DOI:10.26524/jms.2017.1

Journal of Management and Science ISSN: 2249-1260 | e-ISSN: 2250-1819

Parametric regression model for response time in clinical trials – a bayesian approach

DR. N. SUNDARAM

ASSISTANT PROFESSOR, DEPARTMENT OF STATISTICS, PRESIDENCY COLLEGE (AUTONOMOUS), CHENNAI-600 005, TAMILNADU, INDIA.

Abstract

In this paper an attempt has been made to model the censored survival data using Bayesian regressions with Markov Chain Monte Carlo (MCMC) methods. Bayesian Log-Normal (LN) regression model are found to be providing better fit than the other Bayesian regression models namely Exponential (E), Generalized Exponential (GE), Webull (W), Log-Logistic (LL) and Gamma (G).

Key words: Bayesian model; MCMC.

1. Introduction

Models fitted to survival data may involve parametric or semi-parametric or non-parametric forms for the hazard function. This depends on whether this form is defined as that of a known model, or whether it is completely undefined. In the past three decades, a number of regression-type models have been suggested for the analysis (Prentice, 1973; Aitkin and Clayton, 1980). The regression models have been reviewed by Kay (1977) and are introduced at length in the book of Kalbfleisch and Prentice (1980) and Cox and Oakes (1984).

Statistical modeling in Bayesian univariate parametric survival analysis and life testing is too large but some references dealing with applications to medical and industries were thus Box and Tiao (1973), Chen et al.(2000), Ibrahim et al. (2001) and Gelman et al. (2004). Kundu and Gupta (2008) studied only complete data set for Bayesian EE model. Venkatesan and Sundaram (2011) studied the two parameters GE models for censored survival data without covariates compared with Exponential, Weibull and Log-logistic models. This paper focused on, how the independent variable varies from different Bayesian regression models.

.

The section of the paper is arranged as follows. The basic concepts and notations are explained in section 2. In section 3, concept of Bayesian model presented. Real databases presented in section 4. Results and discussion of the performance of study models are presented in section 5. Summary and conclusion of study models are presented in section 6.

2. Notations

The actual survival time of an individual, *t*, can be regarded as the value of a variable *T*, which can take any non-negative value. The random variable T has the distribution function is given by

$$F(t) = P(T < t) = \int_0^t f(u)du, \tag{1}$$

Density function:

$$f(t) = dF(t) / dt.$$

and survival function is defined by:

$$S(t) = P(T \geq t) = 1 - F(t),$$

(2)

The hazard function therefore represents the instantaneous death rate for an individual surviving up to time *t* is given by

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t \mid T \geq t)}{\delta t} \right\} = \frac{f(t)}{S(t)}. \quad (3)$$

It then follows that

$$h(t) = \frac{-d\{\log S(t)\}}{dt},$$

(4)

and

$$S(t) = \exp\{-H(t)\},$$

(5)

Where

$$H(t) = \int_0^t h(u)du.$$

(6)

The most commonly used models are presented in table 1along with survival and hazards.

Table 1Parametric survival models

Models	Density function	Distribution function	Survival function	Hazard function
E	$f_E = \lambda e^{-\lambda t}, t \geq 0, \lambda > 0$	$F_E = 1 - e^{-\lambda t}$	$S_E = e^{-\lambda t}$	$h_E = \lambda$
GE	$f_{GE} = \alpha\lambda(1-e^{-\lambda t})^{\alpha-1}e^{-\lambda t}$ $t \geq 0, \alpha, \lambda > 0$	$F_{GE} = (1-e^{-\lambda t})^{\alpha}$	$S_{GE} = 1-(1-e^{-\lambda t})^{\alpha}$	$h_{GE} = \frac{\alpha\lambda(1-e^{-\lambda t})^{\alpha-1}e^{-\lambda t}}{1-(1-e^{-\lambda t})^{\alpha}}$
W	$f_W = \alpha\lambda(\lambda t)^{\alpha-1}e^{-(\lambda t)^{\alpha}}$ $t \geq 0, \alpha, \lambda > 0$	$F_W = 1 - e^{-(\lambda t)^{\alpha}}$	$S_W = e^{-(\lambda t)^{\alpha}}$	$h_W = \alpha\lambda(\lambda t)^{\alpha-1}$
LL	$f_{LL} = \frac{\alpha\lambda t^{\alpha-1}}{(1+\lambda t^{\alpha})^2}$ $t \geq 0, \alpha, \lambda > 0$	$F_{LL} = \int_0^t f_{LL}(x)dx$	$S_{LL} = \frac{1}{1+\lambda t^{\alpha}}$	$h_{LL} = \frac{\alpha\lambda t^{\alpha-1}}{1+\lambda t^{\alpha}}$
G	$f_G = \frac{\lambda}{\Gamma(\alpha)}(\lambda t)^{\alpha-1}e^{-\lambda t}$ $t, \lambda, \alpha > 0$	$F_G = \frac{1}{\Gamma(\alpha)}\int_0^{\lambda t} u^{\alpha-1}e^{-u}du$ $= I(\lambda t, \alpha)$	$S_G = 1 - F_G$	$h_G = \frac{f_G}{S_G}$
LN	$f_{LN} = \frac{1}{t\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(\log t - \mu)^2}$ $t, \sigma > 0, \mu \geq 0$	$F_{LN} = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$	$S_{LN} = 1 - F_{LN}$	$h_{LL} = \frac{f_{LN}}{S_{LN}}$

3 Bayesian model

The frequently used models for survival analyses are the Cox, AFT, Frailty, Gamma frailty and Bayesian regression. Recently parametric models play an important role in Bayesian survival analysis, since many Bayesian analyses in practice are carried out using parametric models. Theoretically, Bayesian techniques offer simple alternatives to statistical inference and all inferences follow from the posterior model.

In practice, we can obtain the posterior model with straightforward analytical solutions only in the most rudimentary problems. We presented posterior and predictive models and how to carry out Bayesian analysis for parametric regression models. For example, let $t = (t_1,\ldots,t_n)'$ be the *iid* survival times each having a Weibull model denoted by $W(\alpha, \gamma)$ as defined by:

$$f(t/\alpha, \lambda) = \alpha t^{\alpha-1} e^{(\lambda - e^{\lambda} t^{\alpha})}, \quad (7)$$

where $\lambda = log(\gamma)$ and the above expression denoted by $W(\alpha, \lambda)$. The survival function of the Weibull model is $S(t/\alpha,\lambda) = e^{(-e^{\lambda}t^{\alpha})}$ and then we can express likelihood function for the right censored survival data of $W(\alpha, \lambda)$ as:

$$L(\alpha, \lambda / D) = \prod_{i=1}^{n} \{f(t_i/\alpha,\lambda)\}^{\delta_i} \{S(t_i/\alpha,\lambda)\}^{(1-\delta_i)}. \quad (8)$$

Here the conjugate prior for e^{λ} is a Gamma prior if α is assumed known. Both (α, λ) are assumed unknown if no joint conjugate prior is available. In this situation, α and λ to be independent of joint prior and their models Gamma and Normal respectively. We can write the joint posterior model for Gamma prior $G(\alpha_0, k_0)$ and Normal prior $N(\mu_0, \sigma_0^2)$ of (α, λ) (Ibrahim et al, 2001) is given by

$$f(\alpha,\lambda/D) \propto L(\alpha,\lambda/D) f(\alpha/\alpha_0,k_0) f(\lambda/\mu_0,\sigma_0^2) = \prod_{i=1}^{n} \{f(t_i/\alpha,\lambda)\}^{\delta_i} \{S(t_i/\alpha,\lambda)\}^{(1-\delta_i)}. \quad (9)$$

The above joint posterior expression (20) of simplification does not have a closed form hence we can apply MCMC method. The regression model with covariate $\lambda_i = X_i'\beta$ can be written as follows.

$$f(\beta,\alpha/D) \propto \alpha^{\alpha_0+d-1} \exp\left\{\sum_{i=1}^{n} (\delta_i x_i'\beta + \delta_i(\alpha-1)\log(t_i) - t_i^{\alpha} e^{x_i'\beta}) - k_0\left(\alpha - \frac{1}{2}(\beta-\mu_0)\Sigma_0^{-1}(\beta-\mu_0)\right)\right\} \quad (10)$$

where $d = sum\ of\ \delta_i$ and β is the improper prior i.e., $f(\beta) \propto 1$ and we assuming a Normal prior $N_p(\mu_0, \Sigma_0)$ for β and a Gamma prior for α. The joint posterior expression (10) is a non-linear expression so we need to use numerical integration or MCMC. The Weibull Bayesian regression model is fitted by constructing the likelihood function of *n* observations and maximizing this function with respect to unknown parameters, $\beta_1, \beta_2, \ldots, \beta_p$, and α.

Similarly, we also fitted to the other Bayesian regression models respectively Cox, Exponential, Generalized Exponential, Log-logistic, Gamma and Log-Normal. This is done using computer software programme for survival analysis and the results are presented in tables 2 and 3.

4. Databases

For empirical comparison of the different models, we have considered a randomized controlled clinical trial data on Tuberculosis (TB) (Tuberculosis Research Centre (TRC), 2007). The aim of the study is to assess the response time to an 8 month treatment regimen consisting of Ethambutol, Rifampicin, Isoniazid and Pyrazinamide thrice weeks for first two months followed by Isoniazid and Ethambutol daily for next 6 months. The primary outcome variable is sputum culture conversion time. A total 467 patients were considered for this work. Out of these 90% had favourable response and 10% had not responded or lost which constitute the censored observations. Four important covariates were considered for model comparison namely *age, sex, weight on admission* and *percentage of allocated doses* received by each patient.

5. Results and Discussion

In the Bayesian regression model, we used Normal prior for covariates (β) and Gamma prior for alpha (α) with 20000 and 40000 samples and obtain the posterior summaries of covariates. From table 3, we observed that the covariate *dosage* is significant for *W, LL* (at 40000 samples) and *G* (at 20000 and 40000 samples) Bayesian regression models but not significant for other covariates and only *age* is significant for *GE* Bayesian regression model (at 20000 and 40000 samples) but Bayesian *GE* regression model gave the smaller MC standard error. Bayesian *LN* regression model (at 20000 and 40000 samples) is better fit than the other parametric Bayesian regression models based on Deviance information criterion (DIC) (table 2) value and figures (a-f) shows the trace plots of the MC samples, autocorrelation and marginal posterior densities of covariates. We notice that the Bayesian *LN* regression model, the plots indicate that the MC samples are mixing well at 40000 iterations. We observe that, the Bayesian *LN* regression model performs better than the other models (*E, GE, W, LL* and *G*).

Table 2 Model selections for sputum culture conversion data

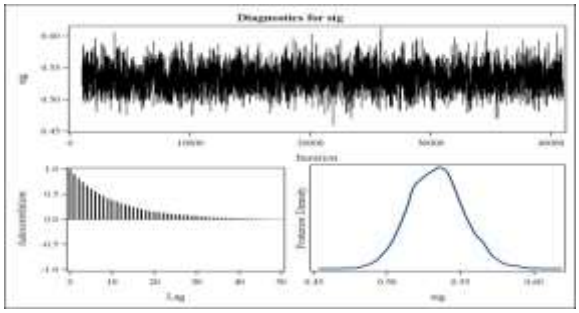
Models	Parameter/ Method	E	EE	W	LL	G	LN
Bayesian Regression (MC Samples-20000)	Scale/Mu	0.267	0.551	1.459	2.723	2.832	0.717
	Shape/Sig	-	2.943	0.271	0.072	1.374	0.534
	DIC	1834	1667	1745	1597	1615	1338@
Bayesian Regression (MC Samples 40000)	Scale/Mu	0.267	0.550	1.459	2.719	2.869	0.716
	Shape/Sig	-	2.941	0.271	0.073	1.382	0.534
	DIC	1834	1667	1745	1597	1615	1338@

Table 3 Posterior Summaries of sputum culture conversion data under different Bayesian Regression Models

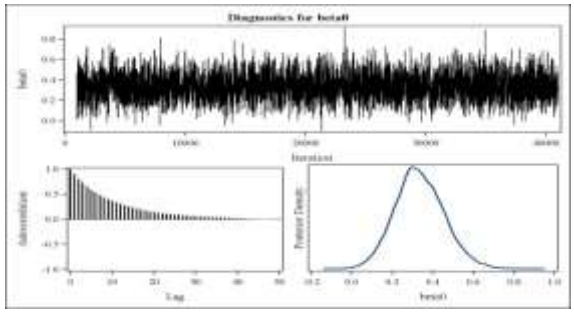
Samples	20000				40000			
Bayesian E Regression Model								
Mean	0.205	0.006	-0.065	-0.283	0.205	0.004	-0.064	-0.284
SD	0.117	0.106	0.107	0.152	0.120	0.107	0.107	0.152
MCSE	0.003	0.003	0.003	0.004	0.003	0.002	0.002	0.003
25%	0.123	-0.064	-0.136	-0.384	0.123	-0.068	-0.136	-0.387
50%	0.207	0.006	-0.064	-0.284	0.207	0.005	-0.065	-0.283
75%	0.288	0.078	0.008	-0.181	0.289	0.078	0.009	-0.182
Bayesian GE Regression Model								
Mean	0.060	0.006*	-0.015	-0.068	0.060	0.006*	-0.015	-0.069
SD	0.027	0.024	0.022	0.033	0.026	0.024	0.022	0.033
MCSE	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
25%	0.042	-0.009	-0.030	-0.090	0.042	-0.009	-0.030	-0.090
50%	0.061	0.006	-0.015	-0.067	0.060	0.006	-0.015	-0.068
75%	0.078	0.022	-0.001	-0.044	0.077	0.022	0.000	-0.046

Bayesian W Regression Model								
Mean	0.196	-0.009	-0.042	-0.197***	0.196	-0.010	-0.042	-0.196
SD	0.083	0.074	0.074	0.105	0.082	0.074	0.073	0.106
MCSE	0.004	0.003	0.003	0.009	0.003	0.002	0.002	0.005
25%	0.138	-0.060	-0.091	-0.268	0.139	-0.061	-0.092	-0.268
50%	0.195	-0.009	-0.041	-0.197	0.196	-0.009	-0.041	-0.195
75%	0.251	0.041	0.007	-0.128	0.253	0.040	0.008	-0.125
Bayesian LL Regression Model								
Mean	0.386	0.158	-0.142	-0.610	0.378	0.164	-0.135	-0.596**
SD	0.200	0.173	0.166	0.250	0.198	0.175	0.168	0.247
MCSE	0.007	0.006	0.005	0.010	0.005	0.004	0.004	0.007
25%	0.258	0.040	-0.249	-0.775	0.247	0.048	-0.245	-0.755
50%	0.386	0.155	-0.141	-0.607	0.375	0.165	-0.133	-0.591
75%	0.517	0.271	-0.027	-0.441	0.509	0.282	-0.022	-0.429
Bayesian G Regression Model								
Mean	0.024	0.014	0.012	0.023***	0.023	0.013	0.011	0.021**
SD	0.025	0.022	0.022	0.030	0.025	0.021	0.021	0.029
MCSE	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001
25%	0.006	-0.002	-0.003	0.001	0.006	-0.001	-0.003	0.001
50%	0.024	0.014	0.012	0.022	0.024	0.013	0.011	0.021
75%	0.041	0.029	0.026	0.044	0.041	0.028	0.025	0.041
Bayesian LN Regression Model								
Mean	-0.112	-0.074	0.003	0.058	-0.111	-0.074	0.000	0.057
SD	0.094	0.079	0.072	0.104	0.093	0.078	0.072	0.103
MCSE	0.003	0.002	0.003	0.004	0.002	0.002	0.002	0.002
25%	-0.175	-0.126	-0.045	-0.011	-0.172	-0.125	-0.048	-0.011
50%	-0.114	-0.074	0.002	0.063	-0.111	-0.075	0.000	0.062
75%	-0.050	-0.021	0.053	0.128	-0.049	-0.021	0.049	0.127

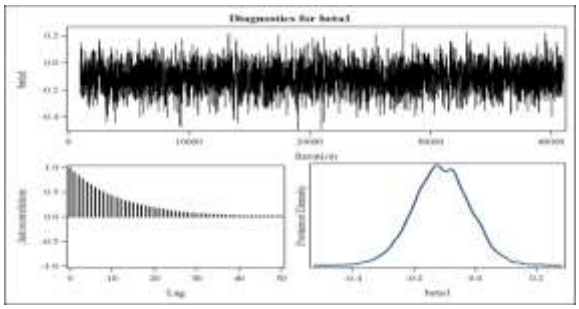
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and @ Better performance



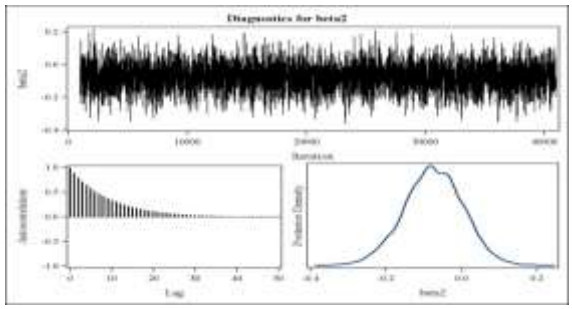
(a) Sigma



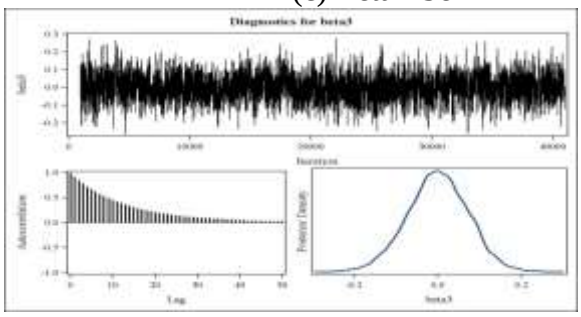
(b) Beta0



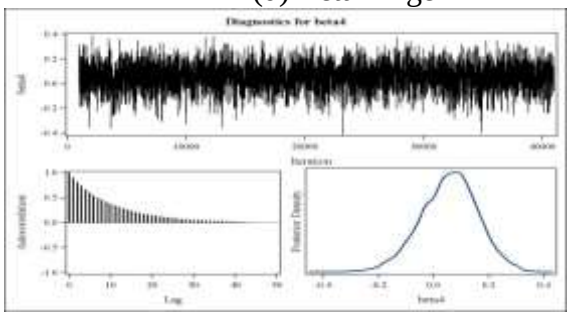
(c) Beta1-Sex



(d) Beta2-Age



(e) Beta3-Weight



(f) Beta4-Dosage

Figure (a-f).Trace plots, Autocorrelations and Marginal posterior densities of Coefficients using Bayesian Log-Normal Regression Model after 40000 iterations for sputum culture conversion data.

6. Summary and Conclusion

In this paper we consider a special case of parametric regression models with unknown shape, scale and regression coefficients of life time for censored data. Using the maximum likelihood with computer software programme, the above said parameter has been estimated. We observed from the discussion of section 5, Bayesian *LN* regression model using MCMC techniques seem to be more appropriate for the study of our right censored tuberculosis sputum culture conversion data when compared to other models. Moreover, the covariate *dosage* is mostly significant for all the models. Overall *LN* model performs well in Bayesian survival models. The performance of *GE, LL, G* and *LN* models improves considerably in low censored situation. Further studies are needed to validate this conclusion under different censoring patterns.

References

1) Aitkin M and Clayton D (1980) The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. App.Statist.29: 156-63.
2) Box GEP and Tiao GC (1973) Bayesian inference in statistical analysis. Wiley classics library edition, published 1992, New York: John Wiley & Sons.
3) Chen MH et al. (2000) Monte Carlo methods in Bayesian computation. New York: Springer-Verlag.
4) Cox DR and Oakes D (1984) Analysis of survival data. Chapman & Hall: London.
5) Gelman A et al. (2004) Bayesian data analysis. Second edition, London: Chapman & Hall.
6) Ibrahim JG et al. (2001) Bayesian survival analysis. . New York: Springer-Verlag.
7) Kalbfleisch JD and Prentice RL (1980) The statistical analysis of failure time data. New York: Wiley and Sons.

8) Kay R (1977) Proportional hazard regression models and the analysis of censored survival data. Appl.Statist.,26: 227-37.
9) Kundu D and Gupta RD (2008) Generalized exponential distribution: Bayesian estimations. Computational Statistics & Data Analysis, 52(4): 1873-83.
10) Prentice RL (1973) Exponential survivals with censoring and explanatory variables. Biometrika, 60: 279-88.
11) Venkatesan P and Sundaram N (2011) Exponentiated Exponential models for survival data. Indian Journal of Science and Technology, 4(8): 923-30.
12) TRC ICMR (2007) Evaluation of a non-rifampicin continuation phase (6HE) following thrice-weekly intensive phase for the treatment of new sputum positive pulmonary tuberculosis. Indian Journal of Tuberculosis, 54:84-90.

*****