International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(4): 94-100

https://www.mathsjournal.com
Received: 03-06-2024
Accepted: 24-06-2024

AM Rangoli
Department of Statistics,
Karnatak University, Dharwad,
Karnataka, India

AS Talawar
Department of Statistics,
Karnatak University, Dharwad,
Karnataka, India

Corresponding Author:
AM Rangoli
Department of Statistics,
Karnatak University, Dharwad,
Karnataka, India

A modified Weibull distribution under competing risks

AM Rangoli and AS Talawar

DOI: https://dx.doi.org/10.22271/maths.2024.v9.i4b.1771

Abstract
In this paper, we introduced a modified Weibull distribution which is capable of handling both unimodal and bimodal scenarios. The hazard function for this distribution exhibits versatile behaviors, including increasing, decreasing, bathtub-shaped, and increasing-decreasing-increasing patterns. This modified distribution is applied in a competing risks framework with two causes of failure. Maximum Likelihood Estimation (MLE) is used for parameter estimation, and due to the lack of explicit solutions, we employed the Newton-Raphson algorithm. To validate the modified Weibull distribution, we conducted a simulation study and analysed a real-life dataset of cancer patients.

Keywords: Modified Weibull distribution, MLE, simulation, competing risks, bimodal

1. Introduction
In many medical or engineering applications, Weibull distribution plays an important role because of its flexible hazard functions, i.e., increasing, decreasing and constant hazard rates. In survival or reliability studies, more than one cause of failure may be directed to a subject at a time, this concept is considered as Competing Risks approach. The competing risks data consists of failure time and cause of failure. Sometimes subjects may be censored. Many authors had given the modification of Weibull distribution in many ways depending upon the situation and datasets, *viz*., Mudholkar et al. (1995), Mudholkar and Hutson (1996), Xie and Lai (1996), Chen (2000), Xie et al. (2002), Lai et al. (2003), Carrasco et al. (2008), Sarhan and Zaindin (2009), Almalki and Yuan (2013), Bebbington et al. (2007), Almalki and Nadarajah (2014), Rangoli and Talawar (2024b) [10, 11, 15, 6, 16, 8, 5, 14, 2, 3, 1, 13]. These modifications work in a unimodal scenario. Rangoli and Talawar (2024a) [12] considered a mixture of two Weibull distribution and Kernel methods for estimation of hazard and survival functions in competing risks scenarios for a bimodal dataset. Bebbington et al. (2007) [3] gave the modification of Weibull distribution within two parameters, which does not include any shape parameter. In this paper we brought modification to the Bebbington et al. (2007) [3] model by adding one shape parameter that works in unimodal as well as bimodal scenarios and has decreasing unimodal and bimodal density curves. Also, this modified Weibull distribution has increasing, decreasing, bathtub, increasing-decreasing-increasing hazard curves.
This paper is organized as follows: We introduce the modified Weibull distribution in Section 2. Order statistics is considered in section 3. The maximum likelihood estimators are provided in Section 4. In Section 5, we discuss simulation studies and real-life application. Finally, Section 6 deals with concluding remarks.

2. Modified Weibull distribution
Under the assumption that there are k causes of failure, we have the following notations
t_i → Lifetime of the i$^{\text{th}}$ unit
t_{ij} →Lifetime of the i$^{\text{th}}$ individual for cause $j, j = 1,2, \dots k$
$F(.)$ →Cummulative Distribution Function (CDF) of t_i
α and γ are scale parameters and β is a shape parameter
$f(.)$ → Probability Density Function (PDF) of t_i
$S(.)$ →Survival Function of t_i
$\lambda(t)$ → Hazard Function of t_i

$F_j(.)$ →Cummulative Distribution Function (CDF) of t_{ij}
$f_j(.)$ → Probability Density Function (PDF) of t_{ij}
$\lambda_{ij}(t)$ → Cause Specific Hazard Function of t_{ij}

Let the CDF is,

$$F(t) = 1 - e^{-e^{\alpha t^{\beta} - \frac{\gamma}{t}}} \text{where } \alpha > 0, \beta > 0, \gamma > 0, t > 0 \quad (1)$$

The PDF corresponding to (1),

$$f(t) = \left(\alpha\, \beta\, t^{\beta-1} + \frac{\gamma}{t^2}\right) e^{\alpha t^{\beta} - \frac{\gamma}{t}}\, e^{-e^{\alpha t^{\beta} - \frac{\gamma}{t}}} \quad (2)$$

Form of the hazard function is,

$$\lambda(t) = \left(\alpha\, \beta\, t^{\beta-1} + \frac{\gamma}{t^2}\right) e^{\alpha t^{\beta} - \frac{\gamma}{t}} \quad (3)$$

And form of the survival function is,

$$s(t) = e^{-e^{\alpha t^{\beta} - \frac{\gamma}{t}}} \quad (4)$$

From Figure 1 and 2 we can see that the density and hazard curves for different values of the parameters. This modified Weibull density tends to Bebbington et al. (2007) [3] modified Weibull distribution when β ≅ 1.

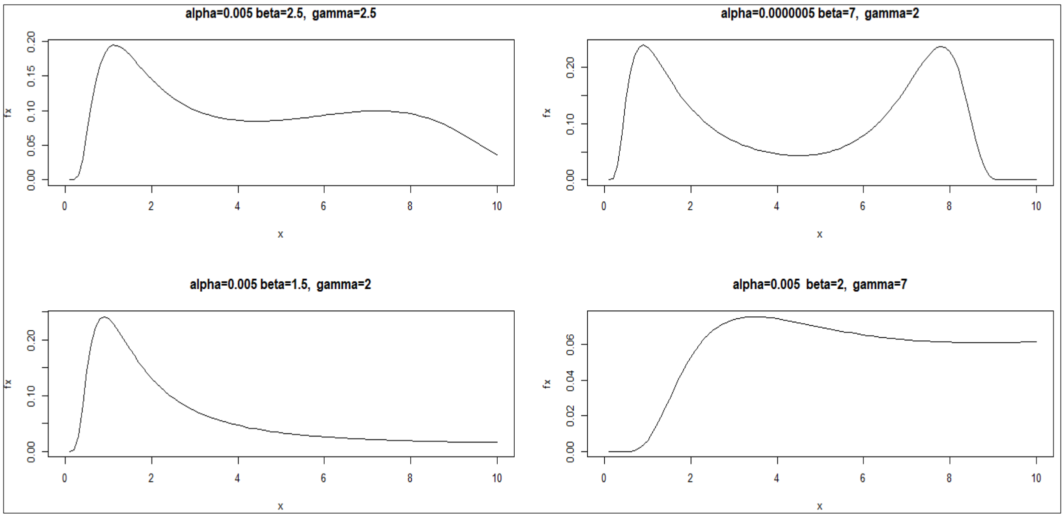

Fig 1: Density curves for different values of parameters.

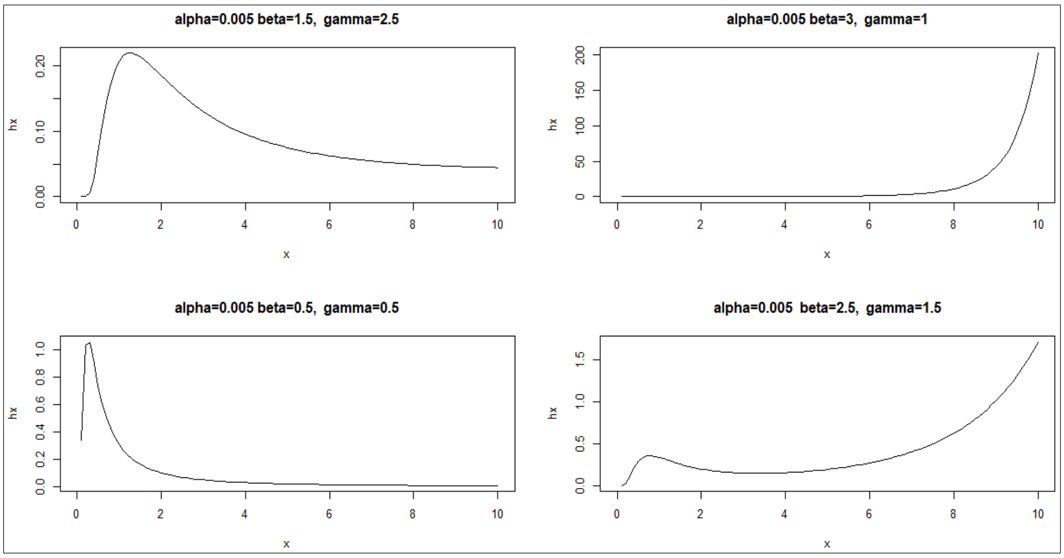

Fig 2: Hazard curves for different values of parameters.

3. Order Statistics

Let $T_1, T_2, \ldots, T_n$ be a random sample of size n from modified Weibull distribution with pdf and cdf defined in (2) and (1) respectively. We give the density of r^{th} order statistic as

$$f_r(t) = \frac{n!}{(r-1)!(n-r)!}\left(1 - e^{-e^{\alpha t^{\beta} - \frac{\gamma}{t}}}\right)^{r-1}\left(e^{-e^{\alpha t^{\beta} - \frac{\gamma}{t}}}\right)^{n-r+1}\left(\alpha\,\beta\,t^{\beta-1} + \frac{\gamma}{t^2}\right)e^{\alpha t^{\beta} - \frac{\gamma}{t}}$$

The density of I^{st} order statistic is given by,

$$f_1(t) = n\left(e^{-e^{\alpha t^{\beta} - \frac{\gamma}{t}}}\right)\left(\alpha\,\beta\,t^{\beta-1} + \frac{\gamma}{t^2}\right)e^{\alpha t^{\beta} - \frac{\gamma}{t}}$$

And the density of n^{th} order statistic is given by,

$$f_n(t) = n\left(1 - e^{-e^{\alpha t^{\beta} - \frac{\gamma}{t}}}\right)^{n-1}\left(e^{-e^{\alpha t^{\beta} - \frac{\gamma}{t}}}\right)\left(\alpha\,\beta\,t^{\beta-1} + \frac{\gamma}{t^2}\right)e^{\alpha t^{\beta} - \frac{\gamma}{t}}$$

4. Maximum Likelihood Estimation

The maximum likelihood estimation (MLE) is widely used among the statistical inference because of its desirable properties like consistency, asymptotic efficiency. Use of maximum likelihood estimation in competing risk set up is different from general approach as illustrated below.

We assume that T_{ij} to follows modified Weibull random variables with parameters $(\alpha_j, \beta_j, \gamma_j)$ $i = 1,2, \ldots n$ $j = 1,2, \ldots k$.

Let $T_1, T_2, \ldots, T_n$ are failure times and the subjects may fail due to various causes. Let us assume that there are k causes which acts independently on the subjects. Let C be the censoring time such that, there will be available only time but not cause of failure. The Likelihood function is given by,

$$L = \prod_{j=1}^{k}\prod_{i=1}^{n}(f_j(t_i))^{\delta_{ij}}\,S(t_i)^{1-\delta_{ij}} \tag{5}$$

$$L = \prod_{j=1}^{k}\prod_{i=1}^{n_j}\left(\lambda_j(t_i)\right)^{\delta_{ij}}S(t_i) \tag{6}$$

$$\delta_{ij} = \begin{cases} 1 \; if \; T_{ij} < C_{ij} \\ 0 \; otherwise \end{cases}$$

The log likelihood can be written as

$$l = logL = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\left(\delta_{ij} * log\left(\lambda_j(t_i)\right)\right) + log\left(S(t_i)\right)\right)$$

$$l = logL = \sum_{j=1}^{k}\left(\sum_{i=1}^{n_j}\left(\delta_{ij} * log\left(\lambda_j(t_i)\right)\right) + \sum_{i=1}^{n} log\left(S(t_i)\right)\right)$$

Now for cause j, the log likelihood is given as, $l_j = \sum_{i=1}^{n_j} log\left(\lambda_j(t_i)\right) + \sum_{i=1}^{n} log\left(S(t_i)\right)$

$$l_j = \sum_{i=1}^{n_j} log\left(\left(\alpha_j\,\beta_j\,t_i^{\beta_j-1} + \frac{\gamma_j}{t_i^2}\right)e^{\alpha_j t_i^{\beta_j} - \frac{\gamma_j}{t_i}}\right) + \sum_{i=1}^{n} log\left(e^{-e^{\alpha_j t_i^{\beta_j} - \frac{\gamma_j}{t_i}}}\right)$$

$$l_j = \sum_{i=1}^{n_j} log\left(\alpha_j\,\beta_j\,t_i^{\beta_j-1} + \frac{\gamma_j}{t_i^2}\right) + \sum_{i=1}^{n_j}\alpha_j\,t_i^{\beta_j} - \sum_{i=1}^{n_j}\frac{\gamma_j}{t_i} - \sum_{i=1}^{n} e^{\alpha_j t_i^{\beta_j} - \frac{\gamma_j}{t_i}}$$

$$\frac{\partial logl_j}{\partial \alpha_j} = \sum_{i=1}^{n_j}\frac{\beta_j\,t_i^{\beta_j-1}}{\alpha_j\,\beta_j\,t_i^{\beta_j-1} + \frac{\gamma_j}{t_i^2}} + \sum_{i=1}^{n_j} t_i^{\beta_j} - \sum_{i=1}^{n} e^{\alpha_j t_i^{\beta_j} - \frac{\gamma_j}{t_i}}\,t_i^{\beta_j}$$

$$\frac{\partial logl_j}{\partial \beta_j} = \sum_{i=1}^{n_j}\frac{\alpha_j\,t_i^{(\beta_j-1)}(\beta_j\log t_i + 1)}{\alpha_j\,\beta_j\,t_i^{\beta_j-1} + \frac{\gamma_j}{t_i^2}} + \sum_{i=1}^{n_j}\alpha_j\,t_i^{\beta_j}\log t_i - \sum_{i=1}^{n} e^{\alpha_j t_i^{\beta_j} - \frac{\gamma_j}{t_i}}\,\alpha_j\,t_i^{\beta_j}\log t_i$$

$$\frac{\partial logl_j}{\partial \gamma_j} = \sum_{i=1}^{n_j}\frac{1}{\alpha_j\,\beta_j\,t_i^{\beta_j-1} + \frac{\gamma_j}{t_i^2}}\,\frac{1}{t_i^2} - \sum_{i=1}^{n_j}\frac{1}{t_i} + \sum_{i=1}^{n}\frac{e^{\alpha_j t_i^{\beta_j} - \frac{\gamma_j}{t_i}}}{t_i}$$

The second order derivatives w.r.t α_j, β_j and γ_j are given in Appendix.
Let $I(\Theta)$ be Fisher Information Matrix of the unknown parameters. $\Theta = (\alpha_j, \beta_j, \gamma_j)$ $j = 1,2, \dots k$ parameters for modified Weibull distribution. The elements of the matrix $I_{ij}(\Theta), i, j = 1,2, \dots k$ can be approximated by $I_{ij}(\hat{\Theta})$, where

$$I_{ij}(\hat{\Theta}) = -\left.\frac{\partial^2 l(\Theta)}{\partial\Theta_i \partial\Theta_j}\right| \text{evaluated at } \Theta = \hat{\Theta}$$

$$I_{ij}(\hat{\Theta}) = \begin{bmatrix} \frac{\partial^2 logl_j}{\partial\alpha_j^2} & \frac{\partial^2 logl_j}{\partial\alpha_j\beta_j} & \frac{\partial^2 logl_j}{\partial\alpha_j\gamma_j} \\ \frac{\partial^2 logl_j}{\partial\beta_j\alpha_j} & \frac{\partial^2 logl_j}{\partial\beta_j^2} & \frac{\partial^2 logl_j}{\partial\beta_j\gamma_j} \\ \frac{\partial^2 logl_j}{\partial\gamma_j\alpha_j} & \frac{\partial^2 logl_j}{\partial\gamma_j\beta_j} & \frac{\partial^2 logl_j}{\partial\gamma_j^2} \end{bmatrix} j = 1,2, \dots, k \qquad (7)$$

The elements of the above matrix were given in Appendix. These equations do not have explicit solutions to estimate the parameter values, so we consider Newton-Raphson method for numerical solution.

4.1 Asymptotic Confidence Bounds

To calculate confidence intervals, we consider asymptotic distribution of the MLE of the parameters (Lawless, 2003 [9]). It is known that the asymptotic distribution of the MLE $\hat{\Theta}$ is given by

$$(\hat{\Theta} - \Theta) \rightarrow N_3(0, I^{-1}(\Theta))$$

Where $\Theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2)$.
The elements of the 3 X 3 matrix of $I^{-1}(.)$, are approximated by $I_{ij}(\hat{\Theta})$,
Therefore, the approximate $100(1-\gamma)\%$ two-sided, confidence interval for Θ is given by

$$\hat{\Theta} \pm Z_{\delta/2}\sqrt{I^{-1}(\hat{\Theta})}$$

Here $Z_{\delta/2}$ is the upper $\delta/2$ th percentile of a standard normal distribution.

4.2 Information Criterion:

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) methods are used to know which of the following distributions fits the data well (Lawless, 2003 [9]). The distribution fits the data well, whose AIC and BIC values are less.

$$AIC = 2K - 2\ lnL$$

$$BIC = Kln(n) - 2\ lnL$$

Where L is the likelihood function, n is the sample size and K is the number of parameters estimated.

5. Results and Discussion

To validate the model we considered the two dataset; Simulated dataset (bimodal) and real life cancer patients dataset (unimodal).

5.1 Simulation Study

To validate the model, we simulated the bimodal dataset using Monte Carlo simulation algorithm. Using the algorithm and with appropriate choice of the parameters, we have simulated 1000 observations. Then by considering these simulated datasets, we generated the competing risks data using the algorithm given below (Beysermann et al., 2001 [4]).

- T be event time generated by Monte Carlo algorithm.
- Run the binomial test for assigning the cause i.e. generating failure cause j,
- $j = 1,2,3, \dots, J$.
- Generate censored observation C. In this case we have used exponential distribution to generate censored observations.
- Now considering min (T, C), and assigning cause if we get T else it will be considered as censored.

In this simulation we have considered the two causes of failure and generated the failure times and the censored times.
Table 1 shows the estimated values using MLE for simulated dataset. Figure 3 shows the density curve of the simulated dataset, from this we can observe that this modified Weibull distribution has bimodal shape. From Figure 4 we can observe the hazard and survival curves of two causes of failure for the simulated dataset. For cause 1 the hazard function is increasing, slight decreasing then again steady increasing. For cause 2 the hazard curve is increasing then decreasing. The survival probability reaches 50% before 5 months for cause 1 whereas for cause 2 it reaches after 10 months.

5.2 Application to cancer data

To validate the model, we have considered the real life data of survival times of cancer patients given in Crowder (2001) [7]. The dataset consists of two causes of failure and consists of censored observations. Table 2 shows the estimated values of the parameters for the cancer dataset. From Table 3 we found that our modified Weibull distribution gives good fit by observing the AIC and BIC values. From Figure 5 we can observe that for cause 1 and cause 2 the hazard curve becomes almost same. Table 4 shows the estimated parameter values, standard error, lower confidence limit and upper confidence limit.

Table 1: Estimated value for the simulated data.

Parameter\Causes	Overall	Cause I	Cause II
α	0.00061604334	0.0001777888	$1.92593*10^{-34}$
β	3.602037	3.9060763989	0.3894937
γ	2.350242	2.2419348289	4.036324

Table 2: Estimated values for the cancer dataset.

Parameter\Causes	Overall	Cause I	Cause II
α	0.001118968	0.002592299	$1.104887*10^{-06}$
β	1.355253491	1.15865758	2.367479
γ	6.676638214	11.958162285	8.160834

Table 3: AIC and BIC values for cancer data.

Distributions	$logl$	AIC	BIC
Bebbington	-686.413	1380.826	1392.009
Modified Bebbington	-665.0875	1342.176	1358.95

Table 4: Estimated parameter value, standard error (SE), lower control limit (LCL) and upper control limit (UCL).

Causes	Parameters	Standard Error	LCL	UCL
Cause 1	α_1=0.002592299	0.000122188	0.002352815	0.002831783
	β_1=1.15865758	0.283940760	0.459352091	1.572379425
	γ_1=11.958162285	1.498186498	9.021770684	14.894553886
Cause 2	α_2=$1.104887*10^{-06}$	$4.502045*10^{-8}$	$1.016649*10^{-6}$	$1.193125*10^{-6}$
	β_2=2.367479	0.7821883	0.8344180	3.900540
	γ_2=8.160834	1.873543	4.488757	11.83291

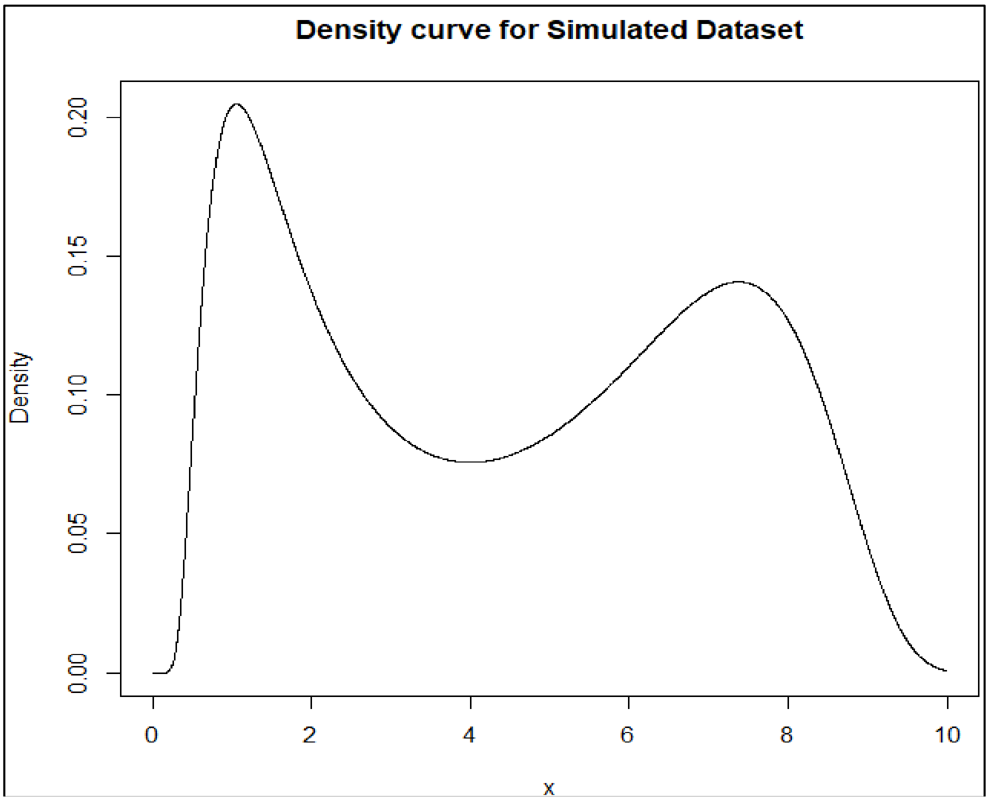

Fig 3: Density curve of the simulated dataset.

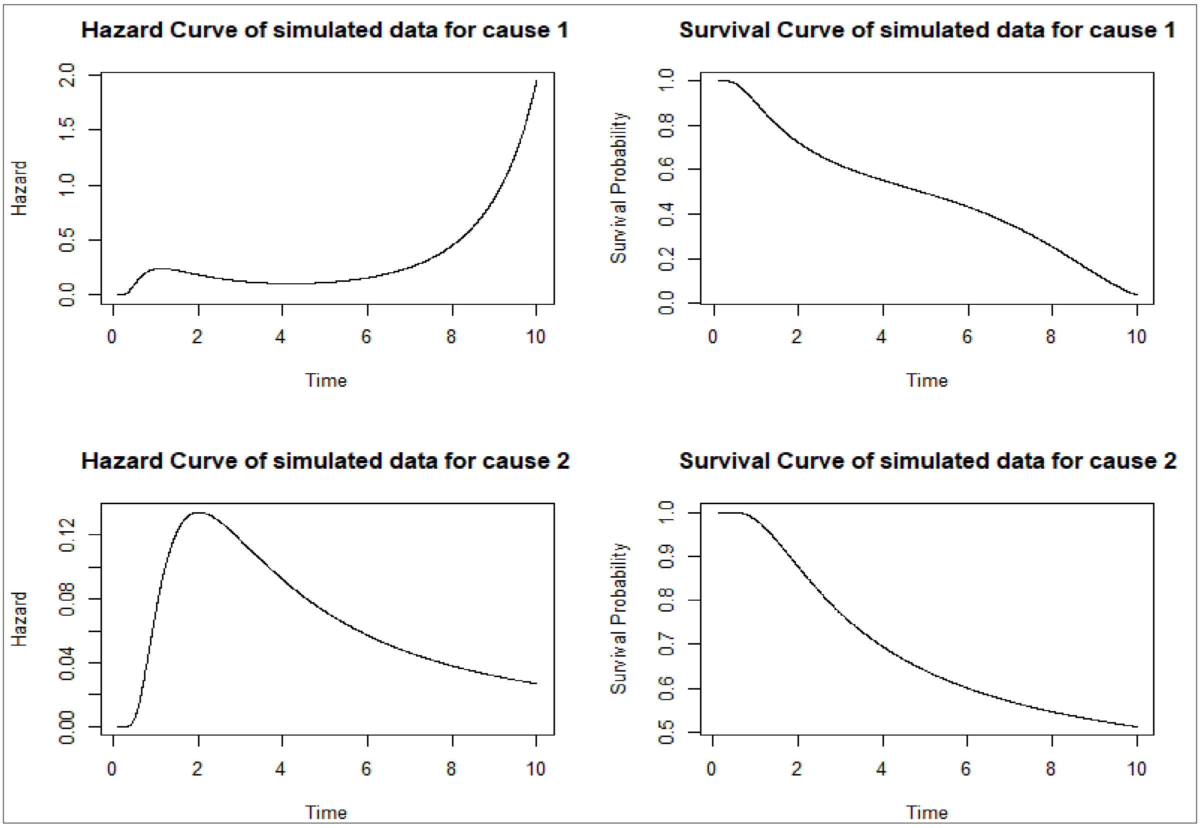

Fig 4: Hazard and Survival curves of 2 causes for the simulated dataset.

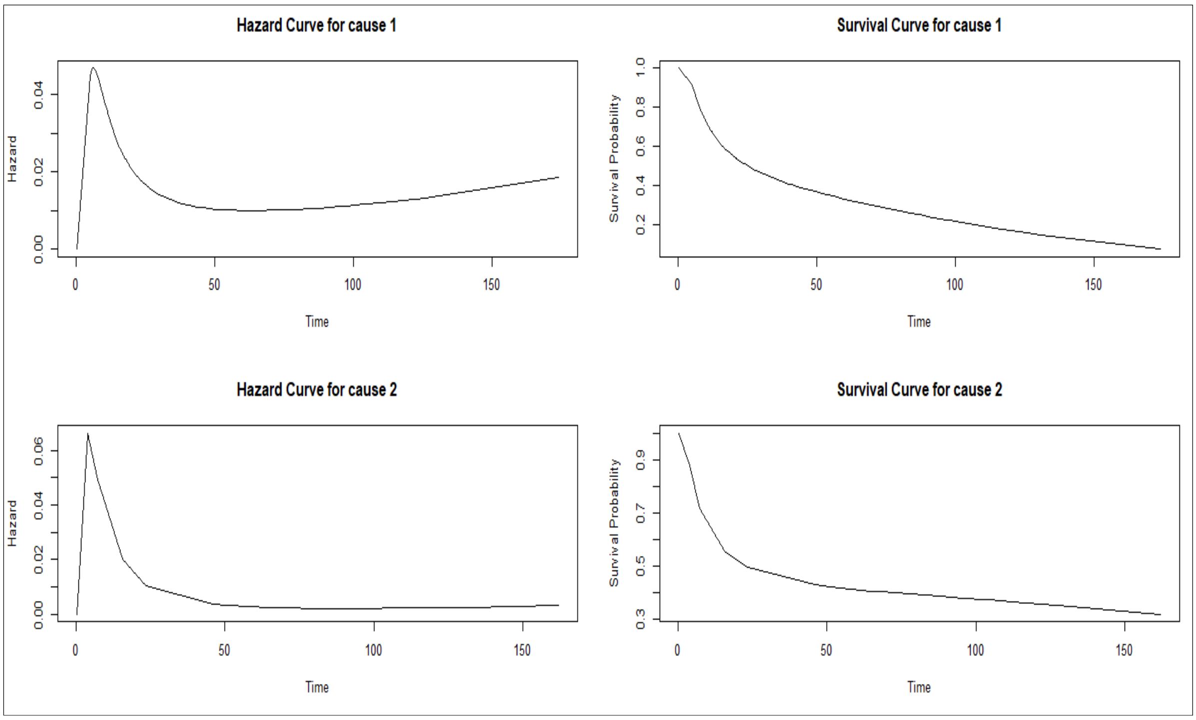

Fig 5: Hazard and survival curves of 2 causes for cancer dataset.

6. Conclusions

The modified Weibull distribution can effectively address both unimodal and bimodal scenarios with one shape parameter. The distribution of the hazard function demonstrates different patterns, including increasing, decreasing, bathtub-shaped, and increasing-decreasing-increasing behaviors, making it highly versatile for various applications. We applied this modified distribution within a competing risks framework involving two causes of failure and demonstrated its practical utility. The simulation studies and the real life dataset justify the application of this modified Weibull distribution in both unimodal and bimodal scenarios. Parameters are estimated for the cancer dataset including the standard error, lower confidence limit and upper confidence limit.

7. Acknowledgment: The first author is thankful to Department of Science and Technology, innovation in science pursuit for inspired research (DST-INSPIRE) for financial support (Fellowship/2021/210203).

8. References

1. Almalki SJ, Nadarajah S. A new discrete modified Weibull distribution. IEEE Transactions on Reliability. 2014;63(1):68-80.
2. Almalki SJ, Yuan J. A new modified Weibull distribution. Reliability Engineering and System Safety. 2013;111:164-170.
3. Bebbington M, Lai CD, Zitikis R. A flexible Weibull extension. Reliability Engineering and System Safety. 2007;92(6):719-726.
4. Beyersmann J, Allignol A, Schumacher M. Competing Risks and Multistate Models with R. Springer Science and Business Media; 2011.
5. Carrasco JM, Ortega EM, Cordeiro GM. A generalized modified Weibull distribution for lifetime modeling. Computational Statistics and Data Analysis. 2008;53(2):450-462.
6. Chen Z. A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. Statistics and Probability Letters. 2000;49(2):155-161.
7. Crowder MJ. Classical Competing Risks. Chapman and Hall/CRC; c2001.
8. Lai CD, Xie M, Murthy DNP. A modified Weibull distribution. IEEE Transactions on Reliability. 2003;52(1):33-37.
9. Lawless JF. Statistical Models and Methods for Lifetime Data. John Wiley and Sons; 2003.
10. Mudholkar GS, Srivastava DK, Freimer M. The exponentiated Weibull family: A reanalysis of the bus-motor-failure data. Technometrics. 1995;37(4):436-445.
11. Mudholkar GS, Hutson AD. The exponentiated Weibull family: Some properties and a flood data application. Communications in Statistics-Theory and Methods. 1996;25(12):3059-3083.
12. Rangoli AM, Talawar AS. Estimation of hazard and survival function for competing risks using kernel and mixture model in bimodal setup. Reliability: Theory and Applications. 2024;19(2):454-466.
13. Rangoli AM, Talawar AS. Inference on modified Weibull type distribution and its application to competing risks data using MCMC. International Journal of Agricultural and Statistical Science. In Press.
14. Sarhan AM, Zaindin M. Modified Weibull distribution. APPS. Applied Sciences. 2009;11:123-136.
15. Xie M, Lai CD. Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. Reliability Engineering and System Safety. 1996;52(1):87-93.
16. Xie M, Tang Y, Goh TN. A modified Weibull extension with bathtub-shaped failure rate function. Reliability Engineering and System Safety. 2002;76(3):279-285.

Appendix

$$nume\beta = \alpha_j\, t_i^{(\beta_j - 1)}(\beta_j \log t_i + 1)$$

$$deno = \alpha_j\, \beta_j\, t_i^{\beta_j - 1} + \frac{\gamma_j}{t_i^2}$$

$$\frac{\partial^2 logl_j}{\partial \alpha_j^2} = -\sum_{i=1}^{n_j} \frac{\left(\beta_j\, t_i^{\beta_j - 1}\right)^2}{\left(\alpha_j\, \beta_j\, t_i^{\beta_j - 1} + \frac{\gamma_j}{t_i^2}\right)^2} - \sum_{i=1}^{n} e^{\alpha_j\, t_i^{\beta_j} - \frac{\gamma_j}{t_i}}\ t_i^{2\,\beta_j}$$

$$\frac{\partial^2 logl_j}{\partial \beta_j^2} = \sum_{i=1}^{n_j} \frac{\left(deno\ \alpha_j\, t_i^{\beta_j - 1} \log t_i(\beta_j \log t_i + 2)\right) - (nume\beta^2)}{deno^2} + \sum_{i=1}^{n_j} \alpha_j\, t_i^{\beta_j} (\log t_i)^2 - \sum_{i=1}^{n} \alpha_j\, (\log t_i)^2\, t_i^{\beta_j} e^{\alpha_j\, t_i^{\beta_j} - \frac{\gamma_j}{t_i}} (\ \ {}_j t_i^{\beta_j} + 1)$$

$$\frac{\partial^2 logl_j}{\partial \gamma_j^2} = -\sum_{i=1}^{n_j} \frac{1}{\left(\alpha_j\, \beta_j\, t_i^{\beta_j - 1} + \frac{\gamma_j}{t_i^2}\right)^2} \frac{1}{x^4} - \sum_{i=1}^{n} \frac{e^{\alpha_j\, t_i^{\beta_j} - \frac{\gamma_j}{t_i}}}{t_i^2}$$

$$\frac{\partial^2 logl_j}{\partial \alpha_j \beta_j} = \frac{\partial^2 logl_j}{\partial \beta_j \alpha_j} = \sum_{i=1}^{n_j} \frac{(deno)\ t_i^{\beta_j - 1}\left(\beta_j \log(t_i) + 1\right) - \beta\ t_i^{\beta_j - 1}\ (nume\beta)}{\left(\alpha_j\, \beta_j\, t_i^{\beta_j - 1} + \frac{\gamma_j}{t_i^2}\right)^2} + \sum_{i=1}^{n_j} t_i^{\beta_j} \log(t_i) - \sum_{i=1}^{n_j} \left(\alpha_j\, t_i^{\beta_j} - 1\right) e^{\alpha_j\, t_i^{\beta_j} - \frac{\gamma_j}{t_i}}\, t_i^{\beta_j} \log(t_i)$$

$$\frac{\partial^2 logl_j}{\partial \alpha_j \gamma_j} = \frac{\partial^2 logl_j}{\partial \gamma_j \alpha_j} = -\sum_{i=1}^{n_j} \frac{\beta_j\, t_i^{\beta_j - 1}}{\left(\alpha_j\, \beta_j\, t_i^{\beta_j - 1} + \frac{\gamma_j}{t_i^2}\right)^2} \frac{1}{t_i^2} + \sum_{i=1}^{n} t_i^{\beta_j - 1}\ e^{\alpha_j\, t_i^{\beta_j} - \frac{\gamma_j}{t_i}}$$

$$\frac{\partial^2 logl_j}{\partial \beta_j \gamma_j} = \frac{\partial^2 logl_j}{\partial \gamma_j \beta_j} = -\sum_{i=1}^{n_j} \frac{nume\beta}{\left(\alpha_j\, \beta_j\, t_i^{\beta_j - 1} + \frac{\gamma_j}{t_i^2}\right)^2} \frac{1}{t_i^2} + \sum_{i=1}^{n} \log(t_i)\, \alpha_j\, t_i^{\beta_j - 1}\ e^{\alpha_j\, t_i^{\beta_j} - \frac{\gamma_j}{t_i}}$$