Central Limit Theorem

There are two ways that you could state the central limit theorem. Either that the sum of IID random variables is normally distributed, or that the average of IID random variables is normally distributed.

The Central Limit Thorem (Sum Version)
Let $X_1, X_2 \dots X_n$ be independent and identically distributed random variables. The **sum** of these random variables approaches a normal as $n \to \infty$:

$$\sum_{i=1}^{n} X_i \sim N(n \cdot \mu, n \cdot \sigma^2)$$

Where $\mu = \mathrm{E}[X_i]$ and $\sigma^2 = \mathrm{Var}(X_i)$. Note that since each X_i is identically distributed they share the same expectation and variance.

At this point you probably think that the central limit theorem is awesome. But it gets even better. With some algebraic manipulation we can show that if the sample mean of IID random variables is normal, it follows that the sum of equally weighted IID random variables must also be normal:

The Central Limit Thorem (Average Version)
Let $X_1, X_2 \dots X_n$ be independent and identically distributed random variables. The **average** of these random variables approaches a normal as $n \to \infty$:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$$

Where $\mu = \mathrm{E}[X_i]$ and $\sigma^2 = \mathrm{Var}(X_i)$.

Central Limit Theorem Intuition

In the previous section we explored what happens when you add two random variables. What happens when you add more than two random variables? For example, what if I wanted to add up 100 different uniform random variables:

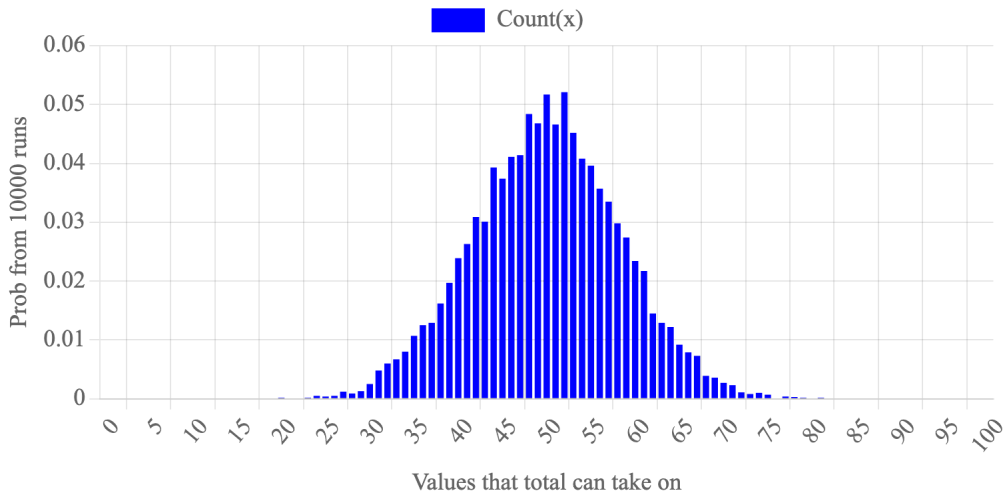
```
from random import random

def add_100_uniforms():
   total = 0
   for i in range(100):
       # returns a sample from uniform(0, 1)
       x_i = random()
       total += x_i
   return total
```

The value, `total` returned by this function will be a random variable. Hit the button below to run the function and observe the resulting value of total:

`add_100_uniforms()` `total: 51.29972`

What does total look like as a distribution? Let's calculate `total` many times and visualize the histogram of values it produces.



10,000 more runs

That is interesting! `total` which is the sum of 100 independent uniforms looks normal. Is that a special property of uniforms? No! It turns out to work for almost any type of distribution (as long as the thing you are adding has finite mean and finite variance, everything we have covered in this reader).

- Sum of 40 X_i where $X_i \sim \text{Beta}(a = 5, b = 4)$? Normal.
- Sum of 90 X_i where $X_i \sim \text{Poi}(\lambda = 4)$? Normal.
- Sum of 50 dice-rolls? Normal.
- Average of 10000 X_i where $X_i \sim \text{Exp}(\lambda = 8)$? Normal.

For any distribution the sum, or average, of n independent equally-weighted samples from that distribution, will be normal.

Continuity Correction

Now we can see that the Binomial Approximation using a Normal actually derives from the central limit theorem. Recall that, when computing probabilities for a normal approximation, we had to to use a continuity correction. This was because we were approximating a discrete random variable (a binomial) with a continuous one (a normal). You should use a continuity correction any time your normal is approximating a discrete random variable. The rules for a general continuity correction are the same as the rules for the binomial-approximation continuity correction.

In the motivating example above, where we added 100 uniforms, a continuity correction isn't needed because the sum of uniforms is continuous. In the dice sum example below, a continuity correction is needed because die outcomes are discrete.

Examples

Example:

You will roll a 6 sided dice 10 times. Let X be the total value of all 10 dice = $X_1 + X_2 + \cdots + X_{10}$. You win the game if $X \leq 25$ or $X \geq 45$. Use the central limit theorem to calculate the probability that you win. Recall that $E[X_i] = 3.5$ and $\text{Var}(X_i) = \frac{35}{12}$.

Let Y be the approximating normal. By the Central Limit Theorem $Y \sim N(10 \cdot E[X_i], 10 \cdot \text{Var}(X_i))$. Substituting in the known values for expectation and variance: $Y \sim N(35, 29.2)$

$$
\begin{aligned}
&\mathrm{P}(X \leq 25 \text{ or } X \geq 45) \\
&\quad = \mathrm{P}(X \leq 25) + \mathrm{P}(X \geq 45) \\
&\quad \approx \mathrm{P}(Y < 25.5) + \mathrm{P}(Y > 44.5) && \text{Continuity Correction} \\
&\quad \approx \mathrm{P}(Y < 25.5) + [1 - \mathrm{P}(Y < 44.5)] \\
&\quad \approx \Phi(\frac{25.5 - 35}{\sqrt{29.2}}) + \left[1 - \Phi(\frac{44.5 - 35}{\sqrt{29.2}})\right] && \text{Normal CDF} \\
&\quad \approx \Phi(-1.76) + [1 - \Phi(1.76)] \\
&\quad \approx 0.039 + (1 - 0.961) \approx 0.078
\end{aligned}
$$

Example:

Say you have a new algorithm and you want to test its running time. You have an idea of the variance of the algorithm's run time: $\sigma^2 = 4\text{sec}^2$ but you want to estimate the mean: $\mu = t\text{sec}$. You can run the algorithm repeatedly (IID trials). How many trials do you have to run so that your estimated runtime = $t \pm 0.5$ with 95\% certainty? LetX_i be the run time of the i-th run (for $1 \leq i \leq n$).

$$0.95 = P(-0.5 \leq \frac{\sum_{i=1}^{n} X_i}{n} - t \leq 0.5)$$

By the central limit theorem, the standard normal Z must be equal to:

$$
\begin{aligned}
Z &= \frac{\left(\sum_{i=1}^{n} X_i\right) - n\mu}{\sigma\sqrt{n}} \\
&= \frac{\left(\sum_{i=1}^{n} X_i\right) - nt}{2\sqrt{n}}
\end{aligned}
$$

Now we rewrite our probability inequality so that the central term is Z:

$$
\begin{aligned}
0.95 &= P(-0.5 \leq \frac{\sum_{i=1}^{n} X_i}{n} - t \leq 0.5) = P(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sum_{i=1}^{n} X_i}{n} - t \leq \frac{0.5\sqrt{n}}{2}) \\
&= P(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sqrt{n}}{2}\frac{\sum_{i=1}^{n} X_i}{n} - \frac{\sqrt{n}}{2}t \leq \frac{0.5\sqrt{n}}{2}) = P(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sum_{i=1}^{n} X_i}{2\sqrt{n}} - \frac{\sqrt{n}}{\sqrt{n}}\frac{\sqrt{n}t}{2} \leq \\
&= P(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sum_{i=1}^{n} X_i - nt}{2\sqrt{n}} \leq \frac{0.5\sqrt{n}}{2}) \\
&= P(\frac{-0.5\sqrt{n}}{2} \leq Z \leq \frac{0.5\sqrt{n}}{2})
\end{aligned}
$$

And now we can find the value of n that makes this equation hold.

$$
\begin{aligned}
0.95 &= \phi(\frac{\sqrt{n}}{4}) - \phi(-\frac{\sqrt{n}}{4}) = \phi(\frac{\sqrt{n}}{4}) - (1 - \phi(\frac{\sqrt{n}}{4})) \\
&= 2\phi(\frac{\sqrt{n}}{4}) - 1 \\
0.975 &= \phi(\frac{\sqrt{n}}{4}) \\
\phi^{-1}(0.975) &= \frac{\sqrt{n}}{4} \\
1.96 &= \frac{\sqrt{n}}{4} \\
n &= 61.4
\end{aligned}
$$

Thus it takes 62 runs. If you are interested in how this extends to cases where the variance is unknown, look into variations of the students' t-test.