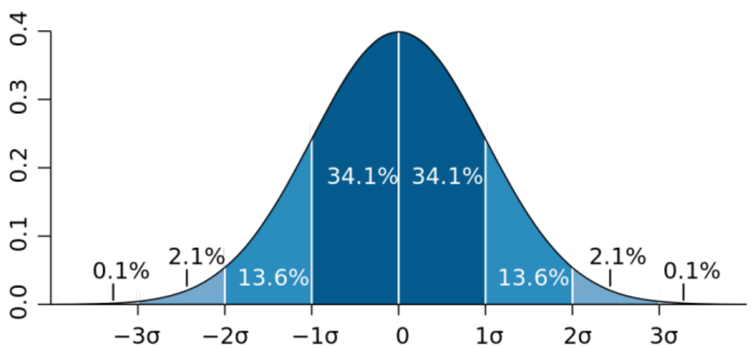
Th2: The Central Limit Theorem

Federico Detomaso

1 Meaning

In probability theory, the central limit theorem (CLT) states that the distribution of a sample variable approximates a normal distribution as the sample size becomes larger, assuming that all samples are identical in size, and regardless of the population's actual distribution shape.

Put another way, CLT is a statistical premise that, given a sufficiently large sample size from a population with a finite level of variance, the mean of all sampled variables from the same population will be approximately equal to the mean of the whole population. Furthermore, these samples approximate a normal distribution, with their variances being approximately equal to the variance of the population as the sample size gets larger, according to the law of large numbers.



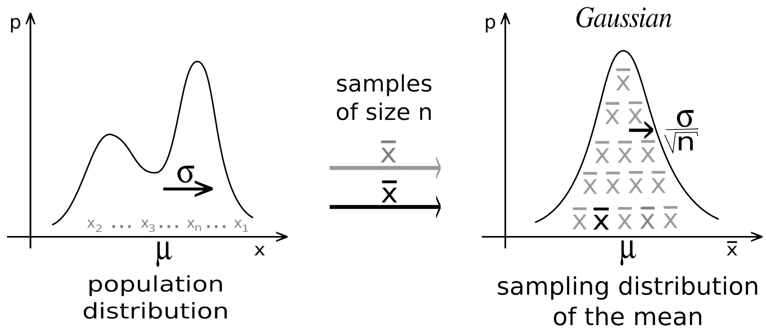
2 Definition

Let $X_1, ..., X_n$ denote a random sample of n independent observations from a population with overall expected value (average) μ and finite variance σ^2 , and let $\overline{X_n}$ denote the sample mean of that sample (which is itself a random variable). Then the limit as $n \longrightarrow \infty$ of the distribution of $\frac{\overline{X_n}-\mu}{\sigma_{\overline{X_n}}}$, where $\sigma_{\overline{X_n}=\frac{\sigma}{\sqrt{n}}}$, is the standard normal distribution.

In other words, suppose that a large sample of observations is obtained, each observation being randomly produced in a way that does not depend on the values of the other observations, and that the average (arithmetic mean) of the observed values is computed. If this procedure is performed many times, resulting in a collection of observed averages, the central limit theorem says that if the sample size was large enough, the probability distribution of these averages will closely approximate a normal distribution.

The central limit theorem has several variants. In its common form, the random variables must be independent and identically distributed (i.i.d.). This requirement can be weakened; convergence of the mean to the normal distribution also occurs for non-identical distributions or for non-independent observations if they comply with certain conditions.

The earliest version of this theorem, that the normal distribution may be used as an approximation to the binomial distribution, is the de Moivre–Laplace theorem.



3 Proof

The proof involves the use of characteristic functions. The characteristic function of a random variable X is defined as:

$$\phi_X(t) = E[e^{itX}] \tag{1}$$

where i is the imaginary unit.

Consider the characteristic function of S_n, the sum of n i.i.d. random variables:

$$\phi_{S_n}(t) = E[e^{itS_n}] = E[e^{it(X_1+X_2+\ldots+X_n)}] \tag{2}$$

Since the random variables are independent, the product of characteristic functions is the product of their individual characteristic functions:

$$\phi_{S_n}(t) = \prod_{i=1}^{n} \phi_{X_i}(t) \tag{3}$$

Continuing with the normalization:

$$\phi_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) = \left[\phi_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n \tag{4}$$

Now, let's express the characteristic function in terms of the moment-generating function (MGF) using the Taylor series expansion:

$$\phi_X(t) = E[e^{itX}] = E\left[\sum_{k=0}^{\infty} \frac{(itX)^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{i^k t^k}{k!} E[X^k] \tag{5}$$

Now, consider the MGF of the normalized sum:

$$M_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) = \left[M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n \tag{6}$$

where $M_X(t)$ is the moment-generating function of a single random variable.

Next, apply the conditions of the CLT. Lindeberg's condition is a common criterion:

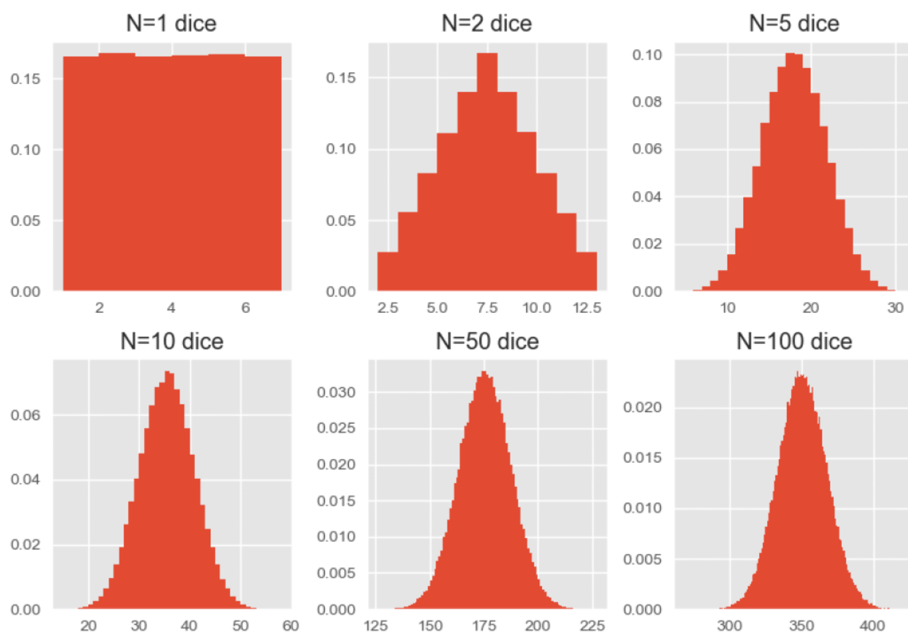
$$\lim_{n\to\infty} \frac{1}{\sigma^2 n} \sum_{i=1}^{n} E\left[(X_i - \mu)^2 \mathbf{1}_{\{|X_i-\mu|>\epsilon\sqrt{n}\}}\right] = 0 \tag{7}$$

This condition ensures that the variance dominates, and no single term in the sum has too much influence.

With Lindeberg's condition satisfied, you can use Levy's Continuity Theorem to conclude that the limiting distribution of the normalized sum is a standard normal distribution.

This completes the proof of the Central Limit Theorem.

4 Simulations



Demonstrating central limit theorem using N numbers of dice (x: sum of the dice values; y: probability).

References

[1] https://en.wikipedia.org/wiki/Central_limit_theorem

[2] https://www.cs.toronto.edu/~yuvalf/CLT.pdf

[3] https://www.gaussianwaves.com/2010/01/central-limit-theorem-2/