IOCMA 2026 Conference

The 2nd International Online Conference on Mathematics and Applications

10–12 June 2026 | Online

MDPI

Validation of the Central Limit Theorem, Minimum Sample Size Estimation for Convergence to Normality, and a Closer Look at Post-Convergence Fluctuation

Shinjon Chakraborty

Department of Statistics, University of Calcutta, Kolkata, India

INTRODUCTION & AIM

The Central Limit Theorem (CLT) is among the most consequential results in mathematical statistics, underpinning much of the inferential machinery used in science and engineering. Formally, if $X_1, X_2, ..., X_n$ are i.i.d. random variables with mean μ and finite variance σ^2, the standardised sample mean $T_n = (\bar{X}_n - \mu) / (\sigma/\sqrt{n})$ converges in distribution to N(0,1) as $n \to \infty$.

The standard textbook rule of "$n \geq 30$" is widely taught but rarely examined carefully. It conflates distributions with very different shapes, treating a near-symmetric Uniform and a heavy-tailed Geometric as if they demanded the same sample size. This paper subjects that rule to systematic scrutiny across ten distributions, both discrete and continuous.

Research Questions

1. Threshold estimation: What is the smallest n at which T_n first passes a formal normality test? We call this the threshold sample size, n*.

2. Post-convergence fluctuation: Once n* is crossed, do all larger samples pass? Empirically, no. We define the optimum sample size, n**, as the last n at which a failure occurs.

3. Test selection: How do Shapiro-Wilk, Kolmogorov-Smirnov, and Anderson-Darling compare in power as functions of n and distributional skewness?

METHOD

Simulation Framework

For each distribution and each sample size n, M = 1000 rows of i.i.d. draws are generated in R. The standardised sample mean T_n⁻¢¡£ is computed from each row, and the Shapiro-Wilk test is applied to the resulting vector of 1000 values. Threshold n* is the first n yielding p-value above $\alpha = 0.05$. Optimum n** is the last $n \leq 5000$ yielding a failure.

Distributions Studied (10 total)

Discrete: Binomial(2, 0.3), Poisson(5), Geometric(0.4), Hypergeometric(6, 8, 4)

Continuous: Exponential(1), Chi-square(3), Gamma(2,3), Beta(2,3), Uniform(0,1), Cauchy(0,1)

Note: Cauchy has no finite moments; analysis uses the sample median with its known asymptotic normal distribution (asymptotic variance = $\pi^2\sigma^2/4$).

Normality Tests Compared

Shapiro-Wilk (SW) based on order-statistic regression; W ≈ 1 favours normality.

Kolmogorov-Smirnov (KS-Lilliefors) supremum of $|F_n(x) - F_0(x)|$; uniform sensitivity across support.

Anderson-Darling (AD) weighted squared deviation; heavier weight on tail discrepancies, which CLT convergence affects last.

Power study design: B = 2000 samples per (distribution, n) pair; $n \in \{10, 20, 30, 50, 75, 100, 150, 200\}$; all code run in R ≥ 4.0 with set.seed(100).

RESULTS & DISCUSSION

Table 1 : Summary of Central Limit Theorem Convergence and Post-Convergence Fluctuation. This table details the threshold sample sizes (n∗) required for initial normality and the optimum sample sizes (n) ensuring stable post-convergence behavior. Results highlight the impact of population skewness and excess kurtosis on Shapiro-Wilk test failure rates across various distributions (α=0.05, M=1000).

DISTRIBUTION	SKEWNESS	EX. KURTOSIS	THRESHOLD (N*)	OPTIMUM (N**)	FAILURE RATE
Uniform(2, 4)	0.000	-1.200	3	1391	3.28%
Beta(3, 4)	0.263	-0.418	5	814	0.40%
Gamma(5, 1)	0.894	1.200	32	3533	12.24%
Exponential(1)	2.000	6.000	65	3811	10.44%
Cauchy(0, 1) [median]	undef.	undef.	34	198	1.88%
Normal(0, 1)	0.000	0.000	1	4340	8.74%

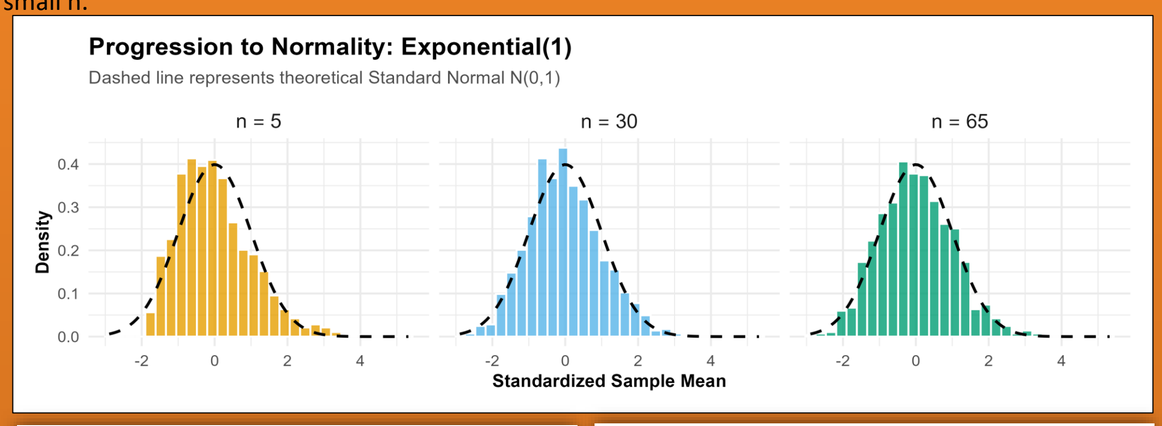
Power Comparison of Normality Tests

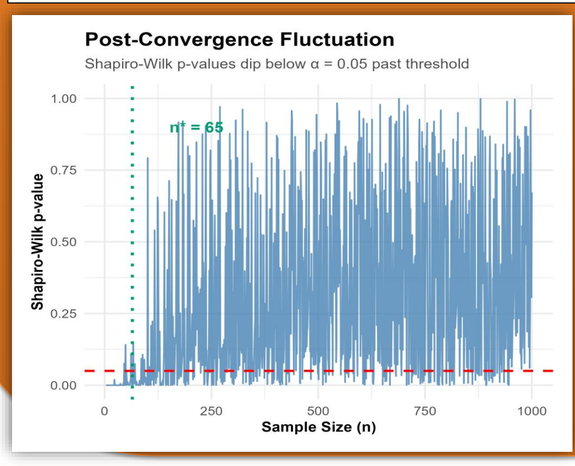
Empirical rejection rates (as functions of n) across Exponential, Gamma, Beta, and Uniform distributions reveal clear ordering at small to moderate sample sizes.

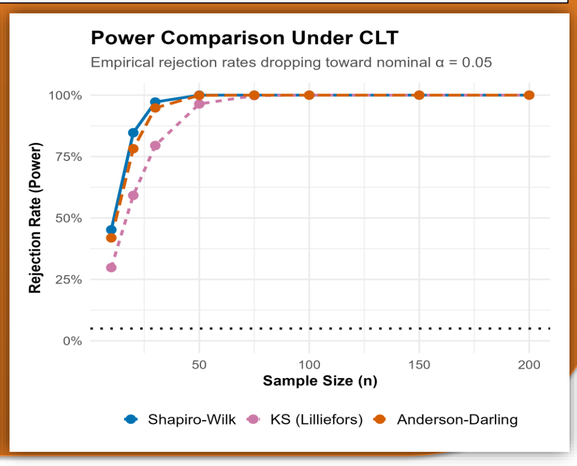
Shapiro-Wilk achieves the highest power for n < 200 against skewed alternatives. Therefore, it is the recommended choice for CLT validation work.

Anderson-Darling outperforms K-S uniformly due to its emphasis on tail deviations, which are precisely what CLT convergence affects last.

K-S (Lilliefors) has the lowest power of the three and is not recommended for CLT convergence studies. For almost symmetric distributions (Uniform, Beta), all tests show rejection rates near α = 0.05 even at small n.







CONCLUSION

The threshold n* ranges from 3 (Uniform) to 152 (Geometric), driven primarily by population skewness and excess kurtosis. The "$n \geq 30$" rule is both an over simplification and frequently wrong. Post-convergence fluctuation is a real finite sample phenomenon present in every distribution studied, including N(0,1) itself. The optimum n** can be orders of magnitude larger than n*, making it a more defensible practical recommendation. No outlier-based theory accounts for all cases of fluctuation. The phenomenon likely arises from the interaction between the Shapiro-Wilk statistic and finite sample near-normality. Shapiro-Wilk and Anderson-Darling clearly outperform KS-Lilliefors for CLT simulation studies.

FUTURE WORK / REFERENCES

Future work: (i) full parameter space sensitivity analysis; (ii) theoretical derivation of the Shapiro-Wilk statistic distribution under near-normal populations; (iii) extension to multivariate CLT and ARMA sequences.

Key references: Shapiro & Wilk (1965) Biometrika 52; Lilliefors (1967) JASA 62; Anderson & Darling (1954) JASA 49; Razali & Wah (2011) J.Stat.Mod.Anal. 2; Feller (1971) Probability Theory Vol.2; Billingsley (1995) Probability and Measure.