**Th2: The CLT Meaning, Proof, Simulations**

Let's approach the concept of the central limit theorem and sampling distributions from a different angle:

Imagine you're conducting an experiment where you randomly select samples from a population and calculate a statistic, like the mean, for each sample. You repeat this process numerous times, creating a collection of means from all the samples (a sampling distribution).

The central limit theorem comes into play here, asserting that if your sample size is sufficiently large, the distribution of these sample means will always resemble a normal distribution. This holds true regardless of the original distribution of the population (whether it's normal, Poisson, binomial, or something else entirely).

A normal distribution is characterized by its symmetrical, bell-shaped curve. As you move away from the center of the distribution, there are fewer observations. The central limit theorem essentially ensures that, with a large enough sample size, the distribution of sample means will exhibit this characteristic normal shape, providing a powerful tool for statistical analysis.

As you increase the sample size, the sampling distribution gradually approaches a normal distribution, aligning its mean with the population mean and its standard deviation with σ/√n, where σ represents the population standard deviation and n is the sample size.

The effect of enlarging the sample size is notable in the standard deviation of the sampling distribution, which diminishes with the square root of the sample size in the denominator. This implies that, as the sample size grows, the sampling distribution becomes more tightly clustered around the mean. In simpler terms, larger sample sizes lead to a more precise and concentrated sampling distribution.

The central limit theorem is vital in statistics for two main reasons—the normality assumption and the precision of the estimates.

**The Central Limit Theorem**

The theorem states that if random variables X₁, …, Xₙ are independent and identically distributed, with a constant fixed mean μ and constant finite variance σ², then the random variable Z approaches the standard normal N(0,1), distribution.

where Z is a random variable that approaches the standard normal distribution

Note that we define:

The proof

The proof is as follows: Let us define a random variable Yᵢ for i = 1, …, n, that is independent and identically distributed such that,

Then, the expected value and variance of each Yᵢ is .

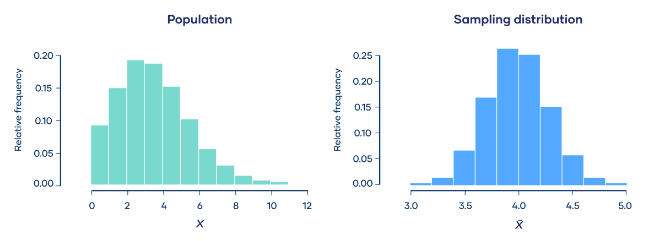
Let us define another random variable, S = Y₁+ … + Yₙ, to be the sum of all Yᵢ’s. Then, the expected value and variance of S is

Let us define our last random variable, Z:

It can be shown that this definition of Z is the same as the Z defined in the Central Limit Theorem.

Example:

A population follows a Poisson distribution (left image). If we take 10,000 samples from the population, each with a sample size of 50, the sample means follow a normal distribution, as predicted by the central limit theorem (right image).



Below you can find the different links from which I took the various information for writing the topic:

<https://statisticsbyjim.com/>

<https://www.scribbr.com/>

<https://medium.com/>