

The Central Limit Theorem: Why the Sample Mean Becomes Normal and Narrows as the Sample Grows

Shaurya R. Chaturvedi¹, Ila K. Bajpai², Nakul V. Upadhyay³, Vandana M. Nigam⁴

ABSTRACT

When a quantity is estimated by averaging a sample, the average itself is uncertain, varying from one sample to another, so it has a distribution of its own; the central limit theorem says that this distribution of the sample mean becomes a normal, bell-shaped one as the sample grows, almost whatever the distribution of the individual values, and narrows as the sample grows, which is why the average of a large sample is both predictable in shape and close to the true value. This study examines the central limit theorem and why the sample mean becomes normal and narrows as the sample grows. The distribution of the sample mean was examined as the size of the sample was increased, for samples drawn from a markedly non-normal population, so that the change in the shape and the spread of that distribution with sample size could be seen. As the sample grew, the distribution of the sample mean changed in two ways: it became more nearly normal in shape, losing the skewness of the population from which the samples were drawn, so that even averages of a strongly non-normal quantity were bell-shaped once the sample was moderately large; and it became narrower, its spread shrinking as the sample grew, in proportion to one over the square root of the sample size, so that the sample mean clustered ever more tightly about the true population mean. The distribution of the sample mean for a large sample was therefore a narrow normal curve centred on the true mean, regardless of the shape of the population. This is why so much of statistics rests on the normal distribution and why larger samples give more reliable estimates, and it underlies confidence intervals and tests built on the sample mean. The study shows that the sample mean becomes normal and narrows as the square root of the sample size, and it discusses the implications for statistical estimation and inference.

KEYWORDS: *Central limit theorem, Sample mean, Sampling distribution, Normal distribution, Standard error, Sample size, Statistical inference, Probability, Estimation*

INTRODUCTION

A great deal of statistics consists of estimating some quantity about a population, its average height, its mean lifetime, its typical value, by drawing a sample and averaging the values in it. The sample mean is then used as an estimate of the true population mean. But the sample is only a part of the population, chosen by chance, so the sample mean is itself a matter of chance: a different sample would give a somewhat different mean, and the sample mean therefore varies from sample to sample (5, 6).

Because it varies, the sample mean has a distribution of its own, the sampling distribution of the mean, describing how the sample mean is spread over the values it might take. Knowing this distribution is essential to using the sample mean sensibly: it tells us how far the sample mean is likely to stray from the true mean, and so how much to trust an estimate from a single sample. The question is what this distribution looks like, and how it depends on the size of the sample (3, 4).

The central limit theorem answers this remarkably generally. It says that, as the sample grows, the distribution of the sample mean becomes normal, the familiar bell shape, almost whatever the distribution of the individual values, and that it narrows, clustering ever more tightly about the true mean. This paper examines the central limit theorem and why the sample mean becomes normal and narrows as the sample grows, examining the distribution of the sample mean as the sample size is increased for a non-normal population. The aim is to show the change in shape and spread with sample size, and to discuss the implications for statistical estimation and inference.

LITERATURE REVIEW

The central limit theorem is among the foundational results of probability, stating that the sum or average of many independent random quantities tends to a normal distribution as their number grows, almost regardless of the distribution of the individual quantities, a result developed and proved in the probability literature (1, 2, 10). Its consequence for statistics, that

the sampling distribution of the mean becomes normal, is set out in all the texts on statistical inference (3, 4).

THE SAMPLING DISTRIBUTION OF THE MEAN

The literature describes the distribution of the sample mean:

- The sample mean varies from sample to sample, with its own sampling distribution (3, 4).
- Its spread, the standard error, is the population spread divided by the square root of the sample size (5, 6).
- By the central limit theorem, its shape becomes normal as the sample grows (1, 2).
- This holds almost regardless of the population's own distribution (1, 7).

A recurring theme is that the sample mean becomes normal and narrows as the sample grows.

The literature establishes the two features of the sampling distribution of the mean: its spread, the standard error, equals the spread of the population divided by the square root of the sample size, so it shrinks as the sample grows; and its shape, by the central limit theorem, approaches the normal as the sample grows, however the population itself is distributed, provided the population has a finite spread (1, 5, 6, 7). It treats the rate of this approach and the conditions under which it holds, and the centrality of this result to estimation and to tests and intervals built on the sample mean, throughout the texts on probability and mathematical statistics (2, 3, 8, 9).

RESEARCH GAP

Although the central limit theorem is foundational, a concise demonstration that shows the distribution of the sample mean both becoming normal and narrowing as the sample grows, for a non-normal population, remains useful as a clear illustration. There is therefore value in a study that displays the sampling distribution at several sample sizes and relates it to estimation and inference.

OBJECTIVES

The objectives of this study are:

- To examine the distribution of the sample mean as the sample size is increased.
- To show how the distribution becomes more nearly normal as the sample grows.

- To show how its spread narrows in proportion to one over the square root of the sample size.
- To relate these to statistical estimation and inference.

METHODOLOGY

1. The Distribution of the Sample Mean

The distribution of the sample mean was examined as the size of the sample was increased, for samples drawn from a markedly non-normal population, so that the change in its shape and spread could be seen, the spread being the population spread divided by the square root of the sample size (5, 6). The elements are summarised in Table 1.

Table 1: Elements of the sampling distribution of the mean

Element	Description
Population	The whole set of values, here non-normal
Sample size	Number of values averaged
Sample mean	Average of a sample, an estimate of the mean
Standard error	Spread of the sample mean across samples

2. Increasing the Sample Size

The distribution of the sample mean was examined at several sample sizes, from small to large, so that the change toward a normal shape and the narrowing of its spread could be seen (1, 2). The narrowing in proportion to one over the square root of the sample size was related to the standard error (5).

RESULTS AND FINDINGS

As the sample grew, the distribution of the sample mean became more nearly normal in shape and narrower in spread, as shown in Figure 1, which displays the distribution of the sample mean for several sample sizes, drawn from a skewed population. The distribution lost the skewness of the population from which the samples were drawn, so that even averages of a strongly non-normal quantity were bell-shaped once the sample was moderately large (1, 2).

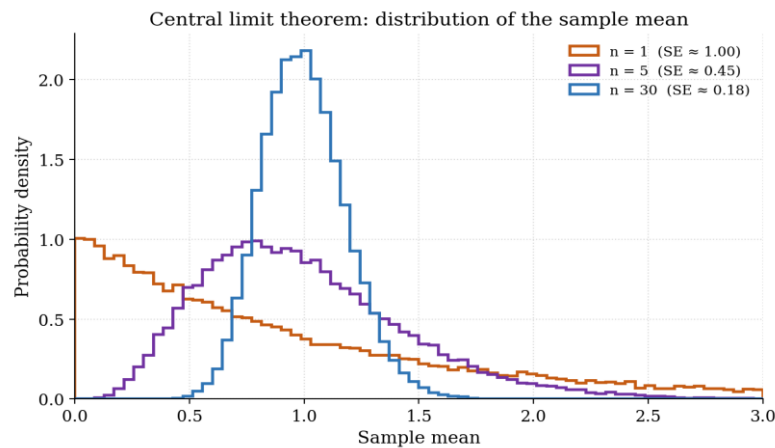


Figure 1: Central limit theorem: distribution of the sample mean

At the same time the distribution became narrower, its spread shrinking as the sample grew, in proportion to one over the square root of the sample size, so that the sample mean clustered ever more tightly about the true population mean (5, 6). The distribution of the sample mean for a large sample was therefore a narrow normal curve centred on the true mean, regardless of the shape of the population. The findings are summarised in Table 2.

Table 2: Distribution of the sample mean with sample size

Sample size	Shape	Spread
Small	Like the population (skewed)	Wide
Moderate	Becoming bell-shaped	Narrower
Large	Nearly normal	Narrow
As sample grows	Approaches normal	Shrinks as $1/\sqrt{n}$

DISCUSSION

The results show that the sample mean becomes normal and narrows as the square root of the sample size, and the two features have different but related explanations. The narrowing comes from averaging: when many values are averaged, their individual highs and lows tend to cancel, so the average varies less than a single value does, and the more values are averaged the more completely the fluctuations cancel; the spread of the average therefore shrinks, and it does so in proportion to one over the square root of the number averaged, the standard error. This is why a larger sample gives a sample mean that strays less from the truth (5, 6).

The approach to a normal shape is the deeper part, and it is the content of the central limit theorem. Whatever the distribution of the individual values, skewed, lumpy, or irregular, the distribution of their average smooths out toward the normal bell shape as more are averaged, because the sum of many independent contributions, no one of which dominates, tends to the normal form. So the sample mean of a large sample is approximately normal even when the underlying quantity is markedly non-normal, as the figure shows for a skewed population whose sample means nonetheless became bell-shaped; this is a remarkable universality, making the normal distribution appear wherever quantities are averaged or accumulated (1, 2, 7).

Together these explain why the normal distribution is so central to statistics and why larger samples are prized, which is the practical import. Because the sample mean is approximately normal with a known spread, the standard error, its behaviour is fully described by two numbers, the true mean and that spread, and the well-known properties of the normal distribution can be used to say how likely the sample mean is to fall within any distance of the truth; this is the basis of confidence intervals and of tests built on the sample mean, which assume the sample mean is normally distributed (3, 9). And because the spread narrows as the square root of the sample size, a larger sample gives a tighter estimate, so to halve the uncertainty of an estimate one must quadruple the sample, a rule that governs how large a sample is needed for a desired precision (5). The diminishing return, that ever larger samples are needed for further gains, follows from the same square-Root law. The practical conclusion is that the central limit theorem makes the sample mean approximately normal and ever more precise as the sample grows, underpinning estimation and inference, the main caveats being that it requires independent observations and a finite population spread and that the figure illustrates the result rather than a specific dataset.

CONCLUSION

The central limit theorem and why the sample mean becomes normal and narrows as the sample grows were examined by examining the distribution of the sample mean as the sample size increased, for a non-normal population. As the sample grew, the distribution of the sample mean became more nearly normal, losing the skewness of the population, and narrower, its spread shrinking as one over the square root of the sample size, so that the sample mean of a large sample was a narrow normal curve centred on the true mean whatever the population's

shape. The study shows that the sample mean becomes normal and narrows as the square root of the sample size, which is why the normal distribution is central to statistics and why larger samples give more reliable estimates, underpinning confidence intervals and tests built on the mean, with the caveats that it requires independent observations and a finite population spread and that the figure illustrates the result rather than a specific dataset.

LIMITATIONS

The following limitations apply to this study:

- The theorem requires independent observations and a population with a finite spread.
- The figure illustrates the result rather than the distribution from a specific dataset.
- For much skewed populations a larger sample is needed before the mean is nearly normal.
- The standard error uses the population spread, which is itself usually estimated from the sample.

FUTURE SCOPE

Future work could extend the present study as follows:

- How large a sample is needed for near-normality for various population shapes.
- The central limit theorem for sums and for statistics other than the mean.
- Behaviour when observations are dependent or the population spread is infinite.
- The role of the theorem in constructing confidence intervals and tests.
- Estimation of the standard error from the sample and its effect on inference.

REFERENCES

1. Feller, W. (1968). *An introduction to probability theory and its applications* (Vol. 1, 3rd ed.). John Wiley & Sons.
2. Ross, S. M. (2014). *A first course in probability* (9th ed.). Pearson.
3. Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Duxbury.
4. Rice, J. A. (2007). *Mathematical statistics and data analysis* (3rd ed.). Thomson Brooks/Cole.
5. Montgomery, D. C., & Runger, G. C. (2018). *Applied statistics and probability for engineers* (7th ed.). John Wiley & Sons.
6. Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2016). *Probability and statistics for engineers and scientists* (9th ed.). Pearson.

7. DeGroot, M. H., & Schervish, M. J. (2012). *Probability and statistics* (4th ed.). Addison-Wesley.
8. Hogg, R. V., McKean, J. W., & Craig, A. T. (2019). *Introduction to mathematical statistics* (8th ed.). Pearson.
9. Wasserman, L. (2004). *All of statistics: A concise course in statistical inference*. Springer.
10. Billingsley, P. (1995). *Probability and measure* (3rd ed.). John Wiley & Sons.

***Author for Correspondence**

Shaurya R. Chaturvedi

E-mail: shaurya.chaturvedi89@gmail.com

¹Associate Professor, Department of Mathematics, Saryu Plains Institute of Technology, Gorakhpur, Uttar Pradesh

²Assistant Professor, Department of Applied Mathematics, Saryu Plains Institute of Technology, Gorakhpur, Uttar Pradesh

³Research Scholar, Department of Statistics, Saryu Plains Institute of Technology, Gorakhpur, Uttar Pradesh

⁴Research Scholar, Department of Mathematics and Statistics, Bundelkhand Ridge College of Engineering, Jhansi, Uttar Pradesh

Received Date: May 19, 2026

Accepted Date: June 02, 2026

Published Date: June 10, 2026

Citation: Shaurya R. Chaturvedi, Ila K. Bajpai, Nakul V. Upadhyay, Vandana M. Nigam. The Central Limit Theorem: Why the Sample Mean Becomes Normal and Narrows as the Sample Grows. *Journal of Engineering Mathematics & Statistics*. 2026; 10(1): 1-8p.