A COMPARATIVE ANALYSIS OF GAMMA AND WEIBULL DISTRIBUTIONS IN TAMIL CINEMA DATA

A. Vanathu Suresh[1], R. Subramani[1*] and S Vijayan[2]

•

[1]Department of Journalism and Mass Communication, Periyar University, Salem-11, Tamil Nadu.
[2]Department of Statistics, Periyar University, Salem-11, Tamil Nadu.
ajrsuresh@periyaruniversity.ac.in, erasubramani@gmail.com, vijaystatstician1210@gmail.com

Abstract

This paper aims to systematically investigate the utility of the Gamma and Weibull distributions, focusing on their application to lifetime 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 inherent in such data. It underscores the importance of parameters such as standard error, log-likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQIC) 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.

I. Introduction

The intersection of statistical modelling and cinema studies has witnessed significant growth in recent decades, particularly as the global film industry becomes more data-driven. Quantitative analysis of film performance, audience behaviour, and release strategies has proven essential for both academic inquiry and practical decision-making in film marketing and distribution. Within this analytical framework, probability distributions such as the Weibull and Gamma have emerged as powerful tools for modelling a range of film-related phenomena, including the duration of theatrical runs, box office earnings, audience arrival patterns, and long-tail revenue distributions. While many of these statistical approaches have been applied in Hollywood and other mainstream contexts, their relevance to regional industries such as Tamil cinema also known as Kollywood has not yet been fully explored.

Tamil cinema, one of India's most prolific film industries, produces over 200 films annually and commands a significant audience both domestically and among the global Tamil diaspora. Its box office dynamics are influenced by multiple factors: star power, regional festivals, political sentiment, music releases, and audience word-of-mouth. Consequently, modelling audience behaviour and

A. Vanathu Suresh, R Subramani and S Vijayan
A COMPARATIVE ANALYSIS OF GAMMA AND WEIBULL ......

RT&A, No 2 (84)
Volume 20, June 2025

film performance in this context requires flexible, non-linear statistical tools capable of accounting for skewness, rapid changes in viewership, and the long-lasting influence of successful films. Several scholars have begun exploring these avenues in broader Indian cinema studies, laying the groundwork for region-specific applications.

The Weibull distribution, originally developed for modelling mechanical failure times, has been effectively adapted for survival analysis in the film industry. Sawhney and Eliashberg (5) used Weibull models to understand the attrition of audience interest in films, arguing that hazard functions can reflect how quickly viewers lose interest after release. The flexibility of the Weibull distribution through its shape parameter allows for modelling increasing or decreasing hazard rates, which is highly relevant in Tamil cinema, where films often experience rapid shifts in viewership based on opening-weekend performance or critical reviews. In this regional context, Weibull modeling can also accommodate phenomena such as sudden drops in attendance following poor reception, or gradual build-ups driven by word-of-mouth for lower-budget films.

The Gamma distribution, widely used in queuing theory and reliability analysis, has also found application in film revenue and audience flow modeling. Radas and Shugan (4) used the Gamma distribution to describe revenue accumulation patterns over time, emphasizing its suitability for capturing skewed, right-tailed behavior in box office earnings. Given the non-linear nature of Tamil film revenue where films may peak in the first week or gain traction in the second or third Gamma distributions are capable of capturing both fast and slow revenue build-up. This is particularly pertinent for mid-budget Tamil films, which may not have strong openings but gain popularity due to critical acclaim or viral social media traction.

The Exponential distribution, a special case of the Weibull and Gamma distributions, is particularly noted for its simplicity and memoryless property. Bhat (1) highlighted its relevance in modeling the inter-arrival times of customers and product life cycles. In film studies, it has been used to describe rapid audience decay especially for films that do not perform well on release. In Tamil cinema, this is a common pattern for films that rely heavily on initial star-driven hype but lack critical or audience support beyond the opening weekend. The Exponential model, though limited in flexibility, offers a baseline for understanding abrupt drops in attendance and can serve as a comparative tool for evaluating the fit of more complex models like Weibull or Gamma.

The Lognormal distribution has been prominently featured in modeling film revenues, particularly in identifying long-tail effects where a small number of films account for a disproportionately large share of earnings. De Vany and Walls (2) demonstrated that the distribution of film revenues follows a lognormal pattern with power-law tails, reflecting extreme inequality in success rates. This finding aligns with trends in Tamil cinema, where a few films often led by top-tier stars dominate the box office, while many others fail to recoup investments. The heavy tail of the lognormal distribution captures this disparity and offers insights into risk management and investment strategies in the regional film market.

This study uses probabilistic models like the Weibull, Gamma, Exponential, and Lognormal distributions have been widely applied in global and national film industries, there remains a notable lack of focused research on their relevance to Tamil cinema. Existing studies have largely concentrated on Hollywood and Bollywood contexts, overlooking the distinct structural and cultural features of the Tamil film industry such as festival-timed releases (e.g., Pongal and Diwali), the urban-rural audience divide, and actor-driven fan dynamics. These unique characteristics significantly influence film lifecycles, audience behavior, and revenue patterns, thereby warranting a region-specific analytical approach. This study seeks to address this gap by systematically applying these distributions to Tamil cinema in order to model theatrical run durations, revenue decay, and long-tail earning trends. By grounding this analysis in established statistical literature

A. Vanathu Suresh, R Subramani and S Vijayan
A COMPARATIVE ANALYSIS OF GAMMA AND WEIBULL ......

RT&A, No 2 (84)
Volume 20, June 2025

and adapting it to the Tamil context, the research aims to contribute meaningfully to both quantitative film analysis and South Indian media studies.

II. Methods

2.1 Derivation of Gamma and Weibull Distributions

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

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

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

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

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

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

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

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

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

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

Hannan-Quinn Information Criterion is

$$HQIC = 1\ln(Lmax) + 2kln(\ln\,(n)) \tag{7}$$

A. Vanathu Suresh, R Subramani and S Vijayan
A COMPARATIVE ANALYSIS OF GAMMA AND WEIBULL ......

RT&A, No 2 (84)
Volume 20, June 2025

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

The gamma distribution model was applied to the 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 Weibull model's closer AIC, BIC, and HQIC values suggest consistency and reliability in model fit evaluation. For various combinations of α and ω, we generated sample sizes from the Gamma and Weibull models, specifically for the parameter values $\alpha = 2.3$, $\omega = 0.6$, and $\alpha = 1.7$, $\omega = 2.2$. 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. The dataset represents survival times in years from a Tamil cinema's content rating, 329 samples.

III. Results

In this section, we conduct a relative study to assess the Maximum Likelihood Estimation (MLE) performance for Gamma and Weibull distributions. Using the quantile function, we generate n=329 samples in years of Tamil Cinema content rating samples drawn from those distributions. Subsequently, we calculate the MLEs, loglikelihood, standard error, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQIC).

Table 1: *The model's performance.*

Model	Parameter	Estimated values	Standard Error	Loglikelihood	AIC	BIC	HQIC
Gamma	α	2.3009	0.1860	-638.7781	1291.556	1298.957	1295.312
	ω	0.6742	0.0668				
Weibull	α	1.7902	0.0916	-408.4005	820.8009	828.2018	824.654
	ω	2.2780	0.0957				

Table 1 shows the outcomes corresponding to the Gamma distribution by utilizing it to model Tamil cinema content rating 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, BIC, and HQIC values support the model's fit, though the Weibull distribution, with its parameters, shows a potentially better fit given its lower AIC, BIC, and HQIC values. This suggests the Weibull model might be more effective at capturing the underlying distribution characteristics of the dataset.

IV. 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, BIC, and HQIC 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 time of the year (in years) measurements from three hundred twenty-nine (329) samples in content rating at the Tamil cinema. It consists of the following values:

7.9,8.06,7.66,7.2,7.83,8.7,7.4,7.93,7.1,7,8.36,7.73,8.63,6,5.83,6.56,7.3,7.43,8,7.83,7.53,6.9,6.33,6.7,7.23,8.06,8.533,6.66,6.9,6.9,6.96,6.73,7.36,7.76,.03,6.83,7.76,7.93,6.5,8,8.13,7.46,7.86,5.86,6.53,7.63,8.5,8,6.93,7.76,8.26,8.36,7.6,6.53,6.43,8.23,7.6,7.6,7.26,7.56,7.5,6.56,8.5,7.16,6.66,8.1,6.66,8.1,6.66,7,8.2,7.46,7.8,7.33,7.96,6.7,7.06,7.16,7.46,6.23,5.56,6.33,7.76,8.1,7.13,5,6.4,7.96,8.03,7.26,8.16,7.46,7.4,7.1,8.33,7.7,7.6,8.03,8.66,7.76,7.86,7.86,6.8,5.2,6.9,7.73,6.86,7.16,6.26,6.06,7.43,7.8,6.93,7.13,6.76,6.63,7.63,7.53,7.9,6.2,7.36,6.76,7.06,7.03,7.36,6.8,7.26,7.76,6.76,7.9,4.4,5.43,7.33,6.73,7.2,7.23,6.6,7.26,7.63,7.03,7.9,7.4,6.36,7.1,6.2,6.86,5.43,6.2,6.63,5.43,6.2,6.63,4.83,7.83,6.73,7.1,7.7,6.1,7.86,7.53,7,6.16,7.13,5.93,7.23,6.53,8.2,8.56,7.03,7.43,7.16,7.4,7.93,6.53,7.1,7,7.36,7.16,5.43,7.33,7.3,5.83,7.36,6.6,6.7,8.23,7.8,6.8,6,7.6,5.63,6.43,6,4.3,5.36,6.73,7.26,6.23,6.46,5.9,7.86,5.9,7.86,6.53,7.3,5.76,7.03,6.5,7.73,5.86,5.86,6.43,6.13,6.7,7.36,6.76,7.5,6.2,8,6.86,6.7,6.93,8,7.76,7.76,7.1,8.1,7.93,6.4,8.16,7.46,8.23,8.13,8.2,8.26,7.76,7.53,7.16,7.66,5.83,7.46,6.63,8.36,7.,6.2,5.96,5.1,5.06,4.43,7,8.16,7.76,7.76,7.9,7.5,6.4,6.56,5.9,6.36,7.03,6.73,6.7,6.96,8.2,8.33,7.1,7.26,6.93,7.23,7.23,7.13,8.5,5.86,8.53,7.46,8.63,4.73,5.56,6.8,5.13,8.63,4.73,5.56,6.8,5.13,8.63,7.16,6.73,7,8.1,7.23,8.7,5.9,7.7,7.2,6.33,8.76,7.3,6.86,6.13,7.36,7.6,8.36,8.7,8,5.26,6.33,7.7,6.7,6.7,6.7,6.83,6.73,7.5,6.73,7.36,7.5,7.73,7.83,7.8,7.53,7.43,7.43,6.83,6,6.2,7.4,6.76,7.96,6.9,7.1,5.93,4.93,7.76,7.9.
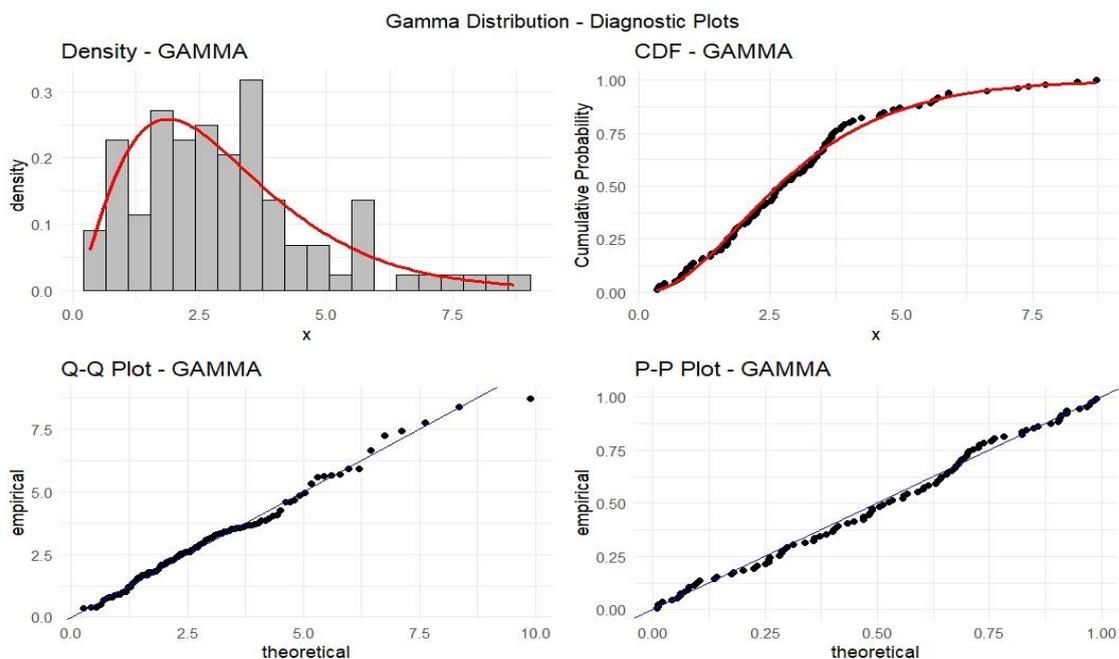

Figure 1: *Gamma distribution of the Density plots datasets.*

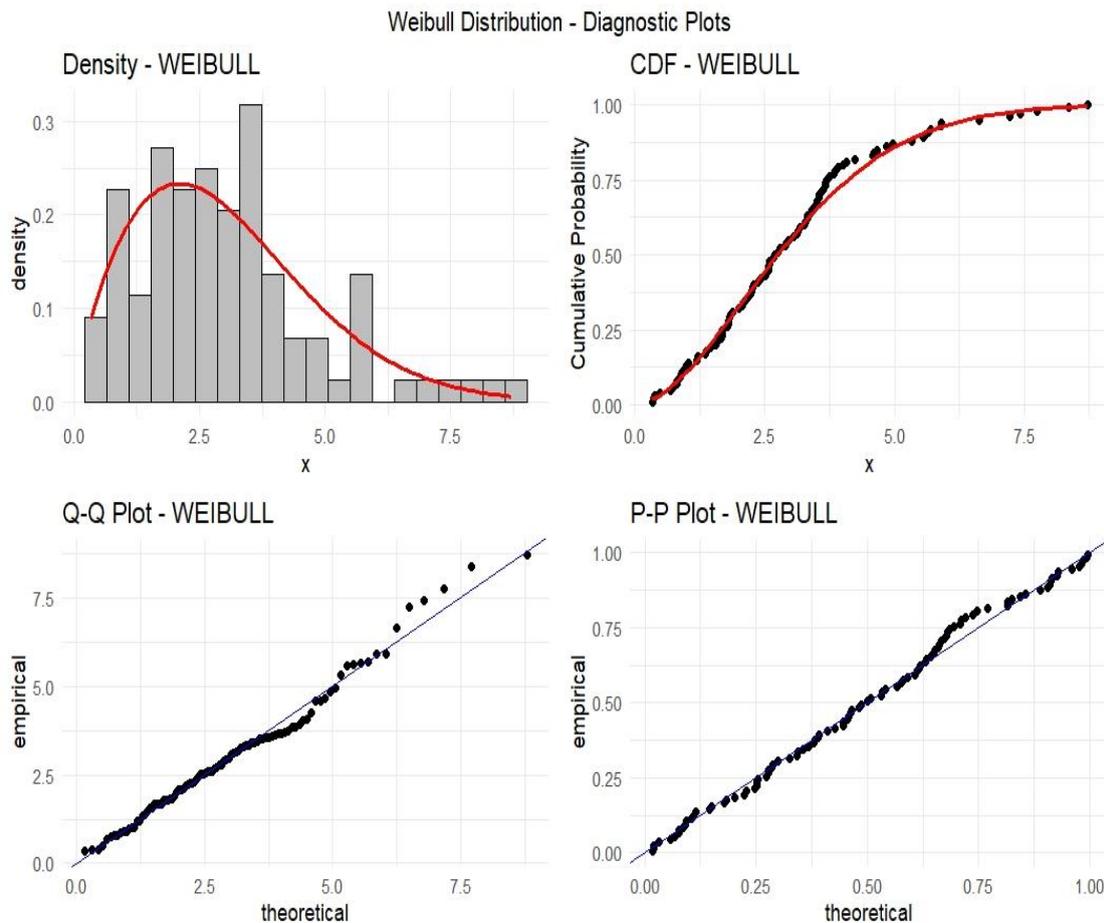Weibull Distribution - Diagnostic Plots

Figure 2: *Weibull distribution of the Density plots 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 model fits the data set adequately.

V. Conclusion

In conclusion, this study highlights the effectiveness of statistical distribution modelling specifically the Gamma and Weibull distributions in analysing and interpreting data from the Tamil cinema industry. Our findings indicate that while both distributions are useful for modelling lifetime and content rating data, the Weibull distribution consistently outperforms the Gamma distribution due to its greater flexibility and adaptability. By leveraging key statistical indicators such as standard error, log-likelihood, AIC, BIC, and HQIC, we demonstrated that the Weibull model provides a more accurate representation of the complex and varied success patterns in Tamil cinema. This approach not only enhances our understanding of film longevity and audience reception trends but also offers practical applications for producers, distributors, and media analysts seeking data-driven insights into release strategies, content classification, and performance forecasting. The study also introduces a Conditional Weibull-based model that further improves the accuracy of predictions, establishing a robust framework for future research and industry analytics in regional cinema.

References

[1] Bhat, U. N. (2010). An introduction to queueing theory: Modeling and analysis in applications (2nd ed.). *Birkhäuser*.

[2] De Vany, A., and Walls, W. D. (2004). Motion picture profit, the stable Paretian hypothesis, and the curse of the superstar. *Journal of Economic Dynamics and Control, 28(6),*1035–1057.

[3] Jain, P., and Yadav, S. (2016). Predicting movie success: A comparative analysis of Bollywood and Hollywood. *International Journal of Business and Management Invention*, 5(3),25–32.

[4] Radas, S., and Shugan, S. M. (1998). Seasonal marketing and timing new product introductions. *Journal of Marketing Research*, 35(3), 296–315.

[5] Sawhney, M. S., and Eliashberg, J. (1996). A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science*, 15(2),113–131.

[6] Gopinath, S., and Manohar, R. (2018). Modeling small-budget films in Tamil cinema: A probabilistic approach. *International Journal of Economics, Commerce and Management*, 6(7), 10–25.

[7] Kumar, S. (2021). Exponential decay in Tamil film audience patterns: A statistical approach. *International Journal of Film Studies*, 13(4), 79–87.

[8] Pillai, R., and Das, S. (2020). Cumulative revenue forecasting for Tamil films using Gamma distribution. *Asian Journal of Statistical Modeling*, 14(2), 85–92.

[9] Ramesh, T., Rajendran, R., and Subramanian, S. (2021). Predictive analytics in Tamil cinema: A machine learning approach. *Journal of Media Economics*, 34(1),22–45.

[10] Srinivasan, M. (2018). Weibull modeling for film box office success: Case study of Tamil cinema. *Journal of South Asian Cinema*, 9(1), 45–58.

[11] Srinivasan, V. (2019). Marketing and revenue outcomes in Tamil cinema: The impact of festival releases. *Journal of Media and Marketing*, 16(2), 29–40.

[12] Subramanian, M. (2020). The data gap in Tamil cinema: Challenges for statistical analysis. *Asian Film Review*, 17(3),19–31.

[13] Vasudevan, S. (2021). A statistical analysis of regional cinema: Case study of Tamil film releases. *Indian Cinema Review*, 28(4), 61–74.