Journal of Probability and Statistical Science
24(1), 1-15 Feb. 2026

Length-Biased MGAMMA Distribution: Statistical Properties and Cancer Data Application

R. Jothika
Department of Statistics
Annamalai University, Tamil Nadu, India

P. Pandiyan
Department of Statistics
Annamalai University, Tamil Nadu, India

ABSTRACT

In this study, we propose the length-biased Mgamma distribution for analyzing lifetime data. We obtain its main mathematical properties and compute the parameter estimates using the maximum likelihood method. The model is then applied to real cancer survival data, and the results show that it provides a better fit than several other lifetime models. Overall, the length-biased Mgamma distribution proves to be a practical and flexible option for survival data analysis.

Keywords: Length-biased Mgamma; Maximum likelihood estimation; Cancer survival data; Lifetime modelling.

1. Introduction

In many real-life situations, the recorded data are influenced by the way observations are selected. When the chance of selecting an observation depends on its size, duration, or magnitude, ordinary probability distributions often fail to describe the data properly. To address this, statisticians use weighted distributions, a concept first introduced by Fisher (1934), who explained how sampling bias can distort observed frequencies, and later formalized by Rao (1965), who established a general mathematical framework for modeling unequal selection probabilities. These models are widely used in medical research, engineering, reliability, ecology, and related fields.

A particularly important subclass of weighted models is the length-biased distribution. This arises when longer lifetimes or larger units have a higher probability of being included in the sample. Length-biased sampling frequently occurs in medical studies, especially in cancer survival analysis, where patients must survive long enough to be observed. Cox (1969) introduced the concept of length-biased sampling by explaining how longer lifetimes are more likely to appear in observed data.

Over the years, several researchers have focused specifically on developing lifetime models that incorporate length bias to more accurately represent survival and reliability data. Afaq et al. (2016) introduced the length-biased weighted Lomax distribution and showed that it is useful for modeling lifetime data. Rather and Subramanian (2018) proposed the length-biased weighted generalized uniform distribution and studied its main statistical properties and parameter estimation. Abd-Elfattah A. M et al. (2021) introduced the length-biased Burr-XII distribution and explored its main properties and parameter estimation for lifetime data. Mustafa and Khan (2023) proposed the length-biased extended Rayleigh distribution and established its statistical properties, showing through real data examples that the model offers improved flexibility compared to existing lifetime distributions. More recently, Alzoubi (2024) introduced the length-biased Benrabia distribution and demonstrated its statistical properties. The paper shows, via simulations and real data applications, that this new distribution can effectively model skewed lifetime data with a better fit than competitor models. Pushkarna and Mustafa (2025) introduced the length-biased weighted Ishita distribution and examined its

- Received December 5, 2025, in final form February 2026.
- R. Jothika (corresponding author) is affiliated with Department of Statistics, Annamalai University, Tamil Nadu, India.
rjothika2023@gmail.com

statistical properties, demonstrating through real data applications that the model provides superior flexibility and improved fitting performance for lifetime data.

In this study, the Length-Biased Mgamma distribution is considered. The Mgamma distribution, proposed by Khan and Ruidas (2025), is a one-parameter lifetime model. The proposed length-biased form increases the flexibility of the model and performs better than several existing lifetime distributions when applied to cancer survival data.

2. Length Biased MGAMMA Distribution (LBMD)

The probability density function (pdf) of Mgamma distribution is given by

$$f(x) = \frac{\theta^3\, x\left(1+\frac{x}{2}\right)e^{-\theta x}}{1+\theta};\ x > 0, \theta > 0 \tag{1}$$

and its cumulative distribution function (cdf) of Mgamma distribution is given by

$$F(x) = 1 - \left[1 + \frac{\theta x}{1+\theta}\left(1 + \theta + \frac{\theta x}{2}\right)\right]e^{-\theta x}; x > 0, \theta > 0 \tag{2}$$

If x be a random variable with probability density function $f(x)$ and $w(x)$ be a non-negative weight function, then the probability density function of the weighted random variable X_w is given by

$$f_w(x) = \frac{w(x)f(x)}{E(w(x))} \quad ; \quad x > 0$$

Where $w(x)$ is a non-negative weight function and $E[w(x)] = \int w(x)f(x)dx < \infty$.

Let $w(x) = x^c$ be the weight function. For c=1, the resulting weighted model corresponds to the Length-Biased Mgamma distribution, with pdf given by

$$f_l(x) = \frac{xf(x)}{E(X)} \quad ; \quad x > \tag{3}$$

Where

$$E(X) = \int_0^\infty x\, f\,(x)dx$$

$$= \int_0^\infty x\left(\frac{\theta^3\, x\left(1+\frac{x}{2}\right)e^{-\theta x}}{1+\theta}\right)dx$$

$$= \frac{\theta^3}{1+\theta}\int_0^\infty x^2\left(1+\frac{x}{2}\right)e^{-\theta x}dx$$

$$= \frac{\theta^3}{1+\theta}\int_0^\infty \left(x^2+\frac{x^3}{2}\right)e^{-\theta x}dx$$

$$= \frac{\theta^3}{1+\theta}\left[\int_0^\infty x^2\, e^{-\theta x}dx + \int_0^\infty \frac{x^3}{2}e^{-\theta x}dx\right]$$

$$E(X) = \frac{\theta^3}{1+\theta}\left[\int_0^\infty x^{3-1}\, e^{-\theta x}dx + \frac{1}{2}\int_0^\infty x^3\, e^{-\theta x}dx\right]$$

Using Gamma function

$$\frac{\Gamma(Z)}{a^Z} = \int_0^\infty t^{z-1}e^{-ta}dt$$

$$E(X) = \frac{\theta^3}{1+\theta}\left[\frac{\Gamma(3)}{\theta^3} + \frac{\Gamma(4)}{2\theta^4}\right]$$

After Simplification, we get

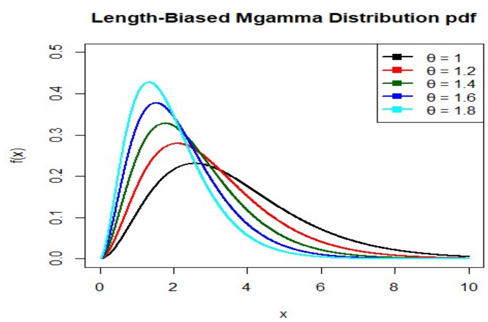
$$E(X) = \frac{2\theta + 3}{\theta(1+\theta)} \quad (4)$$

substituting equation (1) and (4) in equation (3), we obtain probability density function (pdf) of length-biased Mgamma distribution is

$$f_l(x) = \frac{xf(x)}{E(X)}$$

$$f_l(x) = \frac{x\left(\frac{\theta^3\, x\, \left(1+\frac{x}{2}\right) e^{-\theta x}}{1+\theta}\right)}{\left(\frac{2\theta + 3}{\theta(1+\theta)}\right)}$$

$$f_l(x) = \frac{\theta^4\, x^2\, \left(1+\frac{x}{2}\right) e^{-\theta x}}{2\theta + 3} \quad (5)$$



The cumulative distribution function (cdf) of length-biased Mgamma distribution is given by

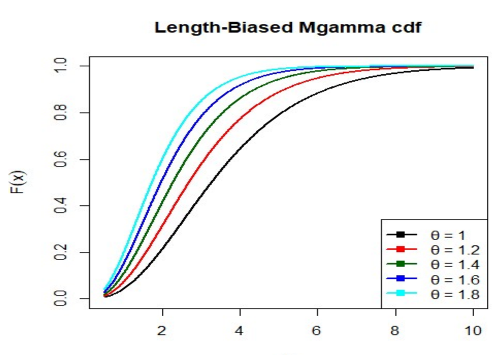
$$F_l(x) = \int_0^x f_l(x)dx$$

$$F_l(x) = \int_0^x \left(\frac{\theta^4\, x^2\, \left(1+\frac{x}{2}\right) e^{-\theta x}}{2\theta + 3}\right) dx$$

$$F_l(x) = \frac{\theta^4}{2\theta + 3}\int_0^x x^2\, \left(1+\frac{x}{2}\right) e^{-\theta x} dx$$

after simplification of the equation (6), we obtain the cumulative distribution function of the length-biased Mgamma distribution as

$$F_l(x) = \frac{2\theta\gamma(3,\theta x) + \gamma(4,\theta x)}{2(2\theta + 3)} \quad (6)$$



3. Reliability Analysis

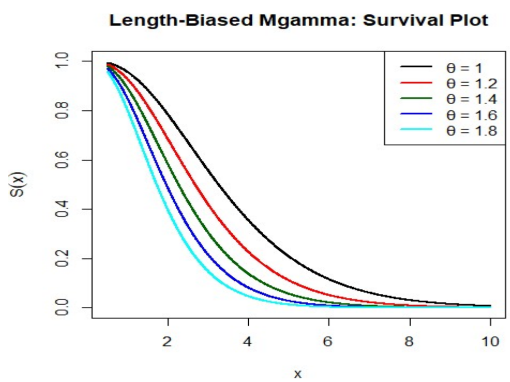
In this section, we discuss the reliability function and hazard function for the proposed length-biased Mgamma distribution.

3.1. Reliability Function

The reliability (survival) function of the length-biased Mgamma distribution is given by

$$S_l(x) = 1 - F_l(x)$$

$$S_l(x) = 1 - \frac{2\theta\gamma(3,\theta x) + \gamma(4,\theta x)}{2(2\theta + 3)}$$



3.2. Hazard Function

$$h(x) = \frac{f_l(x)}{S_l(x)}$$

$$h(x) = \frac{\left(\dfrac{\theta^4\, x^2\, \left(1 + \frac{x}{2}\right) e^{-\theta x}}{2\theta + 3}\right)}{\left(1 - \dfrac{2\theta\gamma(3,\theta x) + \gamma(4,\theta x)}{2(2\theta + 3)}\right)}$$

$$h(x) = \frac{2\left(\theta^4\, x^2\, \left(1 + \frac{x}{2}\right) e^{-\theta x}\right)}{2(2\theta + 3) - (2\theta\gamma(3,\theta x) + \gamma(4,\theta x))}$$

4. Statistical Properties

4.1 Moments

Let X denoted the random variable following LBMD then r^{th} order moments $E(X^r)$ is obtained as

$$E(X^r) = \mu_r{}' = \int_0^\infty x^r f_l(x) dx$$

$$= \int_0^\infty x^r \left(\frac{\theta^4\, x^2\, \left(1 + \frac{x}{2}\right) e^{-\theta x}}{2\theta + 3}\right) dx$$

$$= \frac{\theta^4}{2\theta+3}\int_0^\infty \left(x^{r+2} + \frac{x^{r+3}}{2}\right) e^{-\theta x} dx$$

$$= \frac{\theta^4}{2\theta+3}\left[\int_0^\infty x^{r+2} e^{-\theta x} dx + \frac{1}{2}\int_0^\infty x^{r+3} e^{-\theta x} dx\right]$$

$$= \frac{\theta^4}{2\theta+3}\left[\int_0^\infty x^{(r+3)-1} e^{-\theta x} dx + \frac{1}{2}\int_0^\infty x^{(r+4)-1} e^{-\theta x} dx\right]$$

$$= \frac{\theta^4}{2\theta+3}\left[\frac{\Gamma(\mathrm{r}+3)}{\theta^{r+3}} + \frac{1}{2}\frac{\Gamma(\mathrm{r}+4)}{\theta^{r+4}}\right]$$

$$\mu_r{}' = \frac{2\theta\Gamma(\mathrm{r}+3) + \Gamma(\mathrm{r}+4)}{(2\theta+3)2\theta^r} \tag{7}$$

Putting r=1,2 in equation (7), the mean of LBMD is obtained as

$$\mu_1{}' = \left[\frac{6\theta+12}{\theta(2\theta+3)}\right]$$

$$\mu_2{}' = \left[\frac{24\theta+120}{\theta^2(2\theta+3)}\right]$$

$$\text{Variance } = \mu_2{}' - (\mu_1{}')^2$$

$$= \left[\frac{24\theta+120}{\theta^2(2\theta+3)}\right] - \left[\frac{6\theta+12}{\theta(2\theta+3)}\right]^2$$

$$variance = \left[\frac{(24\theta+120)(2\theta+3) - (6\theta+12)^2}{\theta^2(2\theta+3)^2}\right]$$

Standard Deviation

$$S.D(\sigma) = \sqrt{\frac{(24\theta+120)(2\theta+3) - (6\theta+12)^2}{\theta^2(2\theta+3)^2}}$$

4.2 Moment Generating Function and Characteristic Function

Let X have a LBMD then the MGF of X is obtained as

$$M_X(t) = E(e^{tx})$$

$$= \int_0^\infty e^{tx} f_l(x) dx$$

Using Taylor's series

$$M_X(t) = E(e^{tx}) = \int_0^\infty \left(1 + tx + \frac{(tx)^2}{2!} + \cdots\right) f_l(x;\beta) dx$$

$$= \int_0^\infty \sum_{j=0}^\infty \frac{t^j}{j!} x^j f_l(x) dx$$

$$= \sum_{j=0}^{\infty} \frac{t^j}{j!} \int_0^{\infty} x^j f_l(x) dx$$
$$= \sum_{j=0}^{\infty} \frac{t^j}{j!} \mu_j'$$
$$= \sum_{j=0}^{\infty} \frac{t^j}{j!} \left[\frac{2\theta\Gamma(j+3) + \Gamma(j+4)}{(2\theta+3)2\theta^j} \right]$$

$$M_X(t) = \frac{1}{\theta(2\theta+3)} \sum_{j=0}^{\infty} \frac{t^j}{j!} \frac{2\theta\Gamma(j+3) + \Gamma(j+4)}{(2\theta+3)2\theta^j}$$

In a similar manner, the characteristic function of the LBMD is derived as

$$\Phi_X(t) = \frac{1}{\theta(2\theta+3)} \sum_{j=0}^{\infty} \frac{it^j}{j!} \frac{2\theta\Gamma(j+3) + \Gamma(j+4)}{(2\theta+3)2\theta^j}$$

5. Order Statistics

Let $X_{(1)} \leq X_{(2)} \leq X_{(3)} \ldots . \leq X_{(n)}$ be the order statistics of the random sample taken from LBMD. The pdf of r^{th} order statistics, $X_{(r)}$, is defined as.

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!\,(n-r)!} f_X(x)[F_X(x)]^{r-1}[1 - F_X(x)]^{n-r} \quad (8)$$

Inserting equations (5) and (6) into equation (8), we get the pdf of the rth order statistic X(r) for the LBMD.

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!\,(n-r)!} \left[\frac{\theta^4 x^2 \left(1 + \frac{x}{2}\right) e^{-\theta x}}{2\theta + 3} \right] \left[\frac{2\theta\gamma(3, \theta x) + \gamma(4, \theta x)}{2(2\theta+3)} \right]^{r-1} \left[1 - \frac{2\theta\gamma(3, \theta x) + \gamma(4, \theta x)}{2(2\theta+3)} \right]^{n-r}$$

Consequently, the pdf of the maximum order statistic $X_{(n)}$ for the LBMD is obtained as

$$f_{X_{(n)}}(x) = n \left[\frac{\theta^4 x^2 \left(1 + \frac{x}{2}\right) e^{-\theta x}}{2\theta + 3} \right] \left[\frac{2\theta\gamma(3, \theta x) + \gamma(4, \theta x)}{2(2\theta+3)} \right]^{n-1}$$

Accordingly, the pdf of the minimum order statistic $X_{(1)}$ associated with the LBMD is expressed as

$$f_{X_{(1)}}(x) = \frac{n!}{(r-1)!\,(n-r)!} \left[\frac{\theta^4 x^2 \left(1 + \frac{x}{2}\right) e^{-\theta x}}{2\theta + 3} \right] \left[1 - \frac{2\theta\gamma(3, \theta x) + \gamma(4, \theta x)}{2(2\theta+3)} \right]^{n-1}$$

6. BONFERRONI and LORENZ Curves

The Bonferroni and Lorenz curves are obtained for the LBMD.

$$B(p) = \frac{1}{p\mu}\int_0^q x\ f_l(x)dx$$

And

$$L(p) = \frac{1}{\mu}\int_0^q x\ f_l(x)dx$$

Where $q = F^{-1}(p); q\epsilon[0,1]$

And $\mu = E(X)$

Therefore, the Bonferroni and Lorenz curves of the proposed model are given by

$$\mu = \frac{6\theta + 12}{\theta(2\theta + 3)}$$

$$B(p) = \frac{1}{p\left(\frac{6\theta + 12}{\theta(2\theta + 3)}\right)}\int_0^q x\left(\frac{\theta^4\,x^2\,\left(1 + \frac{x}{2}\right)e^{-\theta x}}{2\theta + 3}\right)dx$$

After solving the above equation, we obtain

$$B(p) = \frac{\theta\gamma(4,\theta q) + \frac{1}{2}\gamma(5,\theta q)}{p(2\theta + 3)}$$

$$L(p) = pB(p)$$

$$L(p) = \frac{\theta\gamma(4,\theta q) + \frac{1}{2}\gamma(5,\theta q)}{(2\theta + 3)}$$

7. Maximum Likelihood Estimation

Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample of size n drawn from the LBMD with parameter θ. The corresponding likelihood function is defined as follows.

$$L(x) = \prod_{i=1}^{n} f_l(x_i)$$

$$= \prod_{i=1}^{n}\left(\frac{\theta^4\,{x_i}^2\,\left(1 + \frac{x_i}{2}\right)e^{-\theta x_i}}{2\theta + 3}\right)$$

$$= \left(\frac{\theta^4}{2\theta + 3}\right)^n \prod_{i=1}^{n} {x_i}^2\,\left(1 + \frac{x_i}{2}\right)e^{-\theta x_i}$$

Then, the log-likelihood function is

$$log\ L = n\,4\,log\ \theta - nlog\ (2\theta + 3) + \sum_{i=1}^{n} log\ {x_i}^2\left(1 + \frac{x_i}{2}\right) - \theta \sum_{i=1}^{n} x_i \tag{9}$$

Deriving (9) partially with respect to θ we have

$$\frac{\partial log\ L}{\partial \theta} = \frac{4n}{\theta} - n\left(\frac{2}{2\theta + 3}\right) - \sum_{i=1}^{n} x_i = 0 \tag{10}$$

Equation (10) provides the maximum likelihood estimating equations for the parameters of the LBMD. As these equations cannot be solved analytically, numerical solutions were obtained using the R software with the observed data.

8. Application

Date Set 1: The data represent the life expectancy (in years) of forty patients with leukemia, a blood cancer, from one of Saudi Arabia's Ministry of Health facilities. The data set is given

Table 1: Blood cancer data: in years, 40 patients' survival time.

0.315	2.211	3.348	4.323	0.496	2.370	3.427	4.381	0.616	2.532
3.499	4.392	1.145	2.693	3.534	4.397	1.208	2.805	3.767	4.647
1.263	2.910	3.751	4.753	1.414	2.912	3.858	4.929	2.025	3.192
3.986	4.973	2.036	3.263	4.049	5.074	2.162	3.348	4.244	5.381

The goodness of fit of the fitted distributions is compared using AIC, BIC, AICC, and the −2 log-likelihood values. These criteria help identify the model that best fits the data.

AIC, BIC, AICC and -2log L can be evaluated by using the formula as follows.

$$AIC = 2K - 2\log L,\quad BIC = k\log n - 2\log L\ \ and\ AICC = AIC + \frac{2k(k+1)}{(n-k-1)}$$

Table 2: Maximum Likelihood Estimates and Model Selection Criteria for Competing Lifetime Distributions

Distribution	MLEs	-2 log L	AIC	BIC	AICC
Length biased Mgamma	$\hat{\theta} = 1.136352$ (0.091998)	146.3923	148.3923	150.0812	148.4976
Mgamma	$\hat{\theta}$ =0.812467 (0.076784)	149.7925	151.7925	153.4814	151.8978
Gamma	$\hat{\alpha} = 3.46469(0.740462)$ $\hat{\beta} = 1.10314(0.25370)$	147.0974	151.0974	154.4752	151.4217
Exponential	$\hat{\theta} = 0.318398$ (0.050343)	171.5563	173.5563	175.2452	173.6616
Lindley	$\hat{\theta} = 0.526921\ (0.06074)$	160.5012	162.5012	164.1900	162.6064

Table 1 shows that the LBMD has the lowest goodness-of-fit values. This means that it explains the data better than the other models used in the study. Therefore, LBMD is suitable for the given dataset.

9. Conclusion

The LBMD was proposed in this study to model lifetime data. Its statistical properties were derived, and the model parameters were estimated using the maximum likelihood method. When applied to real cancer survival data and compared with existing models, the proposed distribution provided a better fit based on standard model selection criteria. Hence, the LBMD is an appropriate model for analyzing lifetime cancer data.

References:

[1]. Fisher, R. A. (1934). The effects of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics*, 6(1), 13–25.

[2]. Rao, C. R. (1965). "*On discrete distributions arising out of method of ascertainment, in classical and Contagious Discrete*", G.P. Patiled; Pergamum Press and Statistical publishing Society, Calcutta. 320-332.

[3]. Cox, D.R. (1968). *Some Sampling Problems in Technology. In: Johnson, N.L. and Smith, H., Eds., New Developments in Survey Sampling*, John Wiley & Sons, New York, 506-527.

[4]. Afaq Ahmad., S.P Ahmad., & A. Ahmed. (2016). Length-Biased Weighted Lomax Distribution: Statistical Properties and Application. *Pakistan Journal of Statistics and operation research*, 12(2), 245-255.

[5]. A. A. Rather, C. Subramanian. (2018). Characterization and Estimation of Length Biased Weighted Generalized Uniform Distribution. *International Journal of Scientific Research in Mathematical and Statistical Science*, 5(5), 72-76.

[6]. Abd-Elfattah A. M, Mahdy. M, Ismail. G. (2021). Length Biased Burr-XII Distribution: Properties and Application. *International Journal of Sciences: Basic and Applied Research*, 56(1), 222-244.

[7]. Abdelfattah Mustafa, & M. I. Khan. (2023). Length Biased Extended Rayleigh Distribution and its Applications. *Journal of Modern Applied Statistical Methods*, 22(2), 203–222.

[8]. Alzoubi, L. M. (2024). The length-biased Benrabia distribution with applications. *Electronic Journal of Applied Statistical Analysis*, 17(2), 255-277.

[9]. Narinder Pushkarna and Mustafa. (2025). Length-biased weighted Ishita distribution: Properties and applications to lifetime data. *Journal of Reliability and Statistical Studies*, 18(1), 213–242.

[10]. M. I. Khan, & Molay Kumar Ruidas. (2025). The Mgamma Distribution: Statistical Properties and Application to Real-life Data. *European Journal of Pure and Applied Mathematics*, 18(4), 1–22.