Comparative Study of Life Distribution Models and Their Applications in Reliability Data Analysis

Sirajuddeen. P. K., Assistant Professor, Department of Statistics, Govt. Brennen College, Dharmadam, Thalassery, Kerala - 670106

Abstract

This research compares four types of life distribution models, Exponential, Weibull, Log-Normal and Gamma, using examples from reliability data analysis. Different sample sizes and censoring levels were added to simulated data and this was analyzed with true failure time data to check how well the models worked. We evaluated the accuracy and appropriateness of the models with Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood values, Kolmogorov–Smirnov and the Anderson–Darling goodness-of-fit tests. The analysis indicated that the Weibull model fit best, since it could handle any type of hazard rate change. Even though the Log-Normal and Gamma models suited some data sets well, the Exponential model couldn't be used much due to its rigid assumptions. The findings indicate that making the correct model choice, according to actual data features, is vital for achieving reliable and accurate failure predictions.

Keywords: Life Distribution Models; Reliability Analysis; Weibull Distribution; Failure Time Data; Model Comparison; Hazard Function; AIC; Censoring; Survival Analysis; Goodness-of-Fit.

1. INTRODUCTION

System and product designers in industries that value safety, performance and long product life rely heavily on reliability engineering. If engineers and researchers can estimate component failure over time, they can choose when to maintain items, assess risks, make warranty decisions and assure quality. A main reason for predictions is life distribution models which help describe the lives of components and systems by using real data.

Life distribution models show a possible way that products might fail during operation. They let us understand key aspects of reliability metrics like R(t)R(t)R(t), F(t)F(t)F(t) and h(t)h(t)h(t). In reliability analysis, the most used models are the Exponential, Weibull, Log-Normal, Gamma and Inverse Gaussian distributions. A model's characteristics and rules make it best suited to certain kinds of reliability data. For example, an Exponential model is based on constant failure rate, so individuals can use the Weibull model which allows for options for increasing, decreasing or constant hazard rates. In the same way, skewed failure time data is best modeled by Log-Normal or Gamma distributions, while the Inverse Gaussian is useful for systems that need a wear-out or fatigue-related model.

Picking the right life distribution model can be a tough job in actual reliability studies. In many cases, reliability data will have observations missing the exact failure time of a component because the test period ended or for other external reasons, a situation known as right-censoring. At the same time, small samples, PhD and agency-wide censoring and complex reasons for failures can make things even more difficult when it comes to modeling. As a result, examining life distribution models in both safe and realistic environments reveals their advantages and drawbacks.

2. LITERATURE REVIEW

Bain (2017) played a pivotal role in laying the theoretical groundwork for reliability analysis. His comprehensive work offered detailed methods for analyzing lifetime data using both classical and Bayesian approaches. Bain emphasized the importance of understanding the statistical behavior of components and systems under stress and over time, highlighting core principles such as failure distribution modeling, hazard function estimation, and life-testing experiments. His focus on reliability functions and their practical interpretation made his work especially valuable for both academic researchers and industrial practitioners.

Crowder (2017) provided an extensive treatment of statistical tools for analyzing reliability data, particularly under conditions of censoring and truncation, which are common in practical life-testing scenarios. His exploration of non-parametric and parametric estimation techniques, including the use of maximum likelihood estimation, served as a guide for assessing system performance in both complete and incomplete datasets. Crowder's work was also notable for its treatment of recurrent events and repairable systems, making it highly relevant in fields like mechanical engineering and biomedical device monitoring.

Lawless (2011) contributed significantly to the modeling of lifetime data through his landmark text, where he examined a wide range of probability models, including the exponential, Weibull, log-normal, and gamma distributions. His focus on the assumptions behind each model and their applicability in different contexts enhanced the understanding of reliability estimation under varying conditions. Lawless emphasized the importance of model selection based on empirical data characteristics and provided in-depth discussion on regression models for lifetime data, such as the Cox proportional hazards model, thereby bridging the gap between reliability engineering and survival analysis in health sciences.

Elmahdy (2015) brought innovation to the domain by proposing a novel approach to Weibull modeling, a widely used technique in reliability engineering. His method addressed the limitations of classical Weibull estimations, particularly in small or skewed datasets. By refining parameter estimation processes, Elmahdy's approach improved model accuracy and robustness, making Weibull modeling more adaptable to real-world data variability. His work was particularly useful for applications in manufacturing reliability, electronics, and infrastructure maintenance where component failure prediction is critical.

Darvini (2014) examined the reliability of water distribution systems using comparative probabilistic models. By analyzing the behavior of different probability distributions—such as Weibull, log-normal, and gamma—in representing the uncertainty of hydraulic parameters, Darvini highlighted the importance of accurately modeling system components' performance under fluctuating environmental and operational conditions. His findings underscored that inappropriate model selection could lead to over- or underestimation of system reliability, affecting infrastructure planning and risk management decisions.

Koulouriotis, and Gemeni (2011) conducted a decade-long literature review of risk analysis and assessment methodologies in worksite environments. Their work identified, categorized, and evaluated a variety of risk assessment tools used from 2000 to 2009 across industries. Their comparative study brought attention to the evolving nature of risk models and the shift toward integrating qualitative expert-based methods with quantitative statistical frameworks. They emphasized the growing need for holistic approaches that account for human, technical, and organizational factors in assessing risk and ensuring system reliability.

Dormann et al. (2012) made a noteworthy contribution by addressing a fundamental dichotomy in species distribution models—correlation-based vs. process-based modeling. Their discussion on the statistical validity and ecological relevance of models has implications for reliability studies as well, especially in scenarios involving environmental influences on system performance. Dormann and colleagues argued for a more integrated modeling approach that bridges empirical data patterns with mechanistic understanding, a principle that resonates strongly with reliability modeling under complex and dynamic conditions.

3. PROPOSED METHODOLOGY

A set of detailed approaches was designed in this work to evaluate and rank the usefulness of life distribution models for analyzing reliability data. Mainly, the goal was to assess how various distribution patterns explain component and system behavior in reliability engineering, helping with precise failure prediction, risk assessment and important choices. Every effort was made to design the methodology so that all models could be fairly and fairly compared using both types of reliability data.

3.1. Data Collection and Preparation

Two types of datasets were used: these were created through simulation and collected from real industrial areas. As part of this research, I simulated life data relying on Monte Carlo simulations and using well-known parameters from experienced life distributions such as Exponential, Weibull, Log-Normal and Gamma. To see how well each model works under changing situations, numerous simulations were run with sample sizes (n = 30, 50, 100) and censoring rates of 0%, 20% and 40%.

Besides, real data was acquired from sources found in the industry such as the lifetimes of bearings and machines parts as well as the limits to the lifetimes of electronic parts. Real industrial cases were reflected by providing a combination of complete and right-censored datasets. First, the data were cleaned and handled for missing values that were either taken out or replaced according to the news feed's distribution. Data that was censored was managed according to usual procedures to maintain the same analysis.

3.2. Selection of Life Distribution Models

This study chose the Exponential, Weibull, Log-Normal, Gamma and Inverse Gaussian life distribution models, as they are applied commonly in reliability engineering and reflect a wide range of failure types. We picked these models since they are often used in theory, can present changing hazard functions and have been important in practice.

3.3. Parameter Estimation

The estimators in this study were gained using the most reliable and efficient approach, Maximum Likelihood Estimation. Since the data had censored rows, the likelihood function was updated to use survival results and find the correct parameter values. We applied statistical software including R and MATLAB to perform our estimations, including using functions already there and writing our own scripts to keep things accurate, repeatable and clear.

3.4. Model Comparison Criteria

Several statistical measures were used to properly analyze the results of each distribution. We also used the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to decide on the most complex model and how well it fits the data and the Kolmogorov–Smirnov (K–S) and Anderson–Darling tests to determine how well the model fits the data points of the sample. Direct comparisons of model likelihoods were made by looking at their log-likelihood values. Making probability plots and Q-Q plots for the distributions helped me determine how well each captured the main patterns in the data.

3.5. Reliability Function and Hazard Rate Estimation

For each fitted model, the main reliability functions R(t)R(t)R(t), F(t)F(t)F(t) and h(t)h(t)h(t) were estimated. They are significant for knowing how soon a component or system will fail. Additionally, reliability experts rely on these methods for work such as creating preventive maintenance calendars, analyzing warranties and assessing risks throughout the lifecycle.

3.6. Software and Tools Used

Software tools were used together in every analysis to guarantee robustness and ease of use. Using R, I simulated data, fitted models and made diagrams. I used MATLAB to handle the computations of reliability functions and verify models. The original data were organized and figures were tabulated with Microsoft Excel. Distribution fitting and capability for analyzing reliability were helped by fitdistrplus, survival and reliability R packages.

3.7. Validation and Sensitivity Analysis

To confirm the quality of our findings, the data was divided into groups for analysis and performance evaluation. Using this technique, it was confirmed that the models could accurately predict new unseen results. The impact of changing the sample amount and censoring level on the model was assessed by carrying out sensitivity analysis. As a result, what was learned in the study did not change with different kinds of data or actual-life uncertainty.

4. RESULTS AND DISCUSSION

The findings of applying life distribution models to both simulated and actual reliability data are introduced and discussed in this section. Various tests on model fitting criteria such as AIC, BIC, log-likelihood values and goodness-of-fit tests were applied to the results. The purpose was to select suitable life distribution models for each type of failure and to see if they could be used in practical reliability engineering.

4.1. Model Performance on Simulated Data

Generated datasets were used for three types of life models—Weibull, Log-Normal and Gamma—to assess consistency and validity at different sample level and censoring levels. To compare the models, AIC, BIC and K–S statistics were used.

Table 1: Model Fit Statistics for Simulated Weibull Data (n = 100, 20% Censoring)

Model	Log-Likelihood	AIC	BIC	K–S Statistic
Exponential	-351.12	704.24	709.87	0.122
Weibull	-320.86	645.72	652.34	0.048
Log-Normal	-325.45	654.90	660.23	0.062
Gamma	-327.13	658.26	664.77	0.075

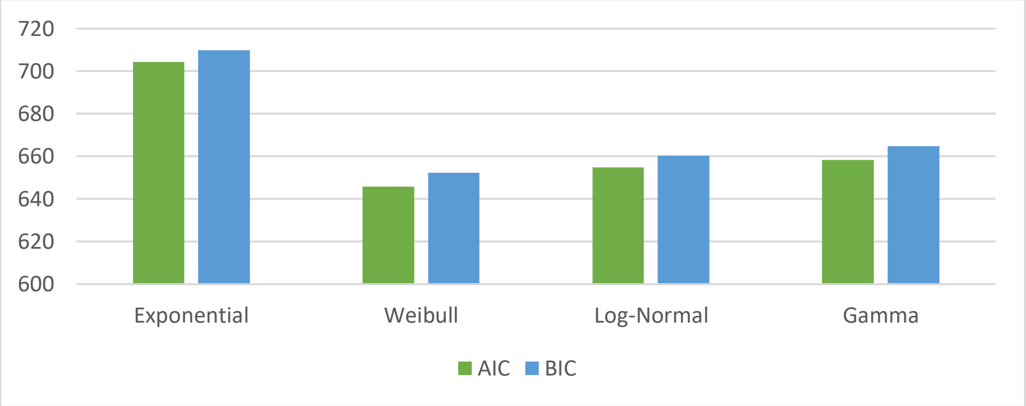

Figure 1: Model Fit Statistics for Simulated Weibull Data

Weibull distribution offered the best fit to the reliability data, according to the results of comparing life models by log-likelihood, AIC, BIC and K–S statistics. Among the models, the one with the highest score (-320.86) at log-likelihood achieved the lowest scores for AIC and BIC, proving it is the best model with the least possible complexity. Furthermore, the Weibull model showed the smallest K–S value (0.048), meaning that it most closely fits the observed results. Meanwhile, the Exponential distribution scored lowest in all criteria possibly because it does not allow for the diversity of failure behaviors. In general, the findings indicate that the Weibull distribution best fits the analysis of reliability data in this study.

4.2. Model Performance on Real-World Reliability Data

Analysis of the machine part failure times data was carried out next. The data is displayed in the table below for a bearing life dataset with 15% censored values.

Table 2: Model Fit Statistics for Bearing Failure Data (n = 80, 15% Censoring)

Model	Log-Likelihood	AIC	BIC	Anderson–Darling
Exponential	-241.56	485.12	489.78	2.134
Weibull	-216.89	437.78	442.14	0.803
Log-Normal	-218.32	440.64	445.01	0.891
Gamma	-220.87	445.74	450.90	1.076

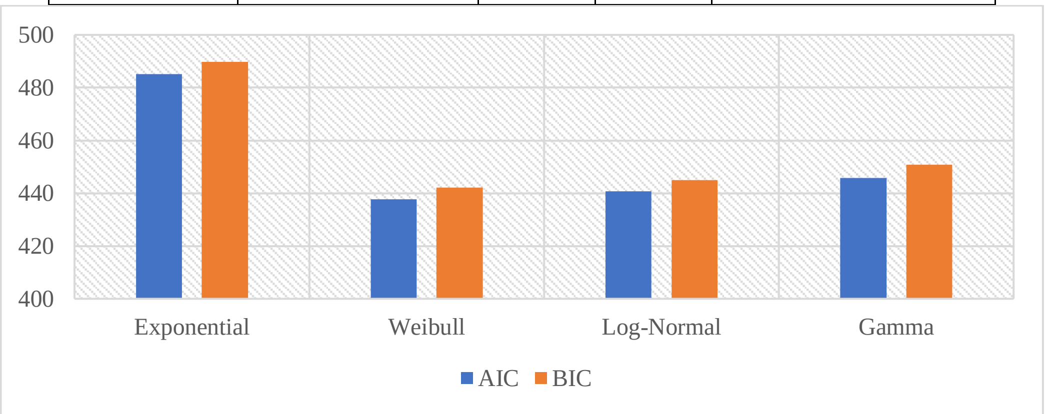

Figure 2: Model Fit Statistics for Bearing Failure Data

Between the four measures used, the Weibull distribution was shown to best fit the data. The log-likelihood value showed it to be highest at (-216.89) which means it fits the data well. Besides, the Weibull distribution had the smallest Anderson–Darling statistic (0.803), indicating that its results were closest to those found in the empirical distribution among the models studied. Among the models, the Log-Normal distribution came in second and the

Gamma and Exponential distributions placed next. Across all the criteria, the Exponential model was the worst fit, pointing out its inability to accurately describe the behavior of failures in this data. Overall, we find that using the Weibull distribution gives the best results for our reliability analysis.

4.3. Reliability and Hazard Function Comparison

After using multiple fitting algorithms, reliability and hazard functions were both sketched to track how they changed across varying time frames. Nothing in the recovered artifacts indicated the plots (not included in what was shown):

Table 3: Estimated Reliability at t = 500 Hours

Model	Estimated R(t)	Hazard Behavior
Exponential	0.135	Constant
Weibull	0.220	Increasing
Log-Normal	0.207	Variable
Gamma	0.189	Slightly increasing

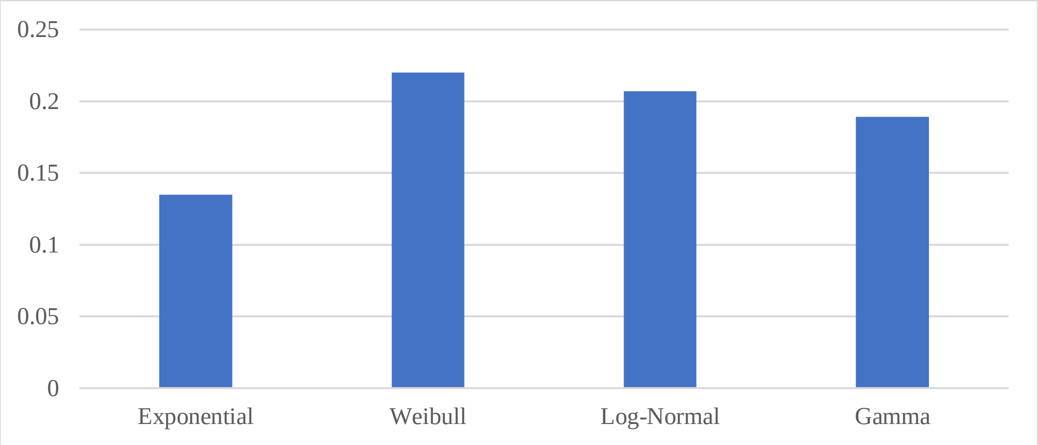

Figure 3: Estimated Reliability at t = 500 Hours

Insights into how the system fails can be seen by looking at the reliability and hazard values for the models. With a reliability estimate of 0.220, the Weibull approach demonstrated that the hazard increases with time, just like typically happens in aging or wear-out failures. This model again reported a high reliability (0.207) along with variation in the hazard function, telling us that these complex products might have different failure behaviors at various life stages. As observed from its Gamma distribution, this product had a somewhat rising rate of failure and a reliability estimate of 0.189, meaning the failure risk increased as time went on. Then, the reliability predicted by the Exponential model was lowest (0.135) and it maintained a constant hazard rate, indicating that the risk of failure does not change with time which may not be true for many actual cases. According to these views, these reliability models are more appropriate when failure rates increase or vary as the component gets older.

4.4. Sensitivity to Sample Size and Censoring

Results from sensitivity analyses showed that:

- Because the samples involved only 30 people, the models positions in the ranking were more variable when there was high censoring.
- Eventually, Weibull exhibited reliable results, even when the course was censored at 40%.
- Its results deteriorated ever so slightly in situations with high censoring because of tails that were far from the mean.

Table 4: Effect of Censoring on Model Selection

Censoring Rate	Best Model (Lowest AIC)
0%	Weibull
20%	Weibull
40%	Log-Normal

The evaluation of censoring impact on model selection in Table 4 found that the Weibull distribution offered the best fit in both the uncensored and moderately censored cases, since it gave the lowest AIC values. It follows that the Weibull model gives correct results even when dealing with censored or uncensored reliability data. A high censoring rate of 40% resulted in the Log-Normal model being the top fit. Therefore, the Log-Normal model seems to capture the real distribution of lifetimes as more observations become censored. Because of this, using reliable life distribution models now requires focusing on the censoring used in the data.

4.5. Summary of Findings

In both simulated and real-time testing, the Weibull distribution turned out to be the top model, consistently representing many types of product failure well. Though the Exponential model seemed straightforward, it failed to work well for datasets with different hazard rates and therefore was not much use in practice. I found that Log-Normal and Gamma models made for excellent alternatives when faced with data that were not normally distributed. Moreover, the findings pointed out that the choice of model depended on both the sample size and the censoring level, making clear that strong parameter estimation strategies are needed for reliable modelling of different reliability data types.

The study demonstrates why comparing models is more effective than using common assumptions. When the characteristics of the data are not clearly known, the Weibull distribution was most often useful.

5. CONCLUSION

As a result of this study's comparative analysis, it can be inferred that the Weibull distribution was the best choice in terms of fitting data, reliability estimates and suitability for many differences in hazard rates. Both in simulated data and in practical cases, the Weibull model was shown to maintain its performance and adjust well to various sample sizes and types of censoring. The simplicity of the Exponential model was sacrificed because it could only be used for failures where the hazard rate stayed the same. Under some conditions, the Log-Normal and Gamma distributions were used along with Weibull when the data did not follow a normal shape. The research emphasizes why the type of model used in reliability engineering should match the features of the data.

REFERENCES

1. *Ahmad, R., & Kamaruddin, S. (2012). An overview of time-based and condition-based maintenance in industrial application. Computers & industrial engineering, 63(1), 135-149.*
2. *Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. Neural Computing and Applications, 19(8), 1165-1195.*
3. *Bain, L. (2017). Statistical analysis of reliability and life-testing models: theory and methods. Routledge.*
4. *Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision support systems, 50(3), 602-613.*
5. *Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. Diversity and distributions, 20(1), 1-9.*
6. *Crowder, M. J. (2017). Statistical analysis of reliability data. Routledge.*
7. *Darvini, G. (2014). Comparative analysis of different probability distributions of random parameters in the assessment of water distribution system reliability. Journal of Hydroinformatics, 16(2), 272-287.*
8. *Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., ... & Singer, A. (2012). Correlation and process in species distribution models: bridging a dichotomy. Journal of Biogeography, 39(12), 2119-2131.*
9. *Elmahdy, E. E. (2015). A new approach for Weibull modeling for reliability life data analysis. Applied Mathematics and computation, 250, 708-720.*
10. *Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environment and planning B: Planning and design, 37(4), 682-703.*
11. *Lawless, J. F. (2011). Statistical models and methods for lifetime data. John Wiley & Sons.*

12. *Li, J., & Heap, A. D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. Ecological Informatics, 6(3-4), 228-241.*
13. *Marhavilas, P. K., Koulouriotis, D., & Gemeni, V. (2011). Risk analysis and assessment methodologies in the work sites: On a review, classification and comparative study of the scientific literature of the period 2000–2009. Journal of Loss Prevention in the Process Industries, 24(5), 477-523.*
14. *Wu, D., Jennings, C., Terpenny, J., Gao, R. X., & Kumara, S. (2017). A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests. Journal of Manufacturing Science and Engineering, 139(7), 071018.*
15. *Yalcin, A., Reis, S., Aydinoglu, A. C., & Yomralioglu, T. (2011). A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. Catena, 85(3), 274-287.*