Search on the web about the Law of large numbers LLN and compare it with Part b of your homework 3 and express in your own words whether your simulation is somehow related with this theorem, and why.

The LLN states that as the size of a sample drawn from a population increases, the sample mean approaches the population mean. In simpler terms, the more observations you have, the closer your average will get to the ‘true’ average of the entire population.

For example, imagine you flip a coin. It can land either heads or tails. Now, if you flip it just once or twice, you might get heads both times, even though we expect heads and tails to be equally likely. So, from just a couple of flips, you might think that getting heads is more common.

But, if you keep flipping the coin, say 1000 times, you’ll likely find that the number of heads and tails becomes closer to even – roughly 500 heads and 500 tails.

The Law of Large Numbers basically says that the more times you repeat a random experiment (like flipping a coin), the closer the average outcome (like the percentage of heads) will get to the expected value (50% heads and 50% tails, in this case).

In simpler words: In the short run, randomness can seem unpredictable and chaotic, but given enough repetitions, things tend to average out to what we expect

The Law of Large Numbers is relevant in the Part b of Homework 3, since it helps explain the behavior of the statistics as the number of attacks (sample size) becomes very large. In particular:

* As the number of attacks (N) becomes larger, the cumulated frequency (f) is likely to stabilize around some value. This is consistent with the LLN, as it suggests that the true proportion of successful penetrations will become more predictable as the sample size (number of attacks) increases.
* The relative frequency (f/number of attacks) will tend to converge to the true probability of penetration as N becomes large. This convergence is in line with the LLN, which tells us that the sample proportion approaches the population proportion.
* The normalized ratio (f/√number of attacks) is a more intricate measure but still reflects the behavior of the system as N grows. This particular normalization with the square root of the number of attacks accounts for variability and could provide a more stable estimate of the penetration probability as N increases.

[Law of Large Numbers - A Deep Dive into the World of Statistics - Machine Learning Plus](https://www.machinelearningplus.com/statistics/law-of-large-numbers/)

Search on the web about the **Central Limit Theorem CLT** and compare it with Part a of your homework 3 and say in your own words whether your simulation is somehow related with this theorem, and why.

The central limit theorem states that if you take sufficiently large samples from a population, the samples’ means will be normally distributed, even if the population isn’t normally distributed.

The central limit theorem relies on the concept of a sampling distribution, which is the probability distribution of a statistic for a large number of samples taken from a population.

For example:

* Suppose that you draw a random sample from a population and calculate a statistic for the sample, such as the mean.
* Now you draw another random sample of the same size, and again calculate the mean.
* You repeat this process many times, and end up with a large number of means, one for each sample.

So, 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, where a normal distribution is a symmetrical, bell-shaped distribution, with increasingly fewer observations the further from the center of the distribution.

The CLT is relevant in the Part a of Homework 3 for two scenarios:

* For a Single System: The "security score" for a single system at each attack can be considered as a sequence of independent, identically distributed random variables. These variables take values of -1 or 1 with probabilities related to the penetration probability p. If we consider the sum of these "security scores" over a large number of attacks (N) for a single system, the CLT implies that this sum (or average) will tend to follow a normal distribution, assuming N is sufficiently large. The mean and variance of this distribution would depend on the penetration probability p.
* For Multiple Systems (M Systems): If we are simulating multiple systems (M systems) in parallel, each with its own sequence of "security scores," then we can think of the "security scores" for each system as independent and identically distributed random variables. As we collect data from multiple systems, the CLT implies that the distribution of the average "security score" across all systems will tend to be approximately normally distributed, again assuming that N (the number of attacks) is sufficiently large.

[Central Limit Theorem | Formula, Definition & Examples (scribbr.com)](https://www.scribbr.com/statistics/central-limit-theorem/)

Based on the CLT, how could you modify ("normalize") the "security score" to obtain an asymptotic convergence to a proper distribution?

To obtain an asymptotic convergence to a proper distribution as per the Central Limit Theorem (CLT), we can modify the code of Part b of Homework 3 in this way:

1. Define the Random Variables: Let Xij be the "security score" for the j-th attack of the i-th system, where Xij takes values of -1 or 1 based on whether the system is penetrated or protected. These random variables are assumed to be independent and identically distributed with a mean of μ (E[Xij] = μ) and a variance of σ2 (Var(Xij) = σ2)
2. Calculate the Sample Means and Normalization: Calculate the sample means of the "security scores" for each system over N attacks. Also, normalize these sample means to z-scores. We should keep track of these sample means and z-scores for each system.
3. Aggregate and Analyze Results: After obtaining the sample means and z-scores for each system, we can analyze and visualize the distribution of the z-scores. If M is reasonably large, the distribution of these z-scores should exhibit properties similar to a standard normal distribution, as suggested by the CLT.

It's important to note that the sample size (N) and the number of systems (M) should be sufficiently large for the CLT to hold. The z-scores should approach a standard normal distribution when these conditions are met. Also, the penetration probability p and the distribution of the "security scores" play a role in determining the properties of the z-scores.