JOURNAL OF STATISTICAL SCIENCES & DATA ANALYTICS
Department of Statistics, Ahmadu Bello University, Zaria, Nigeria | www.jossda.abu.edu.ng www.jossda.com.ng

Statistical Modelling of Positive Continuous Data Using the Gamma Distribution: A Simulation and Real-Life Data Study

I. A. Sadiq[1,*], J. Y. Kajuru[1], Y. Zakari[1], A. Usman[1], S. I. Doguwa[1], N. I. Muhammad[1] and R. Salihu

[1]*Department of Statistics, Ahmadu Bello University, Zaria, Nigeria*
[1]*Department of Statistics, Kaduna State University, Nigeria*

Corresponding Author:
I. A. Sadiq, Department of Statistics, Ahmadu Bello University, Zaria, Nigeria.
isabubakar@abu.du.ng | +234 813 752 6770 | ORCID: 0000-0002-2122-9344

Abstract

The Gamma distribution is a flexible continuous probability distribution widely used for modelling positive, right-skewed data in engineering, reliability, medicine, and biological sciences. This study evaluates the applicability of the Gamma distribution to ten real-life datasets, including ball bearings, air conditioning, Boeing, analgesic, accelerated life test, strain level, strength of glass, vinyl chloride, RT_CT, and virulent tubercle bacilli data, alongside simulated datasets with sample sizes of 100, 200, 300, 400, 500, and 1000. Parameters were estimated using the Maximum Likelihood Estimation (MLE) method, while model adequacy was assessed through goodness-of-fit measures and graphical diagnostics implemented in R. A simulation study was conducted to examine the performance of the MLE across varying sample sizes. The results showed that the Gamma distribution achieved improved model fit with increasing sample size, demonstrating favourable large-sample properties. Among the real-life datasets, the analgesic dataset provided the best fit according to the selected goodness-of-fit criteria. Most datasets exhibited positive skewness and varying degrees of kurtosis, supporting the suitability of the Gamma distribution for modelling non-negative, asymmetric data. The findings demonstrate that the Gamma distribution is an effective model for positively skewed data and performs particularly well for simulated datasets and selected real-life applications. The study highlights the usefulness of Maximum Likelihood Estimation for parameter estimation and reinforces the Gamma distribution as a valuable tool for statistical modelling in reliability, engineering, and biomedical research.

Keywords: Gamma distribution; Maximum likelihood estimation; Goodness-of-fit; Simulation study; Reliability analysis; Real-life data.

1. Introduction

Probability distributions play a fundamental role in statistical modelling by providing mathematical frameworks for describing the behaviour of random variables and supporting statistical inference. Among continuous probability distributions, the Gamma distribution has gained considerable attention because of its flexibility in modelling positive-valued, right-skewed data encountered in many scientific and engineering applications (Johnson, Kotz, & Balakrishnan, 1994). Introduced by Karl Pearson in the late nineteenth century, the Gamma distribution has become one of the most widely used models for analysing lifetime, waiting-time, and reliability data (Pearson, 1893).

The Gamma distribution is a two-parameter continuous distribution defined by a shape parameter (α) and a scale parameter (θ). These parameters determine the variability, skewness, and overall shape of the distribution, enabling it to represent a wide range of positively skewed data. A notable characteristic of the Gamma distribution is its relationship with several important probability distributions. For instance, when the shape parameter equals one, the Gamma distribution reduces to the Exponential distribution, while the Chi-square distribution is a special case of the Gamma family. These relationships have contributed to its widespread application in statistical theory and practice (Casella & Berger, 2002).

The Gamma distribution has been successfully applied in numerous disciplines, including reliability engineering, medicine, finance, hydrology, environmental science, insurance, and biological research. In reliability engineering, it is commonly used to model the lifetime of mechanical systems and electronic components (Lawless, 2003). In medicine and survival analysis, it provides a suitable model for recovery times, incubation periods, and patient waiting times. In hydrology and environmental studies, the Gamma distribution has been widely employed to model rainfall amounts and streamflow characteristics because of its ability to accommodate positively skewed observations. Similarly, in actuarial science and finance, it has been used to describe insurance claim sizes, aggregate losses, and other non-negative financial variables (Johnson et al., 1994).

The usefulness of the Gamma distribution depends largely on accurate estimation of its parameters. Reliable parameter estimates improve statistical inference, prediction, and decision-making across various applications. Several estimation techniques have been proposed for the Gamma distribution, including the Method of Moments (MME), Least Squares Estimation (LSE), Weighted Least Squares Estimation (WLSE), Maximum Likelihood Estimation (MLE), and Maximum Product Spacing (MPS). Among these methods, MLE is widely regarded as the standard approach because of its desirable asymptotic properties, such as consistency, efficiency, and asymptotic normality (Casella & Berger, 2002; Lehmann & Casella, 1998). However, the MPS estimator has emerged as an attractive alternative because of its robustness and favourable performance, particularly when likelihood functions are difficult to optimise or when sample sizes are relatively small (Cheng & Amin, 1983; Ranneby, 1984).

Simulation studies are widely used in statistics to compare the performance of competing estimation methods under controlled experimental conditions. They enable researchers to examine estimator accuracy, bias, and precision across different sample sizes and parameter settings before applying the methods to empirical data (Efron & Tibshirani, 1993). Complementing simulation studies with analyses of real-life datasets provides practical evidence of the suitability of statistical models for describing observed phenomena. Such combined investigations are valuable for assessing both the theoretical and practical performance of probability distributions.

Motivated by the extensive applications of the Gamma distribution and the need for reliable parameter estimation, this study evaluates the performance of the Maximum Likelihood Estimation and Maximum Product Spacing methods using simulated datasets of varying sample sizes and selected real-life datasets. The adequacy of the fitted Gamma models is assessed using appropriate goodness-of-fit measures and graphical diagnostics implemented in R. The findings are expected to provide empirical evidence on the comparative performance of the estimation methods and further demonstrate the usefulness of the Gamma distribution for modelling positive continuous data across diverse application areas.

Parameter estimation plays a central role in statistical modelling because the quality of statistical inference depends largely on the accuracy of the estimated model parameters. Although numerous estimation methods have been proposed for the Gamma distribution, the Maximum Likelihood Estimation (MLE) and Maximum Product Spacing (MPS) methods remain among the most widely used due to their desirable theoretical properties. However, their relative performance may vary depending on sample size and the characteristics of the underlying data.

Most existing studies have focused on individual estimation methods or limited comparisons under specific conditions. Comprehensive evaluations of MLE and MPS using both simulation studies and diverse real-life datasets remain relatively scarce. Consequently, there is a need to compare these estimation techniques systematically and assess their effectiveness in modelling positive continuous data using the Gamma distribution. This study addresses this gap by comparing the performance of MLE and MPS through simulation and empirical applications.

The aim of this study is to evaluate the performance of parameter estimation methods for the Gamma distribution and assess its suitability for modelling positive continuous data.

The specific objectives are to:

i. estimate the parameters of the Gamma distribution using the Maximum Likelihood Estimation (MLE) method;

ii. estimate the parameters of the Gamma distribution using the Maximum Product Spacing (MPS) method;

iii. evaluate the performance of the estimators through simulation studies under different sample sizes; and

iv. assess the adequacy of the Gamma distribution in modelling selected real-life datasets using appropriate goodness-of-fit measures.

This study contributes to the statistical literature by providing a comparative evaluation of the Maximum Likelihood Estimation and Maximum Product Spacing methods for estimating the parameters of the Gamma distribution. The findings will assist researchers and practitioners in selecting appropriate estimation techniques for modelling positive continuous data. Furthermore, the study demonstrates the applicability of the Gamma distribution to a variety of real-life datasets from engineering, reliability, biomedical, and industrial settings. The results are expected to serve as a useful reference for statisticians, researchers, and data analysts involved in probabilistic modelling, reliability analysis, risk assessment, and other applications involving positively skewed continuous data.

Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX

DOI: 10.XXXXX/jossda.2026.1.1.016

A literature review provides a critical evaluation of existing knowledge on a research topic by identifying established findings, theoretical developments, methodological approaches, and research gaps. It serves as the foundation for a study by synthesising previous research, highlighting unresolved issues, and justifying the need for further investigation (Creswell & Creswell, 2018).

The Gamma distribution has been extensively studied because of its flexibility in modelling positive continuous random variables. Since its introduction by Pearson (1893), the distribution has become one of the most important probability models in statistics, reliability engineering, actuarial science, hydrology, medicine, and biological sciences. Johnson, Kotz, and Balakrishnan (1994) presented a comprehensive treatment of the Gamma distribution, discussing its mathematical properties, probability density function, cumulative distribution function, moments, and relationships with other continuous distributions. They showed that the Gamma distribution can accommodate varying degrees of skewness through its shape parameter, making it suitable for modelling lifetime and waiting-time data.

Lawless (2003) emphasised the importance of the Gamma distribution in lifetime and reliability analysis. According to the author, the distribution provides a flexible model for analysing failure times of mechanical and electronic components and has been successfully applied in reliability engineering because of its ability to represent increasing, decreasing, and constant hazard-rate patterns under different parameterisations. Similar applications have been reported in survival analysis and biomedical research, where Gamma models are used to analyse recovery times, incubation periods, and other non-negative continuous outcomes.

Accurate parameter estimation remains a fundamental aspect of applying the Gamma distribution in practice. Among the available estimation techniques, Maximum Likelihood Estimation (MLE) is the most widely used because of its desirable statistical properties. Casella and Berger (2002) showed that MLE is consistent, asymptotically unbiased, asymptotically efficient, and asymptotically normal under regularity conditions. Likewise, Lehmann and Casella (1998) demonstrated that maximum likelihood estimators possess favourable large-sample properties, making them suitable for statistical inference and prediction.

Although MLE performs well for moderate and large samples, numerical optimisation may become challenging in some situations, particularly for complex likelihood functions or small sample sizes. To overcome these limitations, Cheng and Amin (1983) proposed the Maximum Product Spacing (MPS) estimation method as an alternative to MLE. Their study demonstrated that MPS provides reliable parameter estimates by maximising the geometric mean of spacings between successive cumulative distribution function values rather than the likelihood function itself. The authors reported that the method performs well even when maximum likelihood estimation encounters convergence difficulties.

Ranneby (1984) further established the theoretical foundation of the MPS estimator and demonstrated its close relationship with the Kullback–Leibler information criterion. The study proved that MPS estimators are consistent and asymptotically efficient under regularity conditions, making them attractive alternatives to maximum likelihood estimators. Since then, MPS has received considerable attention in the statistical literature because of its robustness and good finite-sample performance.

Simulation studies have become one of the standard approaches for comparing competing estimation methods. Efron and Tibshirani (1993) noted that simulation experiments allow researchers to investigate estimator bias, variance, mean squared error, and overall efficiency under controlled experimental conditions. Such studies provide valuable insight into the behaviour of estimation methods across different sample sizes and parameter configurations before they are applied to empirical datasets.

The Gamma distribution has also been applied successfully to numerous real-life datasets. Johnson et al. (1994) reported applications in rainfall modelling, insurance claims, waiting-time analysis, and industrial reliability. Lawless (2003) further illustrated its usefulness in reliability and survival analysis using engineering and biomedical data. These studies confirm that the Gamma distribution is particularly appropriate for modelling positive continuous variables exhibiting moderate to substantial right skewness.

Despite the extensive literature on the Gamma distribution, relatively few studies have simultaneously compared the performance of the Maximum Likelihood Estimation and Maximum Product Spacing methods using both simulated datasets and multiple real-life datasets. Most published works focus either on the theoretical properties of individual estimators or on applications to specific datasets. Consequently, there is a need for a comprehensive evaluation that combines simulation experiments with empirical applications to assess the relative performance of MLE and MPS and the adequacy of the Gamma distribution for modelling positive continuous data. This study addresses this gap by comparing these two estimation methods across different sample sizes and applying the fitted Gamma model to several real-life datasets using appropriate goodness-of-fit measures.

2. Materials and Methods

2.1. Research Design

This study adopts a quantitative research design to evaluate the performance of the Gamma distribution in modelling positive continuous data. The study comprises two components: a simulation study and an empirical analysis of real-life datasets. The simulation study is used to compare the performance of the Maximum Likelihood Estimation (MLE) and Maximum Product Spacing (MPS) methods under different sample sizes, while the empirical analysis evaluates the adequacy of the Gamma distribution for modelling selected real-life datasets.

All statistical analyses are carried out using R (Version 4.6.0), with parameter estimation, simulation, and goodness-of-fit assessments performed using appropriate R packages and user-defined functions.

2.2. The Gamma Distribution

Let X be a continuous random variable following a Gamma distribution with shape parameter (α) and scale parameter (θ). The probability density function (PDF) is given by:

$$f(x;\alpha,\theta)=\frac{x^{\alpha-1}e^{-x/\theta}}{\Gamma(\alpha)\theta^{\alpha}}, \qquad x>0,\alpha,\theta>0 \tag{1}$$

where $\Gamma(\alpha)=\int_0^{\infty} x^{\alpha-1}e^{-x}dx$ is the Gamma Function.

The cumulative distribution function (CDF) is:

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
Sadiq *et al.* 2026 **DOI**: 10.XXXXX/jossda.2026.1.1.016

$$F(x)=\frac{\gamma(\alpha, x/\theta)}{\Gamma(\alpha)} \tag{2}$$

where $\gamma(\cdot,\cdot)$ denotes the lower incomplete Gamma function.

The mean and variance of the Gamma distribution are respectively given by:

$$E(X)=\alpha\theta \tag{3}$$

and

$$\mathrm{Var}(X)=\alpha\theta^2 \tag{4}$$

2.3. Maximum Likelihood Estimation

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a Gamma distribution. The likelihood function is:

$$L(\alpha,\theta)=\prod_{i=1}^{n}\frac{x_i^{\alpha-1}e^{-x_i/\theta}}{\Gamma(\alpha)\theta^{\alpha}} \tag{5}$$

The corresponding log-likelihood function is:

$$\ell(\alpha,\theta)=(\alpha-1)\sum_{i=1}^{n}\ln x_i-\frac{1}{\theta}\sum_{i=1}^{n}x_i-n\alpha\ln\theta-n\ln\Gamma(\alpha) \tag{6}$$

Differentiating the log-likelihood in equation (6) with respect to the parameters and equating the derivatives to zero yields the likelihood equations. Since these equations have no closed-form solution for the shape parameter, numerical optimisation techniques such as the Newton–Raphson algorithm are employed to obtain the MLEs.

2.4. Maximum Product Spacing Estimation

The Maximum Product Spacing (MPS) estimator is an alternative likelihood-based estimation method proposed by Cheng and Amin (1983) and further developed by Ranneby (1984).

Let $F(x_{(1)})\le\cdots\le F(x_{(n)})$ denote the ordered cumulative probabilities. Define the spacings as:

$$\mathrm{D}_i=F(x_{(i)})-F(x_{(i-1)}) \tag{7}$$

where

$$F(x_{(0)})=0, \qquad F(x_{(n+1)})=1 \tag{8}$$

The objective function is:

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX

DOI: 10.XXXXX/jossda.2026.1.1.016

$$\frac{1}{n+1}\sum_{i=1}^{n+1}\ln(D_i) \tag{9}$$

The MPS estimates are obtained by maximising $M(\alpha,\theta)$ using numerical optimisation algorithms.

2.5. Simulation Study

A Monte Carlo simulation study is conducted to compare the performance of MLE and MPS. Random samples are generated from the Gamma distribution using predetermined parameter values. Sample sizes considered are $n = 100, 200, 300, 400, 500, 1000$. For each sample size, random samples are generated, both estimation methods are applied, parameter estimates are recorded, and the procedure is repeated a large number of times (e.g., 1,000 replications).

2.6. Real-Life Datasets

Ten real-life datasets are analysed to assess the practical performance of the Gamma distribution. The datasets include: Ball Bearings, Air Conditioning, Boeing, Analgesic, Accelerated Life Test, Strain Level, Strength of Glass, Vinyl Chloride, RT-CT, and Virulent Tubercle Bacilli.

Each dataset is examined using descriptive statistics, including the mean, variance, skewness, and kurtosis, before fitting the Gamma distribution.

2.7. Presentation of the Real-Life Datasets

To evaluate the adequacy of the Gamma distribution in modelling positive continuous data, ten benchmark real-life datasets from the reliability, biomedical, engineering, and environmental literature were analysed. These datasets have been widely used in studies involving lifetime distributions and goodness-of-fit comparisons. A brief description of each dataset is presented below.

2.7.1. Ball Bearings Failure Data

This dataset was reported by Lawless (2003) and consists of the endurance times (measured in millions of revolutions) until failure of 23 deep-groove ball bearings tested under laboratory conditions. The observations are:

17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.80, 51.84, 51.96, 54.12, 55.56, 67.80, 68.44, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40.

2.7.2. Air Conditioning Failure Times

This dataset, reported by Linhart and Zucchini (1986), contains the failure times of the air-conditioning system of an aircraft. The dataset comprises 29 observations:

23, 261, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42, 20, 5, 12, 120, 11, 3, 14, 71, 11, 14, 11, 16, 901, 16, 52, 95.

2.7.3. Boeing 720 Aircraft Failure Data

This dataset was reported by Proschan (1963) and records the times between successive failures of the air-conditioning equipment in a Boeing 720 aircraft. The observations are:

74, 57, 48, 29, 502, 12, 70, 21, 29, 386, 59, 27, 153, 26, 326.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
Sadiq *et al.* 2026 **DOI**: 10.XXXXX/jossda.2026.1.1.016

2.7.4. Analgesic Relief-Time Data

Gross and Clark (1975) reported this dataset, which contains the relief times (in minutes) experienced by 20 patients after receiving an analgesic. The observations are:

1.1, 1.4, 1.3, 1.7, 1.9, 1.8, 1.6, 2.2, 1.7, 2.7, 4.1, 1.8, 1.5, 1.2, 1.4, 3.0, 1.7, 2.3, 1.6, 2.0.

2.7.5. Accelerated Life Test Data

Lawless (2003) presented this dataset, which records the failure times (in minutes) of 15 electronic components subjected to an accelerated life test. The observations are:

1.4, 5.1, 6.3, 10.8, 12.1, 18.5, 19.7, 22.2, 23.0, 30.6, 37.3, 46.3, 53.9, 59.8, 66.2.

2.7.6. Yarn Strength Data

This dataset, reported by Lawless (2003), contains the number of cycles to failure for 25 yarn specimens, each measuring 100 cm, tested at a particular strain level. The observations are:

15, 20, 38, 42, 61, 76, 86, 98, 121, 146, 149, 157, 175, 176, 180, 180, 198, 220, 224, 251, 264, 282, 321, 325, 653.

2.7.7. Strength of Glass Data

This dataset was reported by Fuller et al. and consists of the strength measurements of aircraft window glass. The dataset contains 31 observations:

18.83, 20.80, 21.66, 23.03, 23.23, 24.05, 24.321, 25.50, 25.50, 25.80, 26.69, 26.77, 26.78, 27.05, 27.67, 29.90, 31.11, 33.20, 33.73, 33.80, 33.90, 34.76, 35.75, 35.91, 36.98, 37.08, 37.09, 39.58, 44.05, 45.29, 45.381.

2.7.8. Vinyl Chloride Data

This dataset, used by Bhaumik et al. (2009), contains concentrations of vinyl chloride (mg/L) obtained from clean upgradient monitoring wells. The observations are:

5.1, 1.2, 1.3, 0.6, 0.5, 2.4, 0.5, 1.1, 8.0, 0.8, 0.4, 0.6, 0.9, 0.4, 2.0, 0.5, 5.3, 3.2, 2.7, 2.9, 2.5, 2.3, 1.0, 0.2, 0.1, 0.1, 1.8, 0.9, 2.0, 4.0, 6.8, 1.2, 0.4, 0.2.

2.7.9. RT+CT Survival Data

This dataset, reported by Efron (1988), consists of the survival times of patients diagnosed with head and neck cancer who received combined radiotherapy and chemotherapy (RT+CT). The dataset contains 44 observations:

12.20, 23.56, 23.74, 25.87, 31.98, 37.00, 41.35, 47.38, 55.46, 58.36, 63.47, 68.46, 74.47, 78.26, 81.43, 84.00, 92.00, 94.00, 110.00, 112.00, 119.00, 127.00, 130.00, 133.00, 140.00, 146.00, 155.00, 159.00, 173.00, 179.00, 194.00, 195.00, 209.00, 249.00, 281.00, 319.00, 339.00, 432.00, 469.00, 519.00, 633.00, 725.00, 817.00, 1776.00.

2.7.10. Virulent Tubercle Bacilli Data

This dataset was reported by Bjerkedal (1960) and contains the survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli. The observations are:

10, 33, 44, 56, 59, 72, 74, 77, 92, 93, 96, 100, 100, 102, 105, 107, 107, 108, 108, 108, 109, 112, 113, 115, 116, 120, 121, 122, 122, 124, 130, 134, 136, 139, 144, 146, 153, 159, 160, 163, 163, 168, 171, 172, 176, 183, 195, 196, 197, 202, 213, 215, 216, 222, 230, 231, 240, 245, 251, 253, 254, 254, 278, 293, 327, 342, 347, 361, 402, 432, 458, 555.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
Sadiq *et al.* 2026 **DOI**: 10.XXXXX/jossda.2026.1.1.016

3. Results and Discussions

This section presents the results obtained from fitting the Gamma distribution to both real-life and simulated datasets. The analyses include descriptive statistics, parameter estimation using the Maximum Likelihood Estimation (MLE) method, model adequacy based on information criteria, and goodness-of-fit tests.

Table 1. Descriptive analysis of the real-life datasets

Type of Data	N	Mean	Median	Mode	Variance	Skewness	Kurtosis
Ball bearings	23	72.230	67.8	50	1404.783	0.942	0.489
Air conditioning	29	89.586	23	14.11	29549.11	3.854	15.463
Boeing	15	121.27	57	29	23798.78	1.523	0.825
Analgesic	20	1.9	1.7	1.75	0.496	1.720	2.924
Accelerated life test	15	27.547	22.2	15	431.117	0.566	-0.940
Strain level	25	178.32	175	180	17902.64	1.730	4.360
Straight of glass	31	30.813	29.9	27.5	52.631	0.405	-0.715
Vinyl Chloride	34	1.879	1.15	0.5	3.813	1.604	2.005
RT+CT	44	223.477	128.5	100	93286.41	3.384	13.560
Virulent-Tubercle-Bacilli	72	176.819	149.5	108	10702.91	1.342	1.991

Table 1 presents the descriptive statistics for the ten real-life datasets. All datasets consist of positive continuous observations, making them suitable candidates for Gamma distribution modelling. The descriptive measures indicate that the datasets differ considerably in terms of central tendency, variability, skewness, and kurtosis.

The ball bearings, accelerated life test, strength of glass, and virulent tubercle bacilli datasets exhibit moderate positive skewness, suggesting longer right tails than would be expected under a symmetric distribution. The air conditioning and RT+CT datasets display the largest skewness values (3.854 and 3.384, respectively), indicating the presence of extreme observations and pronounced right-skewed distributions.

The kurtosis values further reveal differences in tail behaviour. The accelerated life test and strength of glass datasets have negative kurtosis, indicating relatively flatter (platykurtic) distributions compared with the normal distribution. Conversely, the air conditioning, RT+CT, strain level, analgesic, vinyl chloride, and virulent tubercle bacilli datasets exhibit positive kurtosis, reflecting heavier tails and a greater likelihood of extreme observations. In particular, the air conditioning and RT+CT datasets possess exceptionally high kurtosis values, confirming the presence of substantial outliers.

Generally, the descriptive statistics indicate that most of the datasets are positively skewed with varying degrees of kurtosis, supporting the use of the Gamma distribution for modelling positive continuous data. However, the highly skewed datasets may present greater challenges for achieving an adequate model fit.

Table 2 presents the MLEs of the Gamma distribution parameters together with the corresponding log-likelihood (LL), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Hannan-Quinn Information Criterion (HQIC), and Consistent Akaike Information Criterion (CAIC) for each dataset.

Table 2. Criteria and goodness-of-fit measures, maximum likelihood estimates

Type of data/ model	LL	CAIC	HQIC	AIC	BIC	α	β
Ball bearings	115.579	235.758	235.729	235.158	237.429	9.741	6.958
Air conditioning	436.371	877.204	877.598	876.742	879.477	4.810	6.081
Boeing	149.289	303.579	302.563	302.579	303.995	9.883	7.313
Analgesic	18.514	41.733	41.416	41.027	43.019	0.233	8.897
Accelerated life test	66.639	138.277	137.262	137.277	138.693	9.316	2.229
Strain level	243.626	491.798	491.927	491.251	493.689	9.997	9.998
Straight of glass	107.299	219.027	219.533	218.598	221.466	3.269	9.341
Vinyl Chloride	55.416	115.219	115.873	114.832	117.884	1.778	1.049
RT+CT	747.945	1500.183	1501.214	1499.89	1503.459	8.990	8.038
Virulent-Tubercle-Bacilli	611.417	1227.007	1228.646	1226.833	1231.387	9.990	9.971

The estimated shape α and scale θ parameters vary across datasets, reflecting differences in the underlying distributional characteristics of the observed data. The log-likelihood values indicate the degree to which the Gamma model explains each dataset, whereas the information criteria provide measures of model adequacy after accounting for model complexity.

Because these datasets differ in both sample size and scale, the information criteria should not be compared across datasets. Instead, they are intended for comparing competing probability models fitted to the same dataset. Consequently, the values reported in Table 2 primarily serve as baseline measures of the adequacy of the Gamma model for each dataset rather than as evidence that one dataset is better fitted than another.

Table 3. Test Statistic

Type of data/ model	W	A	One-sample K-S test	
			D	P-value
Ball bearings	0.039	0.218	0.148	0.645
Air conditioning	0.173	1.111	0.369	0.001
Boeing	0.184	1.058	0.378	0.028
Analgesic	0.105	0.623	0.268	0.114
Accelerated life test	0.022	0.176	0.215	0.434
Strain level	0.061	0.412	0.556	0.003
Straight of glass	0.078	0.426	0.142	0.564
Vinyl Chloride	0.046	0.298	0.093	0.930
RT+CT	0.139	0.988	0.509	<0.000
Virulent-Tubercle-Bacilli	0.103	0.623	0.442	<0.000

Table 3 presents the results of the Cramér-von Mises (W), Anderson-Darling (A), and Kolmogorov-Smirnov (KS) goodness-of-fit tests for the Gamma distribution.

Using the Kolmogorov-Smirnov test at the 5% significance level, the Gamma distribution provides an adequate fit for the ball bearings, analgesic, accelerated life test, strength of glass, and vinyl chloride datasets, as their p-values exceed 0.05. Consequently,

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX

DOI: 10.XXXXX/jossda.2026.1.1.016

there is insufficient evidence to reject the null hypothesis that these datasets follow a Gamma distribution.

In contrast, the air conditioning, Boeing, strain level, RT+CT, and virulent tubercle bacilli datasets have p-values below 0.05, indicating statistically significant departures from the Gamma distribution. Therefore, the Gamma model does not adequately describe these datasets.

The Anderson-Darling and Cramér-von Mises statistics provide similar evidence, with smaller values indicating better agreement between the observed and fitted distributions. Generally, the goodness-of-fit tests suggest that the Gamma distribution is appropriate for several, but not all, of the real-life datasets considered.

Table 4. The goodness-of-fit measures of the simulated dataset

Simulated data	LL	CAIC	HQIC	AIC	BIC	α	β
100	74.3695	152.863	154.848	152.739	157.949	0.437	2.170
200	162.785	329.630	332.239	329.569	336.166	0.428	2.284
300	257.287	518.614	521.538	518.573	525.981	0.507	1.922
400	350.574	705.179	708.31	705.149	713.132	0.527	1.865
500	435.423	874.870	878.153	874.845	883.274	0.516	1.873
1000	115.579	235.758	235.729	235.158	237.429	9.741	6.958

Table 4 presents the MLE and information criteria obtained from simulated datasets with sample sizes of 100, 200, 300, 400, 500, and 1000.

As the sample size increases, the parameter estimates become more stable, reflecting the desirable large-sample properties of the Maximum Likelihood Estimator. The increasing log-likelihood values are expected because larger datasets contain more information. Similarly, the information criteria increase with sample size since they depend on the value of the log-likelihood and the number of observations.

Therefore, the information criteria should not be interpreted as indicating that smaller sample sizes provide better model fits. Rather, the simulation results demonstrate that the Gamma distribution can be estimated reliably across all sample sizes, with estimation accuracy expected to improve as the sample size increases.

Table 5. Test Statistic

Sample size	W	A	One-sample K-S test	
			D	P-value
100	0.053	0.329	0.058	0.888
200	0.044	0.366	0.062	0.437
300	0.071	0.474	0.039	0.745
400	0.044	0.311	0.023	0.987
500	0.061	0.425	0.029	0.805
1000	0.061	0.403	0.026	0.504

Table 5 presents the goodness-of-fit statistics for the simulated datasets. The Kolmogorov-Smirnov test produced p-values greater than 0.05 for all sample sizes (100, 200, 300, 400, 500, and 1000). Therefore, the null hypothesis that the simulated data follow a Gamma distribution cannot be rejected.

The Cramér-von Mises and Anderson–Darling statistics are also relatively small across all sample sizes, indicating close agreement between the simulated observations and the fitted Gamma distribution.

These findings confirm that the Gamma distribution adequately models the simulated datasets, which is expected because the data were generated from the Gamma distribution. The results also demonstrate the consistency of the Maximum Likelihood Estimation procedure and provide empirical support for the suitability of the Gamma distribution in modelling positive continuous data.

4. Conclusion

This study evaluated the applicability of the Gamma distribution for modelling positive continuous data using both simulated and real-life datasets. The parameters of the Gamma distribution were estimated using the Maximum Likelihood Estimation (MLE) method, while the adequacy of the fitted model was assessed using descriptive statistics, information criteria, and goodness-of-fit tests.

The descriptive analyses showed that the real-life datasets generally exhibited positive skewness and varying degrees of kurtosis, confirming that they possess characteristics commonly associated with Gamma-distributed data. However, the extent of skewness and tail behaviour varied considerably across the datasets, indicating that the suitability of the Gamma distribution depends on the underlying characteristics of the data.

The goodness-of-fit analyses revealed that the Gamma distribution provided an adequate fit for several of the real-life datasets, particularly the ball bearings, analgesic, accelerated life test, strength of glass, and vinyl chloride datasets. Conversely, the air conditioning, Boeing, strain level, RT-CT, and virulent tubercle bacilli datasets exhibited significant departures from the Gamma distribution according to the goodness-of-fit tests, suggesting that alternative probability distributions may provide better representations of these data.

The simulation study demonstrated that the Maximum Likelihood Estimation method produced stable parameter estimates across different sample sizes. Furthermore, the goodness-of-fit tests confirmed that the Gamma distribution adequately modelled the simulated datasets, supporting the consistency and reliability of the estimation procedure. As expected, estimation performance improved with increasing sample size, illustrating the desirable large-sample properties of the maximum likelihood estimator.

Generally, the findings indicate that the Gamma distribution is an effective model for many types of positive continuous data, particularly those exhibiting moderate positive skewness. Nevertheless, no single probability distribution is universally appropriate for all datasets. Consequently, goodness-of-fit assessment should always accompany model fitting to ensure that the selected distribution adequately represents the observed data. The results of this study contribute to the understanding of the practical applicability of the Gamma distribution and provide useful guidance for researchers and practitioners working with lifetime, reliability, biomedical, and engineering data.

References

Casella, G., & Berger, R. L. (2002). Statistical Inference (2nd ed.). Duxbury.
Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Duxbury Press.

Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.016

Chen and Wang (2019). Application of the Gamma Distribution in Reliability Analysis. Journal of Quality and Reliability

Chen et al. (2021). Application of the Gamma Distribution in Financial Risk Modelling. Journal of Finance and Economics

Chen, J. et al. (2022). Maximum Product Spacing Estimation for Gamma Distribution with Censored Data. Journal of Statistics in Medicine

Chen, L. et al. (2023). Efficient maximum Likelihood Estimation for the Gamma Distribution. Journal of Computational Statistics & Data Analysis

Cheng, R. C. H., & Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society: Series B (Methodological), 45*(3), 394–403.

Cheng, R. C. H., & Amin, N. A. K. (1983). Journal of the Royal Statistical Society, Series B, 45(3), 394–403.

Creswell, J. W., & Creswell, J. D. (2018). Research Design (5th ed.). Sage.

Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman & Hall.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall.

Eric U., Oti M.O.O., Francis C.E. (2021). A Study of Properties and Applications of Gamma Distribution. African Journal of Mathematics and Statistics Studies 4(2), 52-65. DOI: 10.52589/AJMSSMR0DQ1DG

Gupta and Kumar (2020). Gamma Distribution and its Role in Queuing Theory. Journal of Operations Research

Johnson and Thompson (2020). Estimating Shape Parameter in the Gamma Distribution. Journal of Applied Mathematics

Johnson et al. (2018). The Gamma Distribution: Properties and Estimation Techniques. Journal os Statistical Methods

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). Continuous Univariate Distributions, Volume 1 (2nd ed.). Wiley.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions* (Vol. 1, 2nd ed.). John Wiley & Sons.

Johnson, M. et al. (2021). Improved Maximum Likelihood Estimation for the Gamma Distribution Using EM Algorithm. Journal of Statistical Computation and Simulation

Lawless, J. F. (2003). Statistical Models and Methods for Lifetime Data (2nd ed.). Wiley.

Lawless, J. F. (2003). *Statistical models and methods for lifetime data* (2nd ed.). John Wiley & Sons.

Lee and Johnson (2023). Gamma Distribution in Modelling Lifetime Data. Journal of Quality Technology

Lee, H. et al. (2022). Maximum Product Spacing Estimation of the Gamma Distribution: Theory and Applications. Journal of Statistical Inference

Lehmann, E. L., & Casella, G. (1998). Theory of Point Estimation (2nd ed.). Springer.

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation* (2nd ed.). Springer.

Li and Zhang (2021). Gamma Distribution-Based Models for Risk Assessment. Journal of Risk Analysis

Pearson, K. (1893). "Contributions to the Mathematical Theory of Evolution." Proceedings of the Royal Society of London, 54, 329–333.

Pearson, K. (1893). Contributions to the mathematical theory of evolution. *Proceedings of the Royal Society of London*, *54*, 329–333.

Ranneby, B. (1984). "The Maximum Spacing Method." Scandinavian Journal of Statistics, 11(2), 93–112.

Ranneby, B. (1984). The maximum spacing method: An estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics, 11*(2), 93–112.

Rodriguez et al. (2022). Gamma Distribution-Based Models for Rainfall Intensity Analysis. Journal of Hydrology

Rodriguez, A. et al. (2022). Maximum Likelihood Estimation of Gamma Distribution Parameters Using Genetic Algorithms. Journal of Applied Statistics

Sen, S., Maiti, S. S. and Chandra, N. (2016). The xgamma Distribution: statistical properties and application. Journal of Modern Applied Statistical Methods, 15(1), 774-788.

Smith et al. (2019). A comprehensive review of Gamma Distribution Properties. Journal of Statistical Analysis.

Smith, J. et al. (2022). Maximum likelihood Estimation of the Gamma Distribution: A Comparative Study. Journal of Statistical Modelling

Wang et al. (2022). Gamma-mixture models for Data Clustering. Journal of Machine Learning

Wang, Q. et al. (2021). Efficient Maximum Product Spacing Estimation for Gamma Distribution Using Algorithms. Journal of Computational Statistics & Data Analysis

Zhang, L. et al. (2023). Comparison of Maximum Product Spacing and Maximum Likelihood Estimation for the Gamma Distribution. Journal of Communication in Statistics