.

If we want the average roll of a die, we'd do a simple sum:

value on face of die

probability of rolling that face

$$\langle i \rangle = \sum_{i=1}^{6} i \, P_i$$

Suppose we're measuring a property of those faces (average reflectivity or some such nonsense)

reflectivity of face i

probability of rolling that face

$$\langle r \rangle = \sum_{i=1}^{6} r(i) \, P_i$$

What are the equivalents if we have a continuous distribution of outcomes: (e.g. a "Normal" Distribution)

$$\underbrace{p(x)\,dx}_{\text{probability of measuring a value between } x \text{ \& } x+dx} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \qquad -\infty \le x \le \infty$$

Averages:

normalization

$$\langle x \rangle = \int x\,p(x)\,dx \Big/ \underbrace{\int p(x)\,dx}_{\text{normalization}}$$

Properties:

property r depends on x

$$\langle r \rangle = \int r(x)\,p(x)\,dx \Big/ \int p(x)\,dx$$

Some very common continuous distributions:

Gaussian: $p(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$

Exponential: $p(x)\,dx = a e^{-ax}$ $\quad 0 \leq x \leq \infty$

Uniform: $p(x)\,dx = \frac{1}{a}$ $\quad 0 \leq x \leq a$

If we have a normalized distribution,

$\int p(x)\,dx = 1$, we can define some very useful quantities:

Average or Mean: $\langle x \rangle = \int x\, p(x)\, dx$

Second Moment: $\langle x^2 \rangle = \int x^2 p(x)\, dx$

Variance: $\sigma_x^2 = \langle (x - \langle x \rangle)^2 \rangle = \int (x - \langle x \rangle)^2 p(x)\, dx$

$= \langle x^2 \rangle - \langle x \rangle^2$

Standard Deviation: $\sigma_x = \sqrt{\langle (x - \langle x \rangle)^2 \rangle}$

$\sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$

The Central Limit Theorem

If $n_1, n_2, \ldots$ are independent, identically distributed random variables with mean $\langle n \rangle$ an finite, nonzero variance σ^2

$$S_N = n_1 + n_2 + \ldots n_N \quad \leftarrow \text{sum of } N \text{ trials}$$

Then $$\lim_{N \to \infty} P\left(\frac{S_N - N\langle n \rangle}{\sigma \sqrt{N}} \leq x \right) = \Phi(x)$$

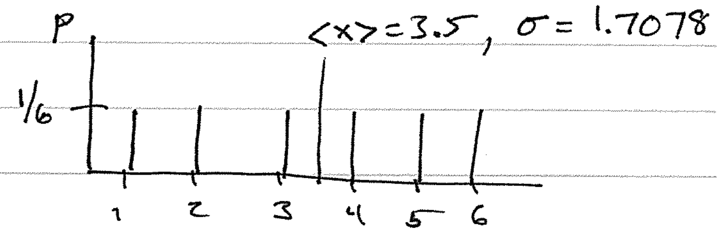
where $\Phi(x)$ is the probability that a normal Gaussian variable will be less than x.

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-(x' - \langle n \rangle)^2 / 2\sigma^2} \, dx'$$

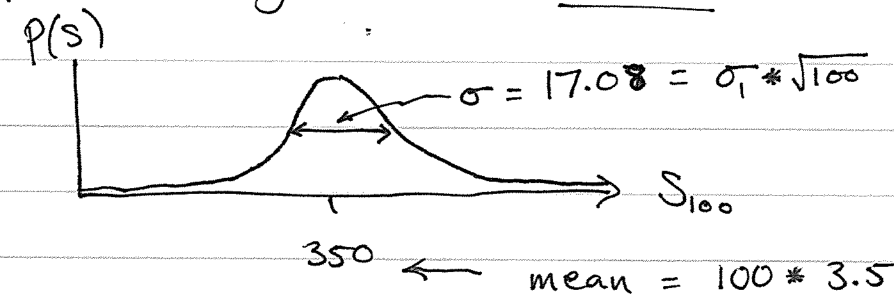
What does this mean?

Consider a fair die:

$P_i = \frac{1}{6}$



If we look at the results of rolling 100 dice and summing these 100 rolls, we would expect to get a Gaussian distribution:



We get this in the limit of many experiments.

The most mind-blowing aspect of the CLT is that it doesn't matter if we're using cheater dice or fair ones. As long as we can find $\langle i \rangle$ & $\langle \sigma \rangle$, the results for summing 100 rolls will <u>still</u> be gaussian!