ADVANCED STATISTICAL APPROACH TO FAILURE DATA WITH GAMMA AND WEIBULL DISTRIBUTIONS

Vijayan S, Kavitha S*

•

Department of Statistics, Periyar University, Salem-11, Tamil Nadu, India.
vijaystatistician1210@gmail.com, pustatkavitha@gmail.com

Abstract

This paper aims to systematically investigate the utility of the Gamma and Weibull distributions, focusing on their application to biomedical datasets and clarifying their mathematical and statistical properties. By analyzing lifetime data across various disciplines, the research emphasizes the effectiveness and flexibility of these distributions in capturing the complexities of biomedical data. It underscores the importance of parameters such as standard error, log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) in value estimation. The findings suggest that both distributions provide valuable insights into the underlying data, with practical implications for reliability engineering and failure analysis. Moreover, the study demonstrates that the Weibull distribution offers a better fit to the given data than the Gamma distribution due to its adaptability, which yields superior results. A key contribution of this study is the proposal of a model based on estimating the Conditional Weibull distribution for feature parameters, which accurately predicts a finite mixture of two-parameter Weibull distributions initially verified on datasets.

Keywords: Gamma distribution, Weibull distribution, Probability density function, Cumulative density function, Akaike information criterion, Bayesian information criterion, Biomedical.

I. Introduction

The Weibull distribution is well-known for its application in reliability engineering and failure analysis. It is extensively used to model biological sciences, weather forecasting, and hydrology data. However, it may not be appropriate for specific applications, especially once hazard rates exhibit bathtub or bimodal shapes. Researchers have developed various modifications and extensions of the Weibull distribution to overcome these limitations and accommodate a broader range of data types [3], introducing the additive Weibull distribution. Using the Generalized Gamma distribution in survival study for breast cancer patients [1]. The generalized Weibull model extends the traditional distribution by incorporating additional parameters, offering greater flexibility in modeling survival data [13]. The main objective is to compare the effectiveness of the Gamma and Weibull distributions in modeling lifetime data across various fields, such as multiple fields used to evaluate their goodness of fit [18]. Numerous studies have also applied Gamma distribution to model wealth inequality [7]. These tests are relevant to various fields, including failure time models and survival

studies, anywhere the failure frequency is constant, and increasing pertinent tests to fields like failure time models and survival analysis are crucial for determining if the failure rate is steady, growing, or falling is vital [5]. This study significantly enhances the understanding and application of the inverse-generalized Weibull distribution, providing more accurate tools for reliability engineering failure analysis [2]. Due to its versatility in the model's lifetime data, the Weibull distribution has long been stable in reliability engineering and various other fields. However, traditional Weibull distribution sometimes fails to capture complex hazard rate shapes truthfully. Researchers have developed multiple modifications and extensions to address these limits [14]. The proposed novel method enhances parameter estimation for finite Weibull distributions, educating the precision and reliability through cost-effective continuing asset management strategies [8]. The proposed technique is evaluated through comprehensive numerical analysis, demonstrating reliability through its effectiveness in outperforming existing advanced methods for handling unfair data [12]. The primary goal is to propose and assess the beta-Weibull distribution efficiency in modelling survival data. By integrating the exponentiated Weibull distributions, this new family modification is a data analysis [4].

The Weibull distribution enjoys widespread use in aerospace, microchip technology, materials, and automotive industries; it is vital to understand components' reliability and failure rates [11]. In today's context of model building and real-life data analysis, numerous lifetime distributions are utilized, with sources [15] and [16] providing foundational ideas and concepts. The Weibull distributions stand out for their flexibility and extensive applicability in modeling diverse data types, particularly excelling in reliability engineering and failure analysis [20]. The parameters of the Weibull-gamma distribution are estimated using the maximum likelihood method [6]. This research is inspired by the recognition that outdated distributions often fail to accommodate the complexities of biomedical datasets. Encompassing, the seeks to contribute to developing two distributions that improve and adapt the intricacies of real-life data. The objective is to methodically analyze and compare these distributions, focusing on their effectiveness in modeling biomedical datasets. Critical parameters such as Estimated Values, Standard Error, Log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) are examined to illuminate these models' statistical characteristics and fitting quality. The ultimate goal is to determine which distribution better captures the nuances of the dataset, thereby advancing the understanding and application of these models in biomedical research.

II. Methods

2.1 Derivation of Gamma and Weibull Distributions

In the section, a new comparison model cumulative distribution (CDF) and probability density function (PDF) of the Gamma and Weibull Distribution family are given as

$$\mathrm{f}(x:\alpha,\beta) = f(x) = \begin{cases} \frac{-x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{1}$$

where $\alpha > 0$ is the scale parameter, $\beta > 0$ is the shape parameter, and $\Gamma(\alpha)$ is the Gamma function. The cumulative distribution function is

$$F(x;\ \alpha,\beta) = f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)}\gamma\left(\alpha,\frac{x}{\beta}\right), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2}$$

Where $\gamma(\alpha, x)$ is the lower incomplete Gamma function.

$$F(x;\ \alpha, k) = f(x) = \begin{cases} 1 - e^{-(x/\alpha)\beta}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3}$$

Where $\alpha > 0$ is the scale parameter, and $\beta > 0$ is the shape parameter.
The cumulative distribution function is given by

$$f(x;\ \alpha, k) = f(x) = \begin{cases} \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{k-1} e^{-(x/\alpha)^{\beta}}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{4}$$

The CDF provides the probability that a random variable X is less than or equal to a specific value x. At the same time, the PDF describes the relative likelihood of the random variable taking on a particular value.

The Bayesian Information criterion is

$$BIC = kln(n) - 2ln(L) \tag{5}$$

k is the number of parameters in the model, n is the number of observations, and L is the maximum likelihood of the model.
The Akaike Information Criterion is given by

$$AIC = 2k - 2ln(L) \tag{6}$$

k is the number of estimated parameters in the model, and L is its maximum likelihood function.

The gamma distribution model was applied to dataset observations to gain insights into its statistical characteristics. The model's shape and scale parameters were accurately estimated, and the negative log-likelihood value indicated a good fit for the data. These statistical measures suggest that the Gamma distribution effectively captures the underlying data distribution, providing a solid foundation for further analysis and interpretation.

The model provided significant insights into the statistical properties of the analyzed dataset. Fitted to observations, the model's shape and scale parameters were estimated with corresponding standard errors, reflecting its fitting quality. These parameters and statistical measures demonstrate the Weibull model's effectiveness in capturing the underlying distribution characteristics, providing a robust tool for further analysis and clarification.

Overall, the Weibull distribution fits the dataset better than the Gamma distribution. The Wei bull model's closer AIC and BIC values suggest consistency and reliability in model fit evaluation. For various combinations of α and β, we generated sample sizes from the Gamma and Weibull models, specifically for the parameter values $\alpha = 2.0$, $\beta = 0.5$, and $\alpha = 1.5$, $\beta = 1$. The value decreases as the sample size increases, as shown in Table 1. Table 1 illustrates that our Gamma and Weibull models better fit the dataset than the KME and KM-IW (α, β) distributions. The dataset represents survival times in days from a two-arm clinical trial, referenced by sources [9] and [17], consisting of heart failure times in days for 299 patients.

III. Results

In this section, we conduct a comparative study to assess the Maximum Likelihood Estimation (MLE) performance for Gamma and Weibull distributions. We generate n = 299 samples drawn from those distributions using the quantile function. Subsequently, we calculate the MLEs, loglikelihood, standard error, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC).

Table 1: *The model's performance Some values of parameters requirements built on the dataset*

Model	Parameter	Estimated values	Standard Error	Loglikelihood	AIC	BIC
Gamma	α	2.0509	0.1560	-658.7781	1321.556	1328.957
	β	0.5442	0.0468			
Weibull	α	1.5902	0.0716	-428.4005	860.8009	868.2018
	β	1.9780	0.0757			

Table 1 shows the outcomes corresponding to the Gamma distribution by utilizing it to model Biomedical datasets. The parameters indicate that the Gamma distribution, with its shape and scale, captures the data distribution with a specific level of skewness and spread. The AIC and BIC values support the model's fit, though the Weibull distribution, with its parameters, shows a potentially better fit given its lower AIC and BIC values. This suggests the Weibull model might be more effective at capturing the underlying distribution characteristics of the dataset.

3. 1 Application

We compare its performance in providing a robust parametric fit to the datasets with that of the Weibull distribution. Metrics such as the log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) are employed for this comparison. The loglikelihood, AIC, and BIC values are computed for the proposed Weibull model for comparison to discern the most suitable model. The model exhibiting the lowest loglikelihood, AIC, and BIC values is deemed the most appropriate match for the provided datasets. The R software is employed for this analytical endeavor, simplifying the necessary calculations and comparisons.

The dataset comprises the summation of time (in days) measurements from two hundred twenty-nine (299) respondents at the heart failure patients. It consists of the following values:
4, 6, 7, 7, 8, 10, 10, 10, 10, 10, 11, 11, 12, 13, 14, 14, 15, 15, 16, 20, 20, 22, 23, 23, 24, 26, 26, 27, 28, 28, 29, 29, 30, 30, 30, 30, 30, 31, 32, 33, 33, 33, 35, 38, 40, 41, 42, 43, 44, 45, 50, 54, 54, 55, 59, 60, 60, 60, 61, 63, 64, 65, 66, 67, 68, 71, 72, 73, 74, 74, 74, 74, 75, 76, 77, 78, 78, 79, 79, 79, 79, 80, 80, 82, 82, 83, 83, 83, 85, 85, 86, 87, 87, 87, 87, 87, 88, 88, 88, 88, 88, 90, 90, 90, 90, 91, 91, 94, 94, 94, 95, 95, 95, 95, 95, 96, 97, 100, 100, 104, 104, 105, 106, 107, 107, 107, 107, 107, 107, 108, 108, 108, 109, 109, 109, 110, 111, 112, 113, 115, 115, 117, 118, 119, 120, 120, 120, 120, 121, 121, 121, 121, 123, 126, 130, 134, 135, 140, 145, 145, 146, 146, 146, 146, 146, 147, 147, 147, 147, 148, 150, 154, 162, 170, 171, 172, 174, 174, 174, 175, 180, 185, 186, 186, 186, 186, 186, 186, 187, 187, 187, 187, 187, 187, 187, 188, 192, 193, 193, 194, 195, 196, 196, 197, 197, 198, 200, 201, 201, 205, 205, 205, 206, 207, 207, 207, 208, 209, 209, 209, 209, 209, 210, 210, 211, 212, 212, 213, 213, 213, 214, 214, 214, 214, 214, 215, 215, 215, 215, 216, 220, 230, 230, 233, 233, 235, 237, 237, 240, 241,

244, 244, 244, 244, 244, 245, 245, 245, 245, 245, 246, 246, 246, 247, 250, 250, 250, 250, 250, 250, 250, 256, 256, 257, 258, 270, 270, 271, 278, 280, 285.

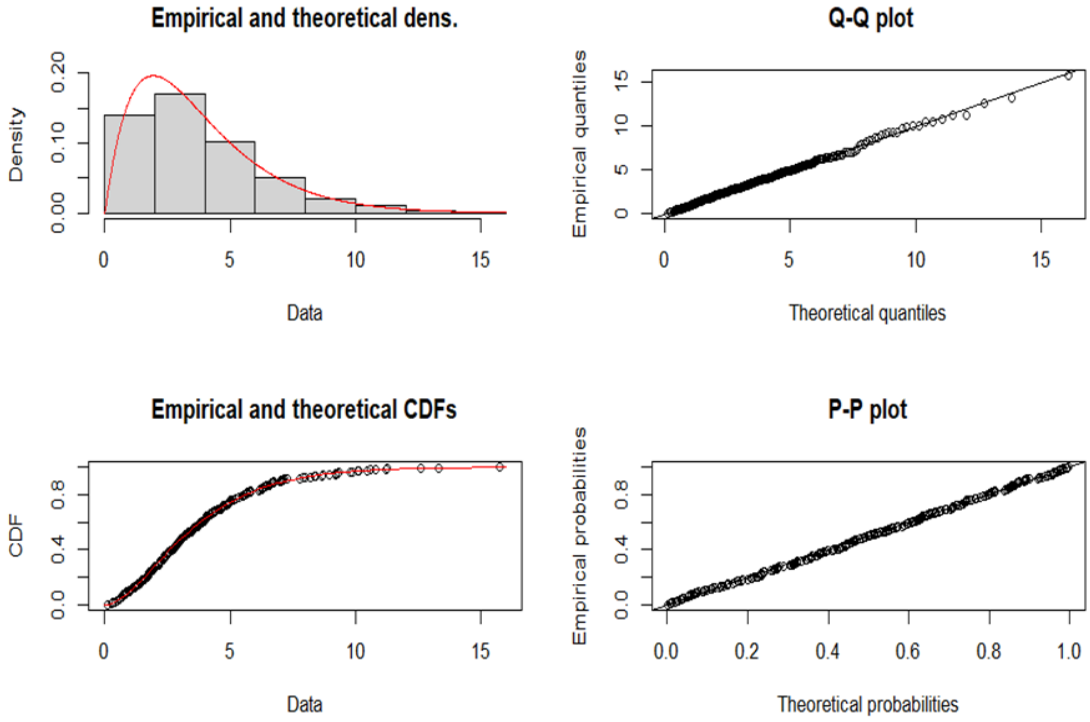

Figure 1: *Gamma distribution Density plots for the datasets.*

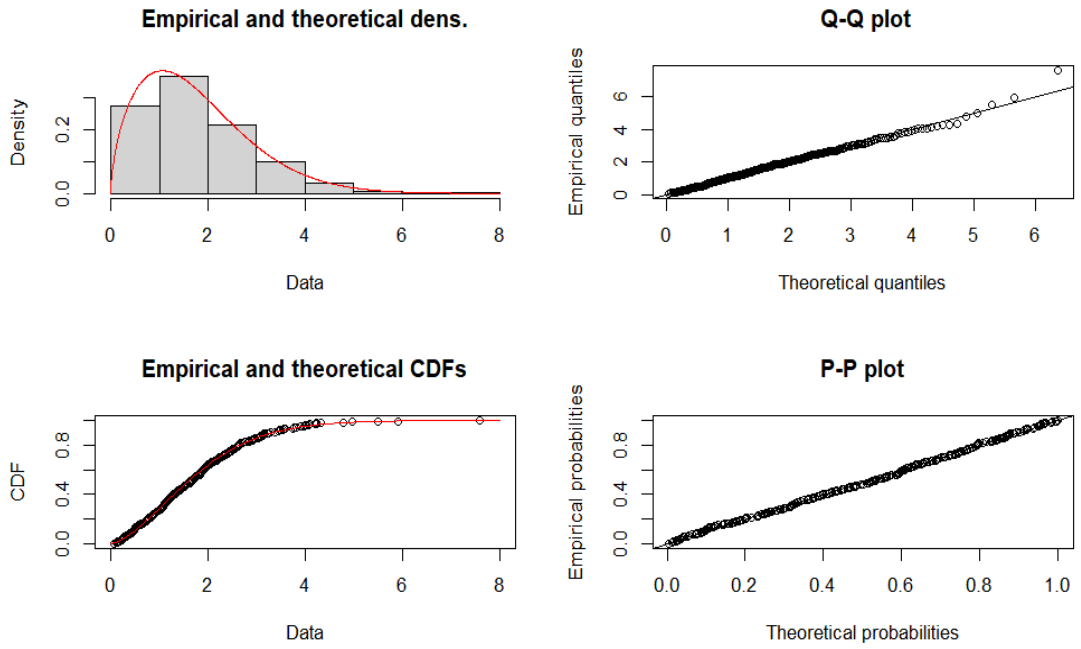

Figure 2: *Weibull distribution Density plots for the datasets.*

Table 1 presents the planned model's superior ability to effectively fit the highly skewed datasets compared to the competing models, as specified by the evaluation metrics employed. Figures 1 and 2 show that the proposed mode fits the data set adequately.

IV. Conclusion

This paper comprehensively compares the Gamma and Weibull distributions for modeling lifetime data. By incorporating key parameters, we improve the flexibility of these distributions, enabling them to capture the complexities of biomedical datasets effectively. Using accurate data, we derived precise closed-form expressions for the probability density function (PDF), cumulative density function (CDF), standard error, log-likelihood function, and parameter estimation. Our analysis demonstrated the practical relevance of these models through Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. This enhances our ability to accurately model and interpret biomedical data, providing a robust tool for future research and analysis.

References

[1] Abadi, A. Amanpour F. Bajdik, C. and Yavari, P. (2012). Breast cancer survival analysis: Applying the generalized gamma distribution under different conditions of the proportional hazards and accelerated failure time assumptions. *International journal of preventive medicine,* 3(9): 644.

[2] Amiradi, A. Jamkhaneh, E.B. and Deiri, E. (2021). A comparison of estimation methods for reliability function of inverses generalized Weibull distribution under new loss function. *Journal of Statistical Computation and Simulation,* 91(13): 2595-2622.

[3] Alshanbari, H.M. Ahamad, Z. El-Bagoury, A. A. A. H. Odhah, O. H. and Rao, G. S. (2024). A new modification of the Weibull distribution: model theory and analyzing engineering data sets. *Symmetry,* 16(5): 611.

[4] Al-Malki, S. J. (2014). Statistical analysis of lifetime data using newly modified Weibull distributions. *The University of Manchester (United Kingdom).*

[5] Basiru, S. O. Khalaf, A.A. Isa, A.M. and Kaigama, A. (2024). On the modeling of biomedical data with exponentiated Gompertz inverse Rayleigh distribution. *Reliability: Theory and Applications,* 19(3 (79): 460-475.

[6] Oluyede, B. Pu. Makubate, B. and Qiu, Y. (2018). The Weibull-Gamma distribution properties and applications. *Entropy,* 47(1): 438.

[7] Chakraborti, A. and Patriarca, M. (2008). Gamma distribution and wealth inequality. *Pramana,* 71: 233-243.

[8] Dong, M. and Nassif, A. B. (2018). Combining modified Weibull distribution models for power system reliability forecast. *IEEE Transactions on Power Systems,* 34(2): 1610-1619.

[9] Efron, B. (1998). Logistic regression, survival analysis, and the Kaplan – Meier curve. *Journal of the American Statistical Association,* 83(402): 414-425.

[10] Elmahdy, E. E. and About Houn, A. W. (2013). A new approach for parameter estimation of finite Weibull mixture distributions for reliability modeling. *Applied Mathematical Modeling.* 37(4): 1800-1810.

[11] Hallinan, Jr, A. J. (1993). A review of the Weibull distribution. *Journal of Quality Technology.* 25(2): 85-93.

[12] Kalmov, F. and Denisov, D. (2020). Gamma distribution-based sampling for imbalanced data. *Knowledge-Based Systems, 207-106368.*

[13] Klakattawi, H. S. (2002). Survival analysis of cancer patients using a new extended Weibull distribution. *Plos one,* 17(2): e0264229.

[14] Lai, C.CD. Xie, M. and Murthy, D. N. P. (2003). A modified Weibull distribution. *IEEE*

Transactions on Reliability. 52(1): 33-37.

[15] Lawless, J. F. (2011). Statistical models and methods for lifetime data. *John Wiley and Sons.*

[16] Lee, E. T. (1986). Statistical methods for survival data analysis. *IEEE Transactions on Reliability,* 35(1): 123-123.

[17] Mudholkar, G. S. Srivastava, D. K. and Kollia, G. D. (1996). A generalization of the Weibull distribution with application to the survival data analysis. *Journal of the American Statistical Association,* 91(436): 1575-1583.

[18] Shanker, R. Shukla, K. K. Shanker, R. and Leonida, T. S. (2016). On modeling of lifetime data using two-parameter gamma and weibull distributions. *Biometrics and Biostatistics International Journal,* 4(5): 1-6.

[19] Wahed, A. S. Luong, T. M. and Jeong, J. H. (2009). A new generalization of Weibull distribution with application to a breast cancer data set. *Statistics in medicine.* 28(16): 2077-2094.

[20] Weibull, W. (1951). A statistical distribution function of broad applicability. *Journal of Applied Mechanics.*