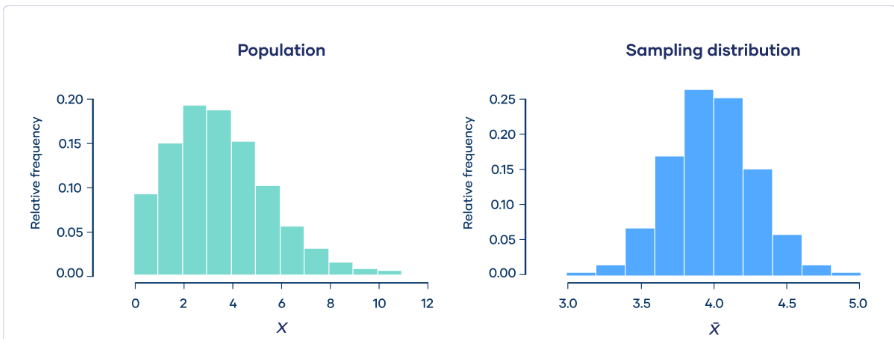
Central Limit Theorem CLT

In probability theory, the central limit theorem (CLT) establishes that, in many situations, for independent and identically distributed random variables, the sampling distribution of the standardized sample mean tends towards the standard normal distribution even if the original variables themselves are not normally distributed.



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).

If we draw a random sample from a population and calculate a statistic for the sample, such as the mean, then we repeat this process many times, and end up with a large number of means, one for each sample, the distribution of the sample means is an example of a sampling distribution.

The central limit theorem says that the sampling distribution of the mean will always be normally distributed, as long as the sample size is large enough. Regardless of whether the population has a normal, Poisson, binomial, or any other distribution, the sampling distribution of the mean will be normal.

A normal distribution is an asymmetrical, bell-shaped distribution, with increasingly fewer observations the further from the center of the distribution.

The parameters of the sampling distribution of the mean are determined by the parameters of the population:

- The mean of the sampling distribution is the mean of the population.
  $$\mu_{\bar{x}} = \mu$$
- The standard deviation of the sampling distribution is the standard deviation of the population divided by the square root of the sample size.
  $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$
- We can describe the sampling distribution of the mean using this notation:

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Where:

- $\bar{X}$ is the sampling distribution of the sample means
- ~ means "follows the distribution"
- N is the normal distribution
- μ is the mean of the population
- σ is the standard deviation of the population
- n is the sample size, it is the number of observations drawn from the population for each sample

The sample size affects the sampling distribution of the mean in two ways.

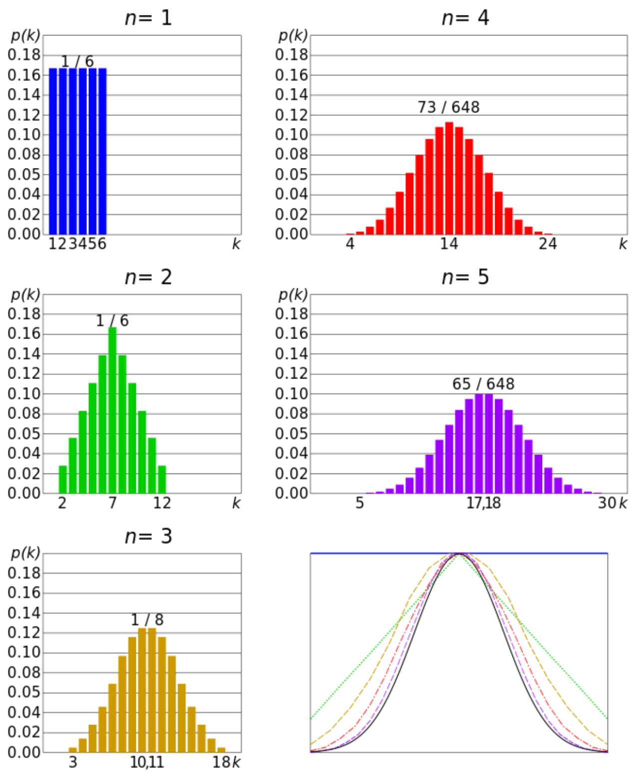
- The larger the sample size, the more closely the sampling distribution will follow a normal distribution.
  When the sample size is small, the sampling distribution of the mean is sometimes uniform.
- It affects the standard deviation of the sampling distribution. Standard deviation is a measure of the variability or spread of the distribution.

The conditions of validity of the central limit theorem are:

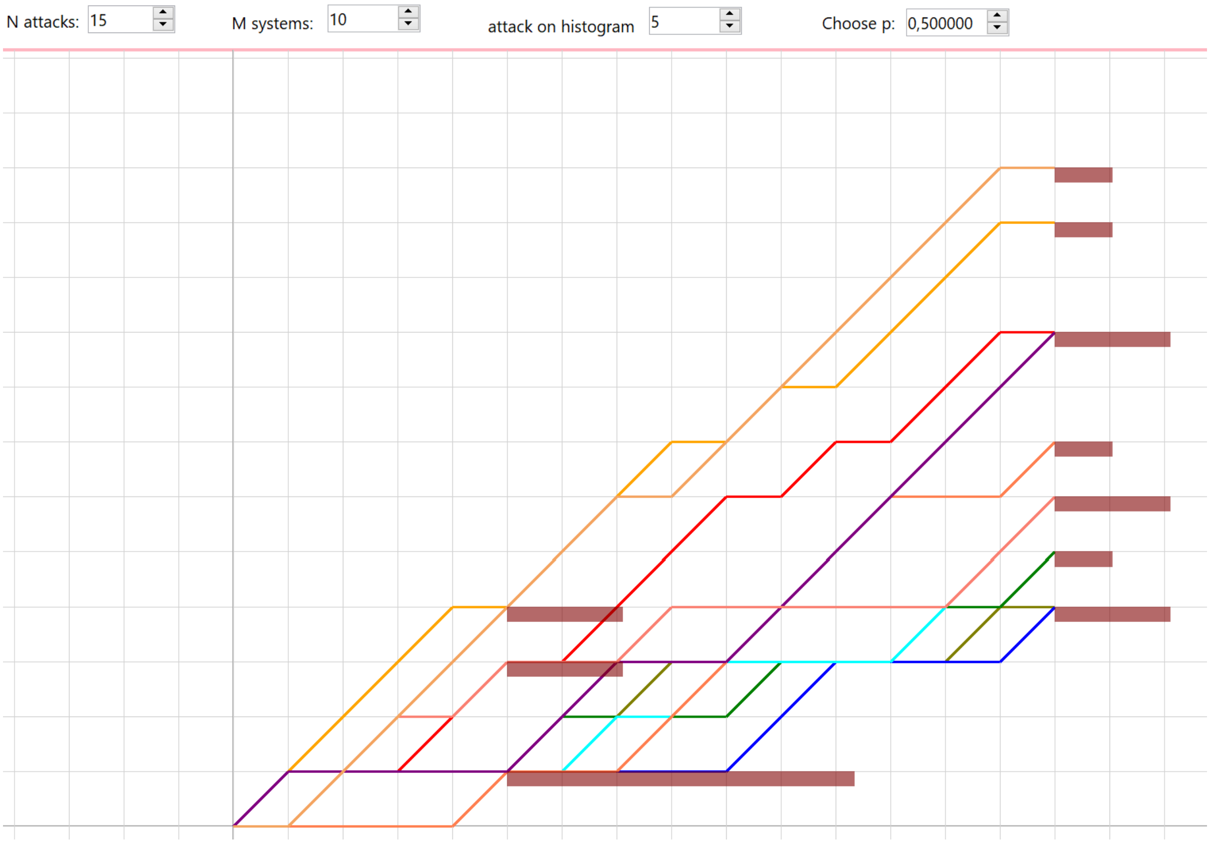
- The sample size is sufficiently large.
- The samples are independent and identically distributed (i.i.d.) random variables.
- The population's distribution has finite variance. Central limit theorem doesn't apply to distributions with infinite variance, such as the Cauchy distribution.

The formal proof of the central limit theorem can be found on this [page].

The below graph shows the comparison of probability density functions p(k) for the sum of n fair 6-sided dice to show their convergence to a normal distribution with increasing n, in accordance to the central limit theorem.



In homework 3 if we choose a small number of systems, we can see that the distribution is uniform, like shown in the following graph.



Instead if we increase the number of systems we can see that the distribution is similar to the normal distribution, and we can see that in the following graph.

This simulation is related to the central limit theorem because, like we can see in the second graph, the normal distribution starts to appear when the three properties written above are valid, in particular:

- The sample size is sufficiently large. In the first graph this property is violated because the number of systems is 10 (too low), instead in the second graph the number of systems is 50 that is sufficiently high to show a normal distribution
- The samples are independent and identically distributed random variables.
- The population's distribution has finite variance.

Bibliography

-https://en.wikipedia.org/wiki/Central_limit_theorem
-https://www.scribbr.com/statistics/central-limit-theorem/#:~:text=The%20central%20limit%20theorem%20states,population%20isn%27t%20normally%20distributed.