A New Generalized Weibull - Gamma Frailty Model in Survival Analysis

[1]G. Prabavathi, [2]R. Elangovan

[1,2]Department of Statistics, Annamalai University
Chidambaram, Tamil Nadu-608002

Abstract: In this study, we have proposed a new generalization of weibull-gamma frailty model is an expanded version of the classical frailty model. The conditional survival function given the frailty is directly modeled in this study as opposed to the standard frailty model, which takes into account modeling of the hazard function. The maximum likelihood technique is used in a simulation study with the E-M algorithm. Then we can apply the suggested model to kidney infection data.

Keywords: Weibull-gamma distribution, frailty model, E-M algorithm, Survival Analysis

1. Introduction

The frailty models in survival analysis presents an extensive overview of the fundamental approaches in the area of frailty model. Frailty model gives a simple way to introduce unobserved heterogeneity and related into models for survival data. In this model examine an unobserved random proportionality factor that changes the hazard function of an individual or related individuals. Various authors are discussed to analyze the frailty model in survival analysis. Beard (1959) presented the first univariate frailty model and later Vaupel et al. (1979) introduced the frailty term in univariate survival models after Clayton (1978) considerably promote its applications to multivariate survival data. Hougaard (1986) discussed a class of multivariate failure time distribution and Kayid et al. (2019) considered a proportional reversed hazards weighted frailty model. Shanubhogue et al. (2017) proposed a new generalization of weibull-exponential frailty model. Balakrishnan et al. (2018) developed a semi parametric likelihood inference for Birnbaum Saunders frailty model in the analysis of real life data. Shanubhogue et al. (2019) proposed a new generalized gamma-exponential frailty model in survival analysis. David (2020) suggested a new correlated inverse Gaussian frailty model with linear failure rate distribution as a baseline distribution that is the best model for analyzing of kidney infection data. Balan et al. (2020) studied that frailty models for survival outcomes and implies that how shared frailty is used to modeled positively dependent outcome in survival data. Alex mota et.al (2021) introduce a weighted Linley frailty model its estimation and application to lung cancer data. Nagaraj et al. (2022) presented to fit Lindley distribution with four frailty models.

In this study, section 2 we first introduce a normal generalized weibull-gamma frailty model and then we estimated a new generalized weibull-gamma frailty model and its properties. In section 3, we discussed the maximum likelihood technique is done with E-M algorithm for estimating the parameters of the proposed model. In section 4, we discussed the simulation study and results. In section 5, we suggested the fitting of model on data in support of the proposed model. In section 6, we discussed the conclusion of the proposed model.

2. A New Generalized Weibull-Gamma Frailty Model

The classical and widely used frailty model makes the assumption that a proportional hazards model, which is dependent on the random effect (frailty) that is, the risk of an increased reliance on an unobservable. The baseline hazard function is multiplicatively affected random variable z. In frailty models, the hazard function is proportional to hazard function of the baseline distribution with constant proportionality as frailty random variable z.

$$h(t) = zh_0(t) \tag{1}$$

In this frailty model, we consider a conditional survival function given as frailty random variable z with δ being unknown parameter, by which we can extent the present frailty model

$$h(t) = z^{\delta} h_0(t) \tag{2}$$

The conditional distribution of T given the frailty z is

$$f\left(t/z\right) = z^{\delta} h_0(t) e^{-z^{\delta}\int_0^t h_0(u)\,du} \tag{3}$$

Then the hazard function of the weibull distribution can be obtained as

$$h_0(t) = \frac{f_0(t,\lambda,\gamma)}{S_0(t)}$$

$$h_0(t) = \lambda\gamma\, t^{\gamma-1} \tag{4}$$

Using equation (4) in equation (3) and we get,

$$f\left(t/z\right) = z^{\delta} \lambda\gamma\, t^{\gamma-1} e^{-z^{\delta}\int_0^t \lambda\gamma\, u^{\gamma-1}\,du}$$

$$f\left(t/z\right)=z^{\delta}\lambda t^{\gamma-1}e^{-z^{\delta}}\lambda\gamma\int_{0}^{t}u^{\gamma-1}\,du$$

$$f\left(t/z\right)=z^{\delta}\lambda t^{\gamma-1}e^{-z^{\delta}}\lambda\gamma\, t^{r}$$

$$f\left(t/z\right)=z^{\delta}\lambda t^{\gamma-1}e^{-z^{\delta}\lambda t^{\gamma}} \tag{5}$$

Mean of the above model in equation (5)

$$E\left(T/z\right)=\int_{0}^{\infty}t\, f\left(t/z\right)dt$$

$$E\left(T/z\right)=\int_{0}^{\infty}t\, z^{\delta}\lambda t^{\gamma-1}e^{-z^{\delta}\lambda t^{\gamma}}\;dt$$

$$E\left(T/z\right)=z^{\delta}\lambda\int_{0}^{\infty}t^{\gamma+1-1}e^{-z^{\delta}\lambda t^{\gamma}}\;dt$$

$$E\left(T/z\right)=\frac{z^{\delta}\lambda}{z^{\delta}\lambda\gamma}\int_{0}^{\infty}\left(\frac{q}{z^{\delta}\lambda}\right)^{\frac{1}{\gamma}}e^{-q}dq$$

$$E\left(T/z\right)=\frac{1}{\gamma}\left(\frac{1}{z^{\delta}\lambda}\right)^{\frac{1}{\gamma}}\int_{0}^{\infty}q^{\left(\frac{1}{\gamma}+1\right)-1}e^{-q}dq$$

$$E\left(T/z\right)=\frac{1}{\gamma\left(z^{\delta}\lambda\right)^{\frac{1}{\gamma}}}\Gamma\frac{1}{\gamma}+1$$

Variance of model is given by

$$E\left(T^{2}/z\right)=\int_{0}^{\infty}t^{2}\; f\left(t/z\right)dt$$

$$E\left(T^{2}/z\right)=\int_{0}^{\infty}t^{2}\; z^{\delta}\lambda t^{\gamma-1}e^{-z^{\delta}\lambda t^{\gamma}}\;dt$$

$$E\left(T^{2}/z\right)=z^{\delta}\lambda\int_{0}^{\infty}t^{\gamma+2-1}\,e^{-z^{\delta}\lambda t^{\gamma}}\;dt$$

Put $z^{\delta}\lambda t^{\gamma}=q,\quad z^{\delta}\lambda\gamma\, t^{\gamma-1}=dq \qquad t^{\gamma}=\frac{q}{z^{\delta}\lambda}\Rightarrow t=\left(\frac{q}{z^{\delta}\lambda}\right)^{\frac{1}{\gamma}}$

$$E\left(T^{2}/z\right)=\int_{0}^{\infty}\left(\frac{q}{z^{\delta}\lambda}\right)^{\frac{2}{\gamma}}e^{-q}dq$$

$$E\left(T^{2}/z\right)=\frac{1}{\gamma}\left(\frac{1}{z^{\delta}\lambda}\right)^{\frac{2}{\gamma}}\int_{0}^{\infty}q^{\frac{2}{\gamma}}e^{-q}dq$$

$$E\left(T^{2}/z\right)=\frac{1}{\gamma}\left(\frac{1}{z^{\delta}\lambda}\right)^{\frac{2}{\gamma}}\int_{0}^{\infty}q^{\left(\frac{2}{\gamma}+1\right)-1}e^{-q}dq \tag{6}$$

Using gamma function to above equation (6), we get

$$E\left(T^{2}/z\right)=\frac{1}{\gamma}\left(\frac{1}{z^{\delta}\lambda}\right)^{\frac{2}{\gamma}}\Gamma\frac{2}{\gamma}+1$$

$$E\left(T^2/z\right)=\frac{\Gamma\frac{2}{\gamma}+1}{\gamma(z^\delta\lambda)^{\frac{2}{\gamma}}}$$

$$\text{Variance }=E\left(\frac{T^2}{z}\right)-\left(E\left(\frac{T^2}{z}\right)\right)^2$$

$$V\left(\frac{T}{z}\right)=\frac{\Gamma\frac{2}{\gamma}+1}{\gamma(z^\delta\lambda)^{\frac{2}{\gamma}}}-\left(\frac{\Gamma\frac{1}{\gamma}+1}{\gamma(z^\delta\lambda)^{\frac{1}{\gamma}}}\right)^2$$

$$V\left(\frac{T}{z}\right)=\frac{1}{\gamma(z^\delta\lambda)^{\frac{2}{\gamma}}}\left(\Gamma\frac{2}{\gamma}+1-\left(\Gamma\frac{1}{\gamma}+1\right)^2\right)$$

$$V\left(\frac{T}{z}\right)=\frac{\Gamma\frac{2}{\gamma}+1-\left(\Gamma\frac{1}{\gamma}+1\right)^2}{\gamma(z^\delta\lambda)^{\frac{2}{\gamma}}}$$

The gamma distribution with two parameters is considered as frailty distribution with $\delta>0$ and assumes that mean $E(z)=1$ and variance $g(z)$ the variance $V(g(z))=\lambda$ it can be estimated from real life data.

$$g(z)=\frac{\lambda^k t^{k-1}e^{-\lambda t}}{\Gamma k},\qquad t>0,\quad \lambda,k>0 \tag{7}$$

Then the joint distribution of t and z is

$$f\left(t,z\right)=z^\delta h_0(t)\,S\left(\frac{t}{z}\right)g(z) \tag{8}$$

$$f\left(t,z\right)=z^\delta\lambda\gamma\, t^{\gamma-1}e^{-z^\delta\gamma t^\gamma}\ \frac{\lambda^k t^{k-1}e^{-\lambda t}}{\Gamma k}$$

$$f\left(t,z,\lambda,\delta,\gamma,k\right)=\frac{z^\delta\lambda^{k+1}\gamma\, t^{k+r-2}e^{-(z^\delta t^\gamma+t)\lambda}}{\Gamma k} \tag{9}$$

Hence, the marginal distribution of T is

$$f(t)=h_0(t)S\left(\frac{t}{z}\right)g(z)dz \tag{10}$$

$$f(t)=\frac{\lambda\gamma\, t^{\gamma-1}}{\Gamma k}\int_0^\infty z^\delta e^{-z^\delta\lambda t^\gamma}\lambda^k\, t^{k-1}e^{-\lambda t}$$

$$f\left(t\right)=\frac{\lambda^{k+1}\gamma\, t^{k+r-2}e^{-\lambda}}{\Gamma k}\int_0^\infty z^\delta e^{-z^\delta\lambda t^\gamma}\, dz \tag{11}$$

Solving equation (10) and we get,

Put $z^\delta\lambda t^\gamma=P\quad\Rightarrow z=\left(\frac{P}{\lambda t^\gamma}\right)^{\frac{1}{\delta}}$

$$\delta\, z^{\delta-1}\lambda t^\gamma dz=dp$$

$$z^{\delta-1}\lambda t^\gamma dz=\frac{1}{\delta}dp$$

$$f(t)=\frac{\gamma\lambda t^{\gamma-1}}{\Gamma k}\int_0^\infty z\, z^{\delta-1}e^{-z^\delta}\lambda t^\gamma\lambda^k t^{k-1}e^{-\lambda t}dz$$

$$f(t)=\frac{\gamma t^{-1}}{\Gamma k}\int_0^\infty\left(\frac{P}{\lambda t^{\gamma}}\right)^{\frac{1}{\delta}}e^{-P}\lambda^k t^{k-1}e^{-\lambda t}\frac{dp}{\delta}$$

$$f(t)=\frac{\gamma t^{k-2}\lambda^k e^{-\lambda t}}{\delta\Gamma k(\lambda t^{\gamma})^{\frac{1}{\delta}}}\int_0^\infty P^{\frac{1}{\delta}}e^{-P}dp$$

$$f(t)=\frac{\gamma t^{k-2}\lambda^k e^{-\lambda t}}{\delta\Gamma k(\lambda t^{\gamma})^{\frac{1}{\delta}}}\int_0^\infty P^{\left(\frac{1}{\delta}+1\right)-1}e^{-P}dp$$

$$f(t)=\frac{\gamma t^{k-2}\lambda^k e^{-\lambda t}}{\delta\Gamma k(\lambda t^{\gamma})^{\frac{1}{\delta}}}\Gamma\frac{1}{\delta}+1$$

$$f(t)=\frac{\gamma t^{k-2-\frac{1}{\delta}}\lambda^{k-\frac{1}{\delta}}e^{-\lambda t}\Gamma\frac{1}{\delta}+1}{\delta\Gamma k} \tag{12}$$

And the conditional distribution of z given $T=t$ is

$$f(z/t)=\frac{g(z)s(z)}{s(t)}$$

$$f(z/t)=\frac{f(t,z)}{f(t)} \tag{13}$$

$$f(z/t)=\frac{z^{\delta}\lambda^{\frac{1}{\delta}+1}t^{\frac{1}{\delta}+r}e^{-z^{\delta}\lambda t^{\gamma}}\delta}{\Gamma\frac{1}{\delta}} \tag{14}$$

Estimates of mean with frailty ($E(T_{wf})$), δ=0 and without frailty ($E(T_{wof})$), δ≠0 are given by

$$E(T_{wf})=\frac{1}{\gamma z^{\frac{\delta}{\gamma}}\lambda^{\frac{1}{\gamma}}}\Gamma\left(\frac{1}{\gamma}+1\right)$$

$$E(T_{wof})=\lambda^{\frac{1}{\gamma}}\Gamma\left(\frac{1}{\gamma}+1\right)$$

The above estimates depend on the parameters with and without frailty respectively.

3. Maximum Likelihood Estimation

In this section we will obtain the maximum likelihood estimates of the parameters of weibull – gamma frailty model using EM algorithm, then the likelihood function for the weibull gamma farilty model is defined as

$$L=\prod_{i=1}^{n}f(t_i,z_i)$$

$$L=\prod_{i=1}^{n}\frac{z_i^{\delta}\lambda^{k+1}\gamma t_i^{k+r-2}e^{-(z_i^{\delta}t_i^{\gamma}\lambda-\lambda t_i}}{\Gamma k}$$

$$L=\left(\frac{\lambda^{k+1}\gamma}{\Gamma k}\right)^n\prod_{i=1}^{n}z_i^{\delta}t_i^{k+r-2}e^{-(z_i^{\delta}t_i^{\gamma}\lambda-\lambda t_i}$$

$$\log L=n\log\left(\frac{\lambda^{k+1}\gamma}{\Gamma k}\right)+\sum_{i=1}^{n}\log z_i^{\delta}+\sum_{i=1}^{n}(k+\gamma-2)\log t_i+\sum_{i=1}^{n}\left(z_i^{\delta}t_i-t_i\right)\lambda\left(\sum_{i=1}^{n}z_i^{\delta}t_i^{\gamma}\right)-\lambda\sum_{i=1}^{n}t_i \tag{15}$$

$$\log L=n(k+1)\log\lambda+n\log\gamma-n\log\Gamma k+\delta\sum_{i=1}^{n}\log z_i+(k+\gamma-2)\sum_{i=1}^{n}\log t_i+\lambda\sum_{i=1}^{n}z_i^{\delta}t_i-\lambda\sum_{i=1}^{n}t_i \tag{16}$$

Taking partial derivatives of the estimates $\delta, \lambda, k, \gamma$ in equation (16), we get

$$\frac{\partial \log L}{\partial \delta} = \sum_{i=1}^{n} \log z_i + \lambda \sum_{i=1}^{n} z_i^{\delta} \log z_i t_i^{\gamma} \quad (17)$$

On simplification we get,

$$n(k+1) - \lambda \sum_{i=1}^{n} t_i = 0$$

$$n(k+1) - \lambda \sum_{i=1}^{n} t_i$$

$$\hat{\lambda} = \frac{n(k+1)}{\sum_{i=1}^{n} t_i}$$

$$\frac{\partial \log L}{\partial k} = n \log \lambda - \frac{n}{\Gamma k} \Psi(k) + \sum_{i=1}^{n} \log t_i \quad (18)$$

$$\frac{\partial \log L}{\partial \gamma} = \frac{n}{\gamma} + \sum_{i=1}^{n} \log t_i + \lambda \sum_{i=1}^{n} z_i t_i^{\gamma} \log t_i \quad (19)$$

We use Newton-Raphson method, Further we use EM algorithm as z is unobservable. This method is used to predict the frailties. The frailty z_i is predicted by

$$z_i = E(z/t)$$

$$E(z/t) = \int_0^{\infty} z\, f(z/t) dz$$

$$E(z/t) = \int_0^{\infty} \frac{z . z^{\delta} \lambda^{\frac{1}{\delta}+1} t^{\frac{1}{\delta}+r} e^{-z^{\delta} \lambda t^{\gamma}} \delta}{\Gamma \frac{1}{\delta}+1} dz$$

$$E(z/t) = \frac{\lambda^{\frac{1}{\delta}+1} t^{\frac{1}{\delta}+r} \delta}{\Gamma \frac{1}{\delta}+1} \int_0^{\infty} z^{\delta+1} e^{-z^{\delta} \lambda t^{\gamma}} dz$$

Put $z^{\delta} \lambda t^{\gamma} = p \quad \Rightarrow \quad z = \left(\frac{p}{\lambda t^{\gamma}}\right)^{\frac{1}{\delta}}$

$$E(z/t) = \frac{\lambda^{\frac{1}{\delta}} t^{\frac{1}{\delta}}}{\Gamma \frac{1}{\delta}+1} \int_0^{\infty} z^{\delta} \lambda t^{\gamma} z\, \delta\, e^{-z^{\delta} \lambda t^{\gamma}} dz$$

$$E(z/t) = \frac{(\lambda t)^{\frac{1}{\delta}}}{\Gamma \frac{1}{\delta}+1} \int_0^{\infty} z\, e^{-p} z\, dp$$

$$E(z/t) = \frac{(\lambda t)^{\frac{1}{\delta}}}{\Gamma \frac{1}{\delta}+1} \left(\frac{1}{\lambda t^{\gamma}}\right)^{\frac{2}{\delta}} \int_0^{\infty} p^{\frac{2}{\delta}} e^{-p}\, dp$$

$$E(z/t) = \frac{1}{(\lambda t^{\gamma})^{\frac{1}{\delta}} \left(\Gamma \frac{1}{\delta}+1\right)} \int_0^{\infty} p^{\left(\frac{2}{\delta}+1\right)-1} e^{-p}\, dp$$

$$E(z/t) = \frac{\Gamma\frac{2}{\delta}+1}{(\lambda t^{\gamma})^{\frac{1}{\delta}}\left(\Gamma\frac{1}{\delta}+1\right)} \quad (20)$$

The E step algorithm requires the estimation of E(T⁄z) and E(ln(T⁄z))

$$E(\ln z/t) = \int_0^{\infty} \ln z f(z/t) dz$$

$$E(\ln z/t) = \int_0^{\infty} \frac{(\ln z)\, z^{\delta} \lambda^{\frac{1}{\delta}+1} t^{\frac{1}{\delta}+r} e^{-z^{\delta}\lambda t^{\gamma}} \delta}{\Gamma\frac{1}{\delta}+1} dz$$

$$E(\ln z/t) = \frac{\lambda^{\frac{1}{\delta}+1} t^{\frac{1}{\delta}+r} \delta}{\Gamma\frac{1}{\delta}+1} \int_0^{\infty} \ln z\, z^{\delta} e^{-z^{\delta}\lambda t^{\gamma}} dz$$

$$E(\ln z/t) = \frac{(\lambda t)^{\frac{1}{\delta}}}{\Gamma\frac{1}{\delta}+1} \int_0^{\infty} \log\left(\frac{p}{\lambda t^{\gamma}}\right)^{\frac{1}{\delta}} \left(\frac{p}{\lambda t^{\gamma}}\right)^{\frac{1}{\delta}} e^{-p} dp$$

$$E(\ln z/t) = \frac{1}{\delta\left(\Gamma\frac{1}{\delta}+1\right)} \int_0^{\infty} \ln\left(\frac{p}{\lambda t^{\gamma}}\right)^{\frac{1}{\delta}} (p)^{\frac{1}{\delta}} e^{-p} dp$$

4. Data Analysis

The data reported by McGilchrist et al. (1991), which corresponds to the recurrence time (in days) to infection, at the point of insertion of the catheter, for kidney patients using portable equipment. The data consisting of recurrence time (in days) for 38 patients is given in table 1.

Table 1. Recurrence Time to Infection Data

Patient Number	Recurrence Time	Event Type	Patient Number	Recurrence Time	Event Type
1	8,16	1,1	20	15,108	1,0
2	23,13	1,0	21	152,562	1,1
3	22,28	1,1	22	402,24	1,0
4	447,318	1,1	23	13,66	1,1
5	30,12	1,1	24	39,46	1,0
6	24,245	1,1	25	12,40	1,1
7	7,9	1,1	26	113,201	0,1
8	511,30	1,1	27	132,156	1,1
9	53,196	1,1	28	34,30	1,1
10	15,154	1,1	29	2,25	1,1
11	7,333	1,1	30	130,26	1,1

12	141,8	1,0	31	27,58	1,1
13	96,38	1,1	32	5,43	0,1
14	149,70	0,0	33	152,30	1,1
15	536,25	1,0	34	190,5	1,0
16	17,4	1,0	35	119,8	1,1
17	185,177	1,1	36	54,16	0,0
18	292,114	1,1	37	6,78	0,1
19	22,159	0,0	38	63,8	1,0

5. Fitting of the Model

In this section, we consider only uncensored observations from the data on recurrence time to infection, at the point of insertion of the catheter, for kidney patients using portable dialysis equipment. In each and every patients have two recurrence time, it is assumed to be independent and describes 1 is uncensored (infection occurs) and 0 is censored (infection not occurs).

Then we will estimate the parameters with help of E-M algorithm implemented by writing an R program. The estimated value of the parameter for the model is given in table 2.

Table 2. Estimates of the Parameters

Models	Parameter	Estimated Value
With Frailty	λ	0.9559351 (168.7459705)
	δ	0.5835266 (1620.6314155)
	γ	0.0010000 (0.2446828)
Without Frailty	λ	1.02355084 (23.67108004)
	δ	0.07526868 (1570.6314155)
	γ	0.00100000 (0.22629419

6. Conclusion

In this paper, we introduced a new frailty model named as a new generalized weibull gamma- frailty model. The weibull distribution has been used in the frailty model as a baseline distribution to obtain a new survival model for the analysis of data. Meanwhile, we obtain the new frailty model by considering the gamma distribution as the model by assuming the mean is equal to 1. The model parameters have been estimated by using EM algorithm and the values are given in table 2. Moreover, the goodness of fit of both frailty models has been examined by using real life data.

References

1. Ashok Shanubhogue and Ankit R. Sinojiya. (2015). A new generalized gamma-exponential frailty model. JP Journal of Biostatistics 12, 1-13.
2. Ashok Shanubhogue and Ankit R. Sinojiya. (2017). A new generalized weibull-exponential frailty model. Open Journal of Statistics 7, 84-91.
3. Balakrishnan, N. and Kai Liu. (2018). Semi parametric likelihood inference for Birnbaum Saunders frailty model. REVSTAT-Statistical Journal 16, 231-255.

4. Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological study of family tendency in chronic disease incidence. Biometrica 65, 141-151.
5. David, D. and Hanagal. (2020). Analysis of kidney infection data using correlated inverse Gaussian frailty model. Statistics and Applications 18, 1-19.
6. Hougaard, P. (1986). A class of multivariate failure time distributions. Biometrica 73, 671-678.
7. Kayid, M., lzadkhah, S., and Abuammoh, A.M. (2019). Proportional reversed hazard rates weighted frailty model. Physics A. 528, 1-14.
8. McGilchrist C. A. and C. W. Aisbett. (1991). Regression with frailty in survival analysis. Biometrics 47, 461-466.
9. Mota, A., Milani, E. A., Calsavara, V. F., Tomazella, V. L., Leao, J., Ramos, P.L., and Louzoda, F. (2021). Weighted lindley frailty model: estimation and application to lung cancer data. Life time data analysis 27, 1-14.
10. Nagaraj, J., Parthasarathy, S., and Ponnuraja, C. (2022). Lindley distribution as frailty models with application to life time data. Advances and Applications in Statistics 75, 119-134.
11. P. Haugaard. (2000). Analysis of multivariate survival data. Springer-Verlag.
12. Theodor, A. Balan and Hein Putter. (2020). A tutorial on frailty model. Statistical Methods in Medical Research 29, 3424-3454.
13. Vaupel, J.W., Manton, K.G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics. The American Statistician 29, 176-185.
14. Wienke, A. (2010). Frailty models in survival analysis. CRC Press.