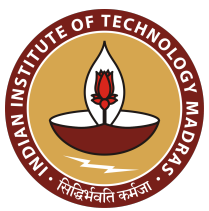
CLT: as a statement on estimator's distribution

Saurav Samantaray

Department of Mathematics

Indian Institute of Technology Madras

October 15, 2026



Fine grain data and their averages

- Suppose we collect observations from a population whose distribution is strongly skewed.
- For example, the population may represent waiting times, income, or the number of successes in repeated trials.
- We collect a sample of size $n = 30$ and calculate the sample mean.
- two distinct questions we can ask:
  1. What does the distribution of the individual observations look like?
  2. What does the distribution of the sample mean look like if we repeat the sampling experiment many times?
- These questions are not equivalent.
- The observations may have a strongly skewed distribution even when the sample mean has an approximately symmetric distribution.

Fine grain data and their averages

Consider the following experiment.

- We collect 30 observations and obtain a sample mean of 2.4.
- We repeat the experiment with a new sample of 30 observations and obtain 1.9.
- A third sample gives 2.1.

Each experiment produces a different sample mean because the observations are random.

- If we could repeat this experiment thousands of times, we would obtain thousands of sample means.
- We could then examine the distribution of these means.

Question

Does the distribution of these sample means have to resemble the distribution of the original observations?

$->$ Let's investigate this question experimentally.

One sample, an estimate; many samples, a distribution

- Let $X_1, X_2, \ldots, X_n$ denote independent observations from a population with mean μ and variance σ^2.
- The sample mean is:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- For a single observed dataset, $\overline{X}_n$ is one numerical value.
- Before observing the data, however, it is a random variable because it depends on the random observations.
- How do we study its distribution computationally?
  By performing the following experiment
  1. Generate n observations from the population.
  2. Calculate their sample mean.
  3. Generate another independent sample of size n.
  4. Calculate its sample mean.
  5. Repeat the procedure R times.

One sample, an estimate; many samples, a distribution

- This gives us

$$\overline{X}_n^{(1)}, \overline{X}_n^{(2)}, \ldots, \overline{X}_n^{(R)}.$$

- A histogram of these values approximates the sampling distribution of the sample mean.
- The key distinction is between the distribution of the original observations and the distribution of the estimator calculated from those observations.

Question

If we increase R, are we increasing the size of each dataset, or are we improving our approximation to the sampling distribution?

- Here, R controls how accurately we see the sampling distribution;
- n controls how much data each individual estimate uses.

The Central Limit Theorem and what does it say?

CLT

Suppose $X_1, \ldots, X_n$ are independent and identically distributed random variables with finite mean μ and finite, non-zero variance σ^2. Then

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1) \qquad \text{as } n \to \infty.$$

Equivalently, for sufficiently large n, the sample mean is often approximated by

$$\overline{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right).$$

This gives us two important pieces of information.

1. The sample mean is centered around the population mean μ.
2. Its variance decreases as the sample size increases.

The Central Limit Theorem and what does it say?

The standard deviation of the sample mean is

$$\mathrm{SE}(\overline{X}_n) = \frac{\sigma}{\sqrt{n}}.$$

- However, the CLT does not tell that the original observations become normally distributed.
- Their population distribution remains unchanged regardless of the sample size.
- It also does not specify a universal sample size beyond which the approximation is accurate.
- The quality of the approximation depends on the population distribution and on what we need the approximation to achieve.

Question

If the observations come from a highly skewed population, why might their average still have an approximately normal distribution?

Computational Exercise: separating n from R

Let's translate the mathematical statement into a computational experiment.

- There are two independent quantities to control:
  1. n: the number of observations in each sample.
  2. R: the number of independent samples generated.
- For a fixed sample size n, we generate R sample means.
- We then repeat the experiment with larger sample sizes while keeping R sufficiently large.
- For example, consider

$$n \in \{5, 10, 30, 100, 500\}, \qquad R = 10\,000.$$

- For every sample size, we generate a histogram of the sample means.
- For populations with finite mean and variance, we can also standardize the means:

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}.$$

Computational Exercise: separating n from R

- If the CLT approximation is accurate, the distribution of Z_n should resemble the standard normal distribution $N(0, 1)$.
- Standardization is useful because it puts different populations and sample sizes on a common scale.

What we will examine:

1. Does the histogram become more symmetric as n increases?
2. Does it become closer to the standard normal distribution?
3. Do all populations approach normality at the same rate?
4. Are there populations for which the usual CLT prediction does not apply?

Problem-1: Bernoulli(0.05) and rare events

Consider independent Bernoulli observations with success probability $p = 0.05$:

$$X_i = \begin{cases} 1, & \text{with probability } 0.05, \\ 0, & \text{with probability } 0.95. \end{cases}$$

The population mean and variance are

$$\mu = p = 0.05, \qquad \sigma^2 = p(1-p) = 0.0475.$$

The sample mean is the observed proportion of successes:

$$\overline{X}_n = \frac{\text{number of successes}}{n}.$$

- When n is small, the number of successes can take only a few possible values.
- For example, with $n = 10$, the sample mean can only be one of

$$0,\ 0.1,\ 0.2, \ldots, 1.$$

Problem-1: Bernoulli(0.05) and rare events

- Most samples will contain zero, one, or two successes.
- The sampling distribution is therefore discrete and can be far from bell-shaped.
- As n increases, the number of possible values increases, and the expected number of successes becomes

$$\mathbb{E}[\text{number of successes}] = np.$$

- For $p = 0.05$, this expectation is only 0.5 when $n = 10$, but it is 25 when $n = 500$.