OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Determining The Relative Normality Convergence Speed Across Different Skewness and Tail Heaviness of Population Distribution

Jiyun Lim
jykim8605@gmail.com

ABSTRACT

This study investigates how the different distributions with varying characteristics and sample sizes converge to the normal distribution at different speeds. Using uniform, exponential, lognormal, and Pareto distributions, we find that the approximation error decreases as the sample size increases, which is consistent with the Central Limit Theorem. However, the rate of convergence differs depending on the shape of the distribution. Distributions with greater skewness and heavier tails tend to converge more slowly. The empirical analysis using IPUMS data shows similar patterns. As the sample size increases, the approximation error decreases, while the scaled error remains relatively stable. Overall, the results suggest that although normal convergence generally occurs, the speed of convergence varies depending on distributional characteristics and may not strictly follow the theoretical rate.

1. INTRODUCTION

Statistical inference is the process of using sample data to draw conclusions and estimate parameters for the population group, through methods such as hypothesis testing, confidence intervals, and estimation. Since analyzing an entire population is often impractical or impossible, statistical inference allows findings from a sample to be generalized to the larger group. This approach is widely used in many fields, including clinical trials, marketing, and finance.

The Central Limit Theorem (CLT) provides a theoretical foundation for these inferences. It states that the distribution of the sample mean tends to approach a normal distribution regardless of the original population's shape, as long as the random variables are independently and identically distributed and the sample size is sufficiently large. In practice, a sample size greater than 30 is often considered adequate. However, since CLT is an asymptotic result, the approximation improves as the sample size increases toward infinity.

Despite its importance, CLT has limitations in real-world situations. Many real datasets typically have consequences of inferential testing or convergence intervals converging with small finite samples from heavy-tailed populations. This raises an important question: How quickly does the sampling distribution approach normality under different population distributions such as exponential, lognormal or Pareto? In addition, how large should the sample size be for the normal approximation to be reliable?

To address these questions, this study considers the Berry-Essen theorem, which provides a way to measure the speed of convergence to normality. The theorem shows that the approximation error decreases at a rate proportional to $\frac{1}{\sqrt{n}}$, with a constant factor determined by a standardized third absolute central moment of the population distribution. This makes it possible to evaluate the accuracy of normal approximation in finite samples and across different distribution shapes.

In this study, simulations are conducted using several non-normal distributions, including exponential, lognormal, and Pareto, to investigate how fast their sample means converge to normality compared to the uniform distribution. In addition, real-world data from the IPUMS Demographic and Health Surveys (DHS) are used to analyze a biological and demographic variable – the age at first birth among women in Kenya. This variable provides an example of real-life data that may exhibit skewness and heavy-tailed behavior. By applying the same framework to this dataset, the study explores how normal approximation performs in practice when the underlying distribution deviates from normality.

2. LITERATURE REVIEW

2.1. Theoretical Background

Statistical Inference is the process of drawing conclusions about a larger population using sample data. In many statistical analyses, the sample mean plays an important role when estimating population characteristics. Therefore, it is important to understand the population distribution and the sampling distribution of the sample mean.

The population distributions explain the original shape of distributions, such as skewness, tailness, or other characteristics of how the original data looks. Based on these characteristics, statistical theories help us understand how the sampling distribution of the sample mean behaves. One of the most important theories in this context is the Central Limit Theorem, which explains how the sample mean approaches a normal distribution under certain conditions.

Central Limit Theorem

The Central Limit Theorem (CLT) states that the sampling distribution of the sample mean becomes approximately normal as the sample size increases, regardless of the shape of the population distribution.

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed (i.i.d.) random variables with a mean μ and a finite σ^2.

The sample mean is defined as: $X_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$

This sample mean is important as it enables us to estimate the population mean.

By the Law of Large Numbers, the sample mean converges to the population mean as the sample size increases: $X_n \to \mu$ as $n \to \infty$, $\overline{X_n} \to \mu$

However, the CLT mainly describes what happens when the sample size becomes large. It does not fully explain how accurate the normal approximation is for finite sample sizes. In practice, real-world data often have skewed shapes or heavy tails, which may affect how quickly the sampling distribution approaches normality. For example, highly skewed distributions may require larger sample sizes to resemble a normal distribution.

Because real data are often finite and non-uniform, we use the Berry-Essen Theorem to study how convergence behaves under different distribution shapes.

Berry-Esseen Theorem

The Berry-Esseen theorem complements these concerns of how much n is enough or how does the different characteristics of distributions affect the convergence speed.

While the CLT guarantees convergence, the Berry-Esseen Theorem gives a quantitative bound on the difference between the distribution of the standardized sample mean and the standard normal distribution.

If $X_1, X_2, \cdots, X_n$ are i.i.d. random variables with the population mean μ, variance σ^2, and finite third absolute moment, then

$$sup_X \left| F_n(X) - \Phi(X) \right| \leq C \bullet \frac{\rho}{\sqrt{n}}$$

where $\rho = \frac{E|(X-\mu)|^3}{\sigma^3}$.

This quantity reflects properties of the population distribution, including skewness and tail heaviness. Distribution with larger values of ρ tend to converge more slowly to normality.

Distribution Considered

In this study, we decided to examine different types of distributions that differ in skewness and tail behavior. These characteristics are important because they influence the convergence speed described by the CLT and Berry-Esseen Theorem.

We selected four distributions:

- Uniform distribution

- Exponential distribution
- Lognormal distribution
- Pareto distribution

Uniform distribution

The uniform distribution has constant probability over an interval [a,b]:

$f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$.

It is symmetric and light-tailed. Because of these properties, the sampling distribution is expected to approach normality relatively quickly.

Exponential Distribution

The exponential distribution describes the time between independent events:

$f\ (x;\ \lambda) = \lambda e^{-\lambda x},\ x \geq 0$

It is skewed but has relatively light tails. Compared to the uniform distribution, convergence to normality may occur more slowly.

Lognormal Distribution

The lognormal distribution arises when the logarithm of a variable follows a normal distribution:

$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} exp(-\frac{(lnx-\mu)^2}{2\sigma^2})$

It is strongly skewed with a heavy right tail. Because of this, larger sample sizes may be required for the sampling distribution to approximate normality.

Pareto distribution

The Pareto distribution is defined as:

$f\left(x; a,\ x_m\right) = \frac{ax_m^a}{x^{a+1}}, x \geq x_m$

It is commonly used to model heavy-tailed phenomena such as income, wealth, or biological extremes.

The shape parameter α controls tail thickness. Smaller values correspond to heavier tails. Because of its heavy-tailed nature, convergence to normality is expected to be slower.

By comparing these four distributions, this study investigates how differences in skewness and tail heaviness influence the speed of the convergence predicted by the CLT and Berry-Esseen Theorem. Furthermore, these theoretical results are applied to real-world demographic data to examine how well normal approximation performs when the underlying data are not normally distributed.

2.2. Empirical and Simulation-Based Studies

In addition to the theoretical results provided by the Central Limit Theorem and the Berry-Esseen theorem, several studies have examined how convergence behaves in finite samples through empirical and simulation-based approaches. These papers highlight that real-world data often deviates from idealized assumptions and requires careful interpretation.

Cluaset, Shlaizi, and Newman (2009) conduct an extensive empirical analysis of power-law distributions using real-world data. Their study shows that heavy-tailed behavior is frequently observed in empirical data and emphasizes the importance of applying appropriate statistical methods when identifying and estimating such distributions. This suggests that distributional features such as tail behavior should be carefully considered in statistical analysis.

Goldstein, Morris, and Yen (2004) examine the problems of estimation methods of power law distribution, which produce biased and inaccurate results. To test the power-law, they conducted an experiment to compare the power-law distribution between different estimation methods that are commonly used. They used the Kolmogorov—Smirnov (KS) Type Goodness-of-Fit Test to measure the Power-Law Distribution. Their study concludes that there are inaccurate and unreliable results in common estimations, and thus MLE and KS-type can be used to reduce these deviating results. They proposed the MLE method to have a more robust estimation and the KS-type test to evaluate the quantitative measurement of goodness-of-fit, which tests if the data uphold the power-law distribution.

Lastly, Newman also exhibits power-law distributions through analyzing real-world data. He used normal histograms, log-log plots, and cumulative distributions to evaluate the power-law distributions of the data. He also designed the mathematical models and showed the tail behaviors of the real-world data. His mathematical model shows that the probability decreases as the sample size increases, with heavy tails of the distributions and scale-free distributions by measuring the exponent ɑ value. Therefore, he proved that the real-world data also follow the power-law distributions, most of which are associated with multiple mechanisms, rather than a single one to analyze.

These multiple prior papers also thoroughly evaluated how the real-world data have a different distribution shape compared to theoretical distributions, and how these deviating values contribute to the experiments.

3. SIMULATION RESULTS

3.1. Simulation Design

The simulation was designed to examine how the sampling distribution of the mean approaches normality under different sample sizes and different characteristics. This simulation aims to illustrate how the Central Limit Theorem operates under different population distributions and how fast each distribution converges to a normal distribution. The distributions considered in this study differ in terms of their skewness and tail heaviness.

In the simulation, the following parameter values were used for each distribution. The uniform distribution was defined as Uniform (0,1). The exponential distribution was specified with rate parameter λ=1. The lognormal distribution was generated with meanlog =0 and sdlog=1. The Pareto distribution was defined with shape parameter α = 5.3 and scale parameter . These parameter choices were selected to represent distributions with increasing levels of skewness and tail heaviness while maintaining finite variance, which is required for the Central Limit Theorem to hold.

To approximate the population parameters, a large number of random variables were first generated from each distribution. Specifically, 5,000 observations were drawn to construct a large population sample, and the population mean and standard deviation were estimated from this sample. These values were later used for standardization.

Next, to examine how the sampling distribution approaches the normal distribution as the sample size increases, samples of size $n = 10,\ 30,\ 50,\ 100,\ and\ 200$ were drawn. For each sample size, the sampling process was repeated 2,000 times in order to construct an empirical sampling distribution of the sample mean for each distribution and sample size combination.

The sample means were then standardized by using population mean and standard deviation. The accuracy of approximation of the normality was measured by Kolmogorov distance

$$D_n = sup_x \left| F_n(x) - F(x) \right|$$

, which represents maximum absolute difference between two cumulative distributions. Here, $F_n(x)$ indicates the empirical cumulative distribution function of the standardized sample means, while $F(x)$ represents the cumulative distribution function of the standard normal distribution. Therefore, a smaller value of D_n indicates a closer approximation to normality.

This procedure was applied to four different distributions that vary in skewness and tail heaviness: the uniform, exponential, lognormal, and Pareto distributions. The uniform distribution has a symmetric shape and the lightest tail among the four distributions. The exponential distribution has mild skewness and a heavier tail than the uniform distribution. The lognormal distribution exhibits strong skewness and a heavier right tail, while the Pareto distribution has the most pronounced skewness and the heaviest tail among the distribution considered. By comparing these distributions, the simulation allows us to examine how skewness and tail heaviness affect the finite-sample behavior of the Central Limit Theorem.

3.2. Simulation Results for the Central Limit Theorem

Table 1 presents the normal approximation error measured using the Kolmogorov distance D_n. A smaller value of D_n indicates that the sampling distribution of the standardized sample mean is closer to the standard normal distribution and therefore has a smaller approximation error.

Table 1. Simulation Results: Normal Approximation Error by Distribution

Distribution	n	reps	D_n
Uniform	10	2000	0.0157
Uniform	30	2000	0.0166
Uniform	50	2000	0.0150
Uniform	100	2000	0.0172
Uniform	300	2000	0.0116
Exponential	10	2000	0.0378
Exponential	30	2000	0.0370
Exponential	50	2000	0.0199
Exponential	100	2000	0.0139
Exponential	300	2000	0.0289
Lognormal	10	2000	0.0836
Lognormal	30	2000	0.0486
Lognormal	50	2000	0.0536
Lognormal	100	2000	0.0420
Lognormal	300	2000	0.0251
Pareto	10	2000	0.0723
Pareto	30	2000	0.0544
Pareto	50	2000	0.0510
Pareto	100	2000	0.0383
Pareto	300	2000	0.0101

Table 1. Simulation Results: Normal Approximation Error by Distribution

Table 1 presents the normal approximation error, measured by Kolmogorov distance D_n across different population distributions and sample sizes. For each distribution and sample size, the sampling distribution of the standardized sample mean was conducted using 2,000 repeated simulations(reps), and the distance from the standard normal distribution was calculated.

Overall, the four distributions show a general tendency of D_n to decrease as the sample size increases. This pattern suggest that the accuracy of the normal approximation improves with larger sample sizes.

For the uniform distribution, the values of D_n remain relatively small across all sample sizes, indicating that the sampling distribution converges quickly to the normal distribution with relatively small approximation error. A slight increase in D_n from 0.0150 to 0.0172 is observed when n is 100, which reflects variation arising from the simulation process. Also, D_n significantly decreases to 0.0116. at $n = 300$, showing a computational error made a significantly small value. Since the Central Limit Theorem is asymptotic, simulations with finite repetitions may show small fluctuations rather than a perfectly monotonic decrease.

In the exponential distribution, the value of D_n generally decreases as the sample size increases. Especially, there is significant decrease in D_n from 0.0370 to 0.0199, where n is 50. This suggests that the sampling distribution becomes substantially closer to the normal distribution at this point.

The lognormal distribution displays a similar pattern but requires larger sample sizes to achieve smaller approximation errors. In particular, a significant decrease in D_n is observed when the sample size increases to n=300, indicating that larger samples are needed for the sampling distribution to approximate normality more closely.

The Pareto distribution also has the largest decrease in distance of 0.0383 and 0.0101, when n is 100 and 300, respectively. Similar to the lognormal distribution, the heavy tail of the Pareto distribution requires larger sample sizes for the sampling distribution to approximate the normal distribution accurately.

Comparing the four distributions suggests that the symmetry and tail behavior of the population distribution influence the speed of normal approximation. The uniform distribution, which is symmetric and light-tailed, converges most quickly to the normal distribution. The exponential distribution, which introduces moderate skewness, converges more slowly. The lognormal and Pareto distributions, which are highly skewed and have heavier tails, require larger sample sizes before the sampling distribution closely resembles a normal distribution.

The occasional increases in D_n observed in some cases do not indicate errors in the simulation rather than reflect natural variation arising from a finite number of simulation repetitions.

Figure 1. Finite-Sample Normal Approximation Error Across Distribution

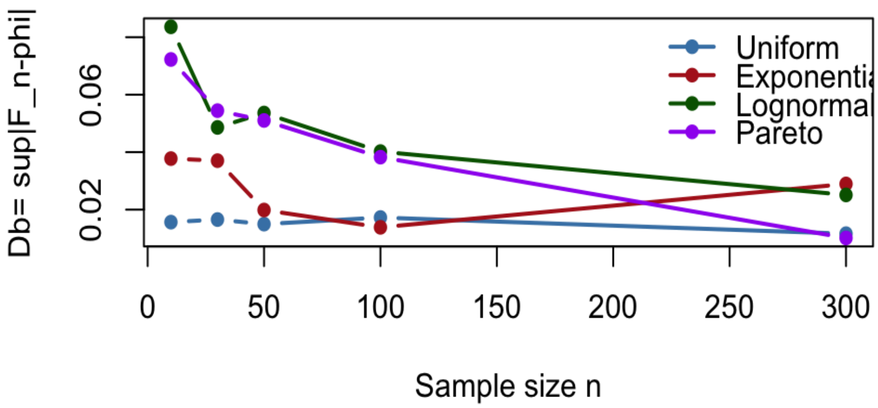

Figure 1. Finite-Sample Normal Approximation Error Across Distribution. The figure plots the Kolmogorov distance between the empirical distribution of the standardized sample mean and the standard normal distribution for each distribution as the sample size increases.

Figure 1 shows how the value of D_n changes for four different distributions as the sample size increases. In general, the value of D_n decreases as the sample size n increases for all four distributions. This shows that the sampling distribution of the standardized sample mean gradually approaches the normal distribution as the sample size becomes larger, which is consistent with the Central Limit Theorem.

In uniform distribution, the values of D_n are the smallest across all sample sizes. Even with a small sample size, the sampling distribution already appears close to the normal distribution. This suggests that the symmetric shape and light tail of the uniform distribution lead to a smaller approximation error and faster convergence to normality compared to the other distributions.

In the Exponential distribution, the value of D_n is larger than that of the uniform distribution but still shows a decreasing trend as the sample size increases. Since the exponential distribution is slightly skewed and has a heavier tail than the uniform distribution, the approximation error is somewhat larger and the convergence to the normal distribution occurs at a slightly slower rate.

In the lognormal distribution, it displays the biggest value of D_n when the sample size is small. This reflects the strong skewness and heavier right tail of the distribution. As the sample size increases, the value of D_n gradually decreases, but the convergence to normality is noticeably slower than in the uniform and exponential distributions.

Lastly, the Pareto distribution shows relatively large values of D_n for small sample sizes and a slower convergence pattern. Because the Pareto distribution has a very heavy tail and strong skewness, larger samples are required for the sampling distribution of the mean to approach the normal distribution.

Overall, Figure 1 shows that as the sample size increases all distributions showed decreasing value of D_n. However, distributions with greater skewness and heavier tail tend to have larger approximation errors, particularly when the sample size is small, and they convergence to normality more slowly.

3.3. Simulation Results for the Berry-Esseen Theorem

Table 1 and Figure 1 shows how approximation error denoted by D_n changes with increasing sample size. However, simply observing the decrease in D_n does not reveal the theoretical rate at which the convergence occurs.

To examine the speed of convergence more clearly, the simulation results are interpreted through the Berry-Esseen theorem. The Berry-Esseen theorem provides a quantitative bound on the error in the Central Limit Theorem and states that the approximation error decreases at a rate proportional to $\frac{1}{\sqrt{n}}$.

To evaluate this theoretical rate of convergence, the simulation calculates the scaled quantity $D_n\sqrt{n}$, which multiplies the approximation error by the square root of the sample size. If the Berry-Esseen rate holds, the value of $D_n\sqrt{n}$ should remain relatively stable as the sample size increases.

By examining this scaled measure, the analysis allows for a clearer comparison between the empirical convergence behavior observed in the simulation and the theoretical rate suggested by the Berry-Esseen theorem.

Table 2. Scaled Normal Approximation Error Across Distributions

Distribution	n	reps	D_n	$D_n\sqrt{n}$
Uniform	10	2000	0.0157	0.0495
Uniform	30	2000	0.0166	0.0907
Uniform	50	2000	0.0150	0.1058
Uniform	100	2000	0.0172	0.1725
Uniform	300	2000	0.0116	0.0116
Exponential	10	2000	0.0378	0.1194
Exponential	30	2000	0.0370	0.2029
Exponential	50	2000	0.0199	0.1404
Exponential	100	2000	0.0139	0.1385
Exponential	300	2000	0.0289	0.5012
Lognormal	10	2000	0.0836	0.2644

Lognormal	30	2000	0.0486	0.2662
Lognormal	50	2000	0.0536	0.3791
Lognormal	100	2000	0.0420	0.4020
Lognormal	300	2000	0.0251	0.4349
Pareto	10	2000	0.0723	0.2285
Pareto	30	2000	0.0544	0.2982
Pareto	50	2000	0.0510	0.3609
Pareto	100	2000	0.0383	0.3827
Pareto	300	2000	0.0101	0.1756

Table 2. Scaled Normal Approximation Error Across Distributions. Table 2 presents the scaled normal approximation error $D_n\sqrt{n}$ across different population distributions and sample sizes. By multiplying the Kolmogorov distance D_n by $\sqrt{n}$, the table adjusts for the theoretical convergence rate suggested by the Berry-Esseen theorem.

Table 2 presents the value of $D_n\sqrt{n}$ for each distribution and sample size. If the Berry-Esseen theorem describes the convergence behavior well, these values are expected to remain relatively stable as the sample size increases.

The uniform distribution shows the smallest values of $D_n\sqrt{n}$ across all sample sizes. This indicates that the symmetric shape and light tail of the uniform distribution allow the sampling distribution to converge quickly and steadily to the normal distribution. The computational error happened at $n = 3$, sharply dropping the value of $D_n\sqrt{n}$ to 0.0116.

The exponential distribution also shows relatively stable values of $D_n\sqrt{n}$, although the values are slightly larger than those of the uniform distribution. This suggests that the mild skewness and heavier tail of the exponential distribution lead to a somewhat slower convergence rate.

In the lognormal distribution, the values of $D_n\sqrt{n}$ are noticeably larger than those of the uniform and exponential distributions. This reflects the strong skewness and heavier tail of the distribution, which increase the approximation error and slow down the convergence toward normality.

Lastly, the Pareto distribution shows relatively large values of $D_n\sqrt{n}$. In particular, the scaled error tends to increase when the sample size becomes larger. This pattern suggests that the heavy-tailed nature of the Pareto distribution can produce greater variability in the approximation error.

Overall, the results show that the uniform distribution has the smallest and most stable values of $D_n\sqrt{n}$. Indicating the fastest convergence to normality. As the distributions become more skewed and heavy-tailed, the scaled approximation error becomes larger and less stable. These results suggest that skewness and tail heaviness play an important role in the finite-sample behavior of the Central Limit Theorem.

Figure 2. Scaled Approximation Error Across Distributions

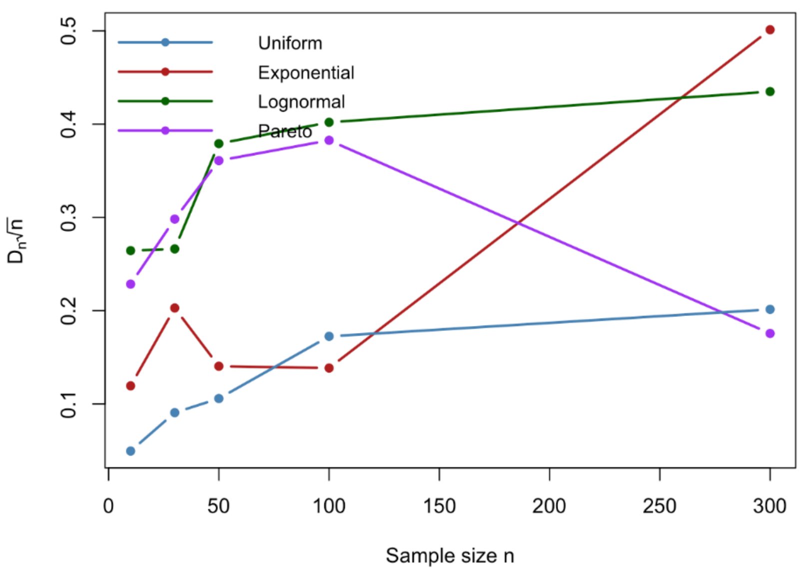

Figure 2. Scaled Approximation Error Across Distributions. Figure 2 plots scaled normal approximation error, $D_n\sqrt{n}$, across different population distributions and sample sizes. By rescaling the Kolmogorov distance D_n by $\sqrt{n}$, the figure illustrates whether the convergence error follows the $\frac{1}{\sqrt{n}}$ rate suggested by the Berry-Esseen theorem.

All four distributions show a similar trend of increasing value of $D_n\sqrt{n}$ as the sample size increases. Although the values are not perfectly constant, they remain within a similar range, which is broadly consistent with the convergence rate suggested by the Berry-Esseen theorem.

The uniform distribution has the smallest value of $D_n\sqrt{n}$ among all four distributions. This indicates its symmetric shape and light tail allow the sampling distribution to approach normality with relatively small error. Even though the values slightly increase as the sample size increases, the overall magnitude remains low, suggesting fast and stable convergence.

The exponential distribution also exhibits an overall increasing trend over the increasing sample size. The value of $D_n\sqrt{n}$ is a slightly larger than the uniform distribution due to its little skewness and light tail.

While the overall trend follows the theoretical expectation, some fluctuation is observed, especially at larger sample sizes. This variation is likely due to simulation noise from finite repetitions rather than a failure of the theoretical results.

For the lognormal distribution, the values of $D_n\sqrt{n}$ are consistently higher than those of the uniform and exponential distributions. This suggests that stronger skewness and heavier tails lead to larger approximation errors. As a result, the convergence to normality is slower compared to distributions with lighter tails.

The Pareto distribution also shows relatively large values of $D_n\sqrt{n}$, along with noticeable variation across sample sizes. Since the Pareto distribution has a heavy tail, extreme values can have a stronger influence on the sample mean, which leads to less stable approximation behavior.

Overall, the results indicate that while the scaled error does not remain perfectly constant, it stays within a comparable range across sample sizes. At the same time, distributions with greater skewness and heavier tails tend to have larger values of $D_n\sqrt{n}$, implying slower convergence to the normal distribution.

4. EMPIRICAL APPLICATION

In the previous section, we focused on simulation experiments through R programming to examine whether the Central Limit Theorem and the Berry-Esseen theorem hold under different distributional shapes and characteristics. The results showed that as the sample size increases, all four distributions converge toward normality. We also found that greater skewness and heavier tails slow down the rate of convergence.

However, these simulation results are based on artificially generated data. In practice, real-world data often have more complex and irregular distributional features. Therefore, it is important to examine whether similar patterns appear in actual data. To address this, we apply our analysis to real-world data from the IPUMS International database. Specifically, we use survey data from Kenya on women's age at first birth. The goal is to investigate whether the convergence patterns observed in the simulation also appear in real data.

4.1. Data Description

The data used in this analysis are microdata from the IPUMS International database, which provides standardized survey data across different countries' populations. This study uses Demographic and Health Survey (DHS) data from Kenya, which contains information on women's demographic characteristics and reproductive history.

The main variable is women's age at their first birth. This variable is expected to exhibit a non-normal distribution, making it suitable for examining convergence behavior in finite samples. Furthermore, the study considers the women's individual sample weight to account for the survey design. DHS data are collected using a complex sampling design, so simple unweighted averages may not accurately represent the population. The sample weights adjust for the unequal selection probabilities, nonresponse, and population representation. These weights are reported with six decimal places and are typically divided by 1,000,000 before use. Applying these weights allows the analysis to better reflect the underlying population.

Table 3: Summary Statistics of Age at First Birth

N	Mean	Median	S.D.	Minimum	Maximum	Skewness	Kurtosis
77,381	19.1877	19.0000	3.7888	10.0000	46.0000	0.6651	4.0437

Table 3: Summary Statistics of Age at First Birth. Table 3 presents summary statistics for the age at first birth variable using the Kenya DHS sample from IPUMS International. The table reports the number of observations (N), mean, median, standard deviation (S.D.), minimum and maximum values, as well as measures of skewness and kurtosis.

Table 3 presents the summary statistics of the data. The sample size is 77,381. The mean age at first birth is 19.1877, and the median is 19.0000, indicating that the central tendency of the distribution is around 19 years. The standard deviation is 3.7888, suggesting a moderate level of variation in the data.

The minimum value is 10.0000, and the maximum value is 46.0000, indicating that the range of data is relatively wide. The skewness is 0.6651, which indicates a positively skewed distribution, meaning that the right tail is longer. The kurtosis is 4.0437, which is higher than the value of 3 for a normal distribution. This suggests that the distribution has heavier tails than a normal distribution.

Overall, these results indicate that the distribution of age at first birth deviates from normality. Therefore, it provides a suitable real-world example for analyzing the Central Limit Theorem.

Figure 3: Distribution of Age at First Birth in Kenya

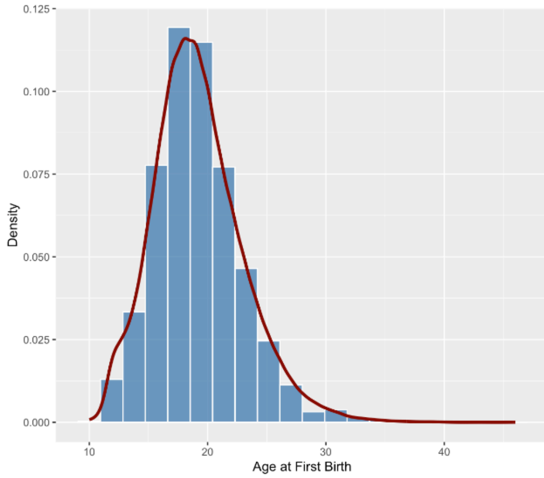

Figure 3: Distribution of Age at First Birth in Kenya. Figure 3 plots the distribution of age at first birth in the Kenya DHS sample.

Figure 3 reflects the summary statistics shown in Table 3 and illustrates the overall distribution of age at first birth. The distribution is 1right-skewed, with values ranging from 10 to 46. The highest density is observed between age 18 and 20, which is consistent with the mean and median values. As age increases, the density gradually decreases, forming a long right tail. This pattern confirms the presence of positive skewness in the data. This distribution is not symmetric and differs from a normal distribution.

Overall, Figure 3 shows that the data have a non-normal shape with right-skewness, which makes it appropriate for examining the behavior of the Central Limit Theorem in real-world data.

Then, we conduct an additional resampling-based simulation using the empirical data, since the real-world data do not perfectly follow a Pareto distribution. We focus on the upper tail of the distribution by selecting observations above the 90th percentile, corresponding to ages greater than 24. This reduces the sample size to 9,605 observations, with values ranging from 24 to 46. Compared to the full sample, with a mean of approximately 19, this subset represents a more right-skewed and tail-focused distribution.

Table 4: Estimated Pareto Tail Index for Age at First Birth

Method	Number of Observation	P90	α
OLS	9,605	24	10.6
MLE	9,605	24	11.8
Hill	9,605	24	10.9

Table 4: Estimated Pareto Tail Index for Age at First Birth. Table 4 presents estimates of the Pareto tail index(α) for the upper tail of the age at first birth distribution using three different methods: ordinary

least squares (OLS), maximum likelihood estimation (MLE), and Hill estimator. The tail is defined as observations above the 90th percentile.

To examine whether this tail follows a Pareto-like structure, we estimate the tail index(α) using three methods: Ordinary Least Squares(OLS), Maximum Likelihood Estimation (MLE), and the Hill estimator. The estimated values are 10.6, 11.8, and 10.9, respectively. Although the estimation methods differ, the results are relatively close, suggesting that the upper tail of the distribution exhibits Pareto-like behavior. The relatively large values of α indicate that the tail is not extremely heavy, but still displays right-skewed characteristics.

Table 5: Normal Approximation Error Based on Empirical Tail Resampling

n	reps	D_n	$D_n\sqrt{n}$
10	2000	0.077	0.244
30	2000	0.034	0.186
50	2000	0.032	0.231
100	2000	0.031	0.309
300	2000	0.014	0.250

Table 5: Normal Approximation Error Based on Empirical Tail Resampling. Table 5 presents the normal approximation error, measured by the Kolmogorov distance D_n, based on resampling from the empirical tail distribution of the Kenya data. For each sample size n, 2,000, repeated simulations (reps) are conducted, and the sampling distribution of the standardized sample mean is compared to the standard normal distribution. The table also reports the scaled error $D_n\sqrt{n}$ to examine the convergence rate suggested by the Berry-Esseen theorem.

Table 5 reports the results of the resampling-based simulation using the empirical tail distribution. As the sample size increases, the value of D_n generally decreases, indicating that the sampling distribution of the standardized sample mean becomes closer to the normal distribution. For example, D_n decreases from 0.077 at n=10 to 0.014 at n=300, which is consistent with the Central Limit Theorem.

However, the scaled measures $D_n\sqrt{n}$ do not remain perfectly constant across different sample sizes. Instead, it fluctuates within a certain range, reflecting variability due to finite sample size and the underlying distributional characteristics. This suggests that while the general convergence toward normality is observed, the exact convergence rate does not strictly follow the theoretical bound implied by the Berry-Esseen theorem.

Overall, these findings suggest that the empirical data exhibit convergence patterns broadly consistent with the Central Limit Theorem, while showing some deviations from the idealized convergence rate predicted by the Berry-Esseen theorem.

5. CONCLUSION

This paper examines how the sampling distribution of the mean behaves under different population distributions and finite sample sizes. Through simulations using uniform, exponential, lognormal, and Pareto distributions, the results show that all distributions converge toward the normal distributions as the sample size increases.

However, the speed of convergence differs depending on the characteristics of the distribution. Distributions with more skewness and heavier tails tend to converge more slowly, while symmetric and light-tailed distributions converge more quickly. In particular, the uniform distribution shows the fastest convergence, followed by the exponential distribution, while the lognormal and Pareto distributions require larger sample sizes to approach normality.

The empirical analysis using Kenya DHS data from IPUMS shows a right-skewed distribution with moderate tail heaviness. Tail analysis suggests a Pareto-like structure, supported by consistent estimates from OLS, MLE, and the Hill estimator. Resampling-based simulation further shows that the sampling distribution approaches normality as the sample size increases.

At the same time, some deviations from the theoretical results are observed. Due to finite sample size, the approximation error does not decrease perfectly monotonically, and the scaled error does not remain constant. Especially, in the real world, there are more data with many small sample sizes and heavy-tailed populations used for statistical testing and confidence intervals. These results indicate that while the Central Limit Theorem provides a useful approximation in practice, the exact rate of convergence may vary depending on the data and may not strictly follow the Berry-Esseen bound.

Overall, the findings suggest that normal approximation improves as the sample size increases, but its accuracy and convergence speed depend on the underlying distributional characteristics.

REFERENCES

Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the american mathematical society*, *49*(1), 122-136.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, *51*(4), 661-703.

Esseen, C.-G. (1942). On the Liapunoff limit of error in the theory of probability. *Arkiv för Matematik, Astronomi och Fysik*, *A28*(1), 1–19

Feller, W. (1991). *An introduction to probability theory and its applications, Volume 2* (Vol. 2). John Wiley & Sons.

Gnedenko, B. V., & Kolmogorov, A. N. (1968). *Limit distributions for sums of independent random* *variables* (Vol. 2420). Addison-wesley.

Goldstein, M. L., Morris, S. A., & Yen, G. G. (2004). Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, *41*(2), 255-258.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 1163-1174.

IPUMS International. (2023). IPUMS International database. University of Minnesota.

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary physics*, *46*(5), 323-351.

The DHS Program. (2023). Kenya DHS data.