Jurnal Teknologi

Full paper

Multiplicative Piecewise Gamma in Survival Data Analysis

Noraslinda Mohamed Ismail*, Zarina Mohd Khalid, Norhaiza Ahmad

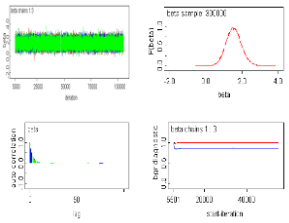
Department of Mathematical Sciences, Faculty of Science, 81310, UTM Johor Bahru, Johor, Malaysia

*Corresponding author: noraslinda@utm.my

Article history

Received :13 October 2013
Received in revised form :
28 January 2014
Accepted :4 February 2014

Graphical abstract



Abstract

The proportional hazard model is the most general of the regression models since it is not based on any assumptions concerning the nature or shape of the underlying survival distribution. The model assumes that the underlying hazard *rate* is a function of the covariates (independent variables) and there are no assumptions about the nature or shape of the hazard function. Proportional hazards model in survival analysis is used to estimate the effects of different covariates which was influenced by the survival data. This paper proposes the new multiplicative piecewise gamma in the hazard function using OpenBugs Statistical Packages. The proposed model is a flexible survival model for any types of non-informative censored data in estimating the parameters using Bayesian approach and also an alternative model to the existing model. We used the Markov Chain Monte Carlo method in computing the Bayesian estimator on leukemia data. The results obtained show that the proposed model can be an alternative to the existing multiplicative model since it can estimate the parameters using any types of survival data compared to the existing model that can only be used for leukemia data.

Keywords: Piecewise gamma; hazard function; Bayesian; MCMC

Abstrak

PLEASE PROVIDE

Kata kunci: **PLEASE PROVIDE**

© 2014 Penerbit UTM Press. All rights reserved.

1.0 INTRODUCTION

Survival analysis techniques are important tools for analyzing the data that belongs to field of medicine, engineering, marketing etc. Survival analysis is dealing with models, methods and is used for analyzing data of life times. Survival data are generally dealt with lifetimes from some initial event at time zero to some terminal event of interest and the data is basically would be an independent non-negative random variable, say *T*. In the era of the 1950's, a major advancement in survival analysis took place where Kaplan and Meier [1] proposed their famous estimator of the survival curve. David Cox [2] introduced the proportional hazards model which incorporates covariates later in 1972.

A general class of semi-parametric hazards regression model for survival data have been proposed by Chen and Jewell [3] which include Cox proportional hazards model, the accelerated failure time model and the accelerated hazards model. Their new models can yield more accurate prediction of an individual's survival process and are flexible. While a covariate's effect can be identified by separating two components namely a time scale change on hazard progression and a relative hazard ratio.

Semi-parametric Bayesian analyses of proportional hazard models recently have become computationally feasible due to modern technology and advancement in computing techniques such as the Gibbs sampler and other Markov Chain Monte Carlo (MCMC) methods. Arjas and Gasbarra [4] considered simple right censored data with a common unknown hazard rate in which the hazard rate is modelled nonparametrically. Gibbs Sampler is used in their study to generate the sample paths of the hazard rate from the posterior distribution.

An overview of Bayesian semiparametric methods for the Cox model has been provided by Sinha and Dey [5]. The

Bayesian method gives an advantage where we can join the baseline hazard and the regression coefficients that can be used accurately to compute the target posterior quantities using MCMC simulation techniques. Based on either the hazard or the intensity function, they investigated the potential of Bayes methods for the analysis of survival data using semiparametric models. The nonparametric part of every model is assumed to be a realization of a stochastic process while the parametric part is assumed to have a prior distribution with possibly unknown hyperparameters which may include a regression parameter or a parameter quantifying the heterogeneity of a population.

Ibrahim and Chen [6] developed a class of semi-parametric informative prior distributions for the Cox model. They specified a non-parametric prior for the baseline hazard rate and a parametric prior for the regression coefficients via the development of novel MCMC techniques for sampling from the posterior distribution of the parameters. Their approach seems to be a useful approach to this problem since it is difficult to specify meaningful prior distributions for the parameters in each model task and requiring contextual interpretations of a large number of parameters.

Proportional hazards model is modelled from a classical perspective by obtaining the partial likelihood approach to estimate the unknown parameters but recently, the model has become the most common from a Bayesian perspective ever since Sinha and Dey [5] had proposed an excellent paper in this area. It has been widely used in survival analysis for such realistic models. Fully Bayesian computations of multi-level or hierarchical model are now possible using simulation techniques. This new development has motivated the use of Bayesian methods in survival analysis and Gibbs Sampling is one of the new numerical algorithms which allow the obtaining of samples from posterior of interest. Gibb Sampling or Gibbs Sampler is commonly used as a means of statistical inference, especially in Bayesian inference, for obtaining a sequence of observations which are approximated from a specified multivariate probability distribution when direct sampling is difficult**.** Specialized software packages called BUGS [7, 8] are created for implementing MCMC-based analyses of full probability models. These packages will treat all unknowns as random variables.

In this paper, we used the proportional hazards model that has been used extensively since 1972. In a fully parametric model, the lifetime distribution has been assumed to belong to a family of parametric distributions and reducing the regression problem of estimating the parameters from the data. This paper describes the use of freely available software for the analysis of complex statistical models using MCMC techniques, called OpenBUGS [9].

The proposed baseline hazard function in Cox model is slightly different compared to the original Cox model which has been proposed by Kalbfleisch [10]. He proposed gamma process prior as the baseline hazard function in Cox model and he fixed the value of *c* (a specification of the weight attached to a guess) and *r* (a guess at the failure rate per unit time). In this paper, we extend our previous work using both *c* and *r* having gamma and uniform distributions, respectively by setting the shape parameter in gamma process prior to the baseline hazard function to have a piecewise function that has a Gamma distribution. The purpose of this extension is to make the model more useful and flexible using hyperparameters instead of using non-informative prior, and the most important it can be used to any types of survival data in estimating the parameters.

2.0 INFERENCE PROCEDURE

A random variable for the survival time of an individual, *T* with vector covariates **x** follows a multiplicative Piecewise Gamma model if its hazard function has the following form

$$\lambda(t|\mathbf{x}) = \lambda_0(t)\exp\left(\beta'\mathbf{x}(t)\right), \quad (1)$$

where $\lambda_0(t)$ is an unknown baseline hazard function that follows the Gamma distribution with mean, $\frac{\varpi}{\theta}$ and variance, $\frac{\varpi}{\theta^2}$.

Usually, the counting process analysis is based on the modeling of the intensity function in survival data. The counting process, $N_{ij}(t)$ which is observed in the i^{th} individual $(i = 1,2,\dots,N)$ and in the j^{th} cluster $(j = 1,2,\dots,K)$, is counting the number of failures which have occurred up to time *t*. $dN_{ij}(t)$ is the counting process increments in the time interval $[t, t+dt)$ and assumed to be independent Poisson random variables with means, $I_{ij}(t)dt$, where

$$dN_{ij}(t) \sim \text{Poisson}\left(I_{ij}(t)dt\right).$$

The new failure rate is then seen to be an interval and defined as

$$I_{ij}(t)dt = Y_{ij}(t)\lambda\left(t\middle|\mathbf{x}_{ij}\right) = Y_{ij}(t)d\Lambda\left(t\middle|\mathbf{x}_{ij}\right),$$

where $Y_{ij}(t)$ is an observed process and will take the value 1 or 0 according to whether or not subject or individual *i* is observed at time *t*.

The multiplicative intensity model which was adopted by Cox's model is given by

$$I_{ij}(t)dt = Y_{ij}(t)\exp\left(\beta\mathbf{x}_{ij}\right)d\Lambda_0(t),$$

where $d\Lambda_0(t)$ is the increment or jump in the integrated baseline hazard function occurring during the time interval $[t;\ t+dt)$.

Kalbfleisch [10] proposed gamma process prior for the baseline hazard function by assuming $d\Lambda_0(t)$ having a gamma distribution with mean $d\Lambda_0^*(t)$ and variance $d\Lambda_0^*(t)/c$, which can be written as

$$d\Lambda_0(t) \sim \text{Gamma}(cd\Lambda_0^*(t)\,, c).$$

We use the same baseline hazard function which is having a gamma distribution but with different mean and variance. We proposed our model using the gamma distribution with mean $d\Lambda_0^*(t)/c$ and variance $d\Lambda_0^*(t)/c^2$ which can be written as

$$d\Lambda_0(t) \sim \text{Gamma}(d\Lambda_0^*(t)\,, c).$$

The original Cox model is proposed by Kalbfleisch which used a gamma process prior as the baseline hazard function with both *c* and *r* are fixed values. Later Ismail *et al.* [11] proposed an alternative to make it more flexible assuming that both *c* and *r* are having their own distributions. Ayman and Anis [12] proposed the baseline hazard function as the non-parametric part of the model to be a non-negative polygonal function with the vertices located at times $a_0 = 0 < a_1 < \cdots < a_T < a_{T_{\max}+1}$, where the polygonal takes the values $\tau_0 = 0 < \tau_1 < \cdots < \tau_T < \tau_{T_{\max}+1}$, respectively and becomes constant over time $a_{T_{\max}}$. Ismail *et al.* [13] once again proposed an alternative which suggested a combination of both parametric and nonparametric functions as the baseline hazard function by adopting the idea of Beamonte and Bermudez [14].

Once again we adopt both ideas and proposed the piecewise gamma baseline hazard function for Cox model, slightly different from the original proposed by Kalbfleisch. Since the conjugate prior for the Poisson mean is the gamma distribution, it would be convenient if $\Lambda_0(\,)$ were a process in which the increments $d\Lambda_0(t)$ are assumed to be the conjugate independent increments prior.

In this paper, we assume the mean, $d\Lambda_0^*(t)$ to have a piecewise function,

$$d\Lambda_0^*(t) = \begin{cases} \tau_{j-1} + \dfrac{(\tau_j - \tau_{j-1})(t - a_{j-1})}{(a_j - a_{j-1})} & \text{if } a_{j-1} \le t \le a_j \,; j = 1,2,\dots,k \\ \tau_{j-1} & \text{if } t > a_k \end{cases}$$

that will take the values $\tau_0, \tau_1, \dots, \tau_k$ with the vertices that will be located at times $a_0, a_1, \dots, a_k$ and it will become constant over time a_k.

3.0 RESULTS AND DISCUSSION

Leukemia data will be used in this analysis, where the effect of 6-MP (6-Mercaptopurine) therapy for the duration of remissions induced by adrenal corticosteroids has been studied as a model for testing of new agents in leukemia patients. Patients in remission were assigned randomly to maintenance therapy with either 6-MP or placebo.

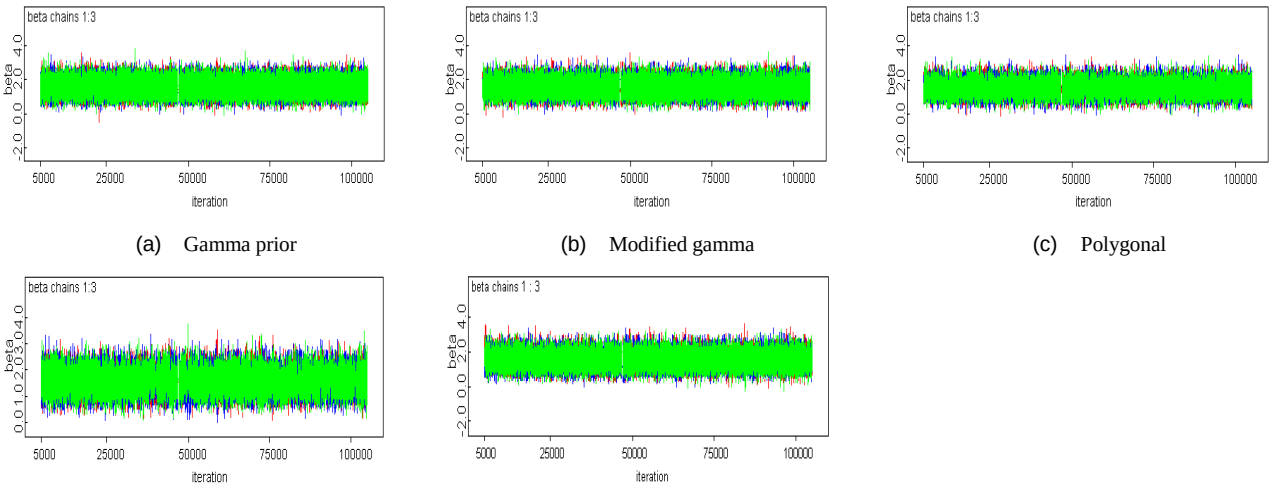
In our previous work, other than proposing a new model, analysis was also made on other existing models using Leukemia data and some comparisons are made on it. We do the analysis for four different types of the baseline hazard function in Cox regression, namely gamma prior baseline hazard function, modified gamma baseline hazard function, polygonal baseline hazard function and gamma polygonal baseline hazard function. In this paper, the same approach is used to justify that the proposed model can be an alternative to the existing models in estimating the parameter. We extend our previous work by extending the baseline hazard model to have a piecewise gamma model. The aim of this paper is to introduce the new piecewise gamma baseline hazard function as an alternative to the existing multiplicative model using the BUGS software program.

Table 1 Summaries of parameter estimation for Cox Regression with different baseline hazard function.

Baseline Hazard Function	Mean $\widehat{\boldsymbol{\beta}}$	Standard Deviation	MC Error	95% CI	DIC	- Log Likelihood
Gamma process	1.545	0.4189	0.001847	(0.7571, 2.402)	232.6	106.35
Modified gamma	1.538	0.4121	0.002079	(0.7611, 2.385)	209.2	102.40
Polygonal	1.520	0.4143	0.002920	(0.7378, 2.370)	211.2	101.75
Gamma polygonal	1.576	0.4056	0.003602	(0.8055, 2.399)	208.3	102.15
Piecewise gamma	1.532	0.4134	0.002294	(0.7466, 2.370)	215.0	102.45

The analysis started by choosing three parallel chains with different starting values for each model and they were carried out simultaneously. Each chain performed 100,000 iterations after 5,000 iterations for burn-in to obtain convergence to the posterior distribution. One out of every 100th values is used to reduce the autocorrelation of the chain. The convergence of the chains can be monitored via the Brooks-Gelman-Rubin (BGR) convergence-diagnostic graph.

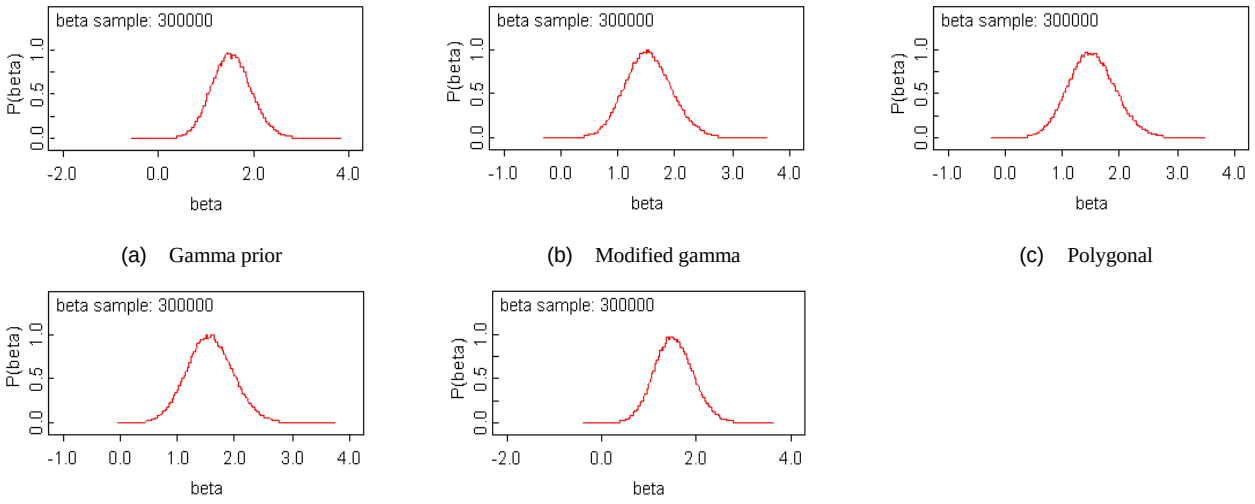
Table 1 shows the summaries of parameter estimation for Cox regression using different types of baseline hazard functions including the proposed model. The parameter estimation for all types is quite similar including the log-likelihood and deviance information criterion. Figure 1(a) to (e) show the posterior trace plots for 100,000 iterations for each of three generated samples while Figure 2(a) to (e) show the density plots associated with the coefficient of the covariate. The convergence of the parameters has been achieved since auto-correlations decreases only after considering a lag equal to 50 and this indicates a good convergence of the parameter space with a reasonably small number of iterations. This can be seen in Figure 3(a) to (e).



(a) Gamma prior (b) Modified gamma (c) Polygonal

(d) Gamma polygonal (e) Piecewise gamma

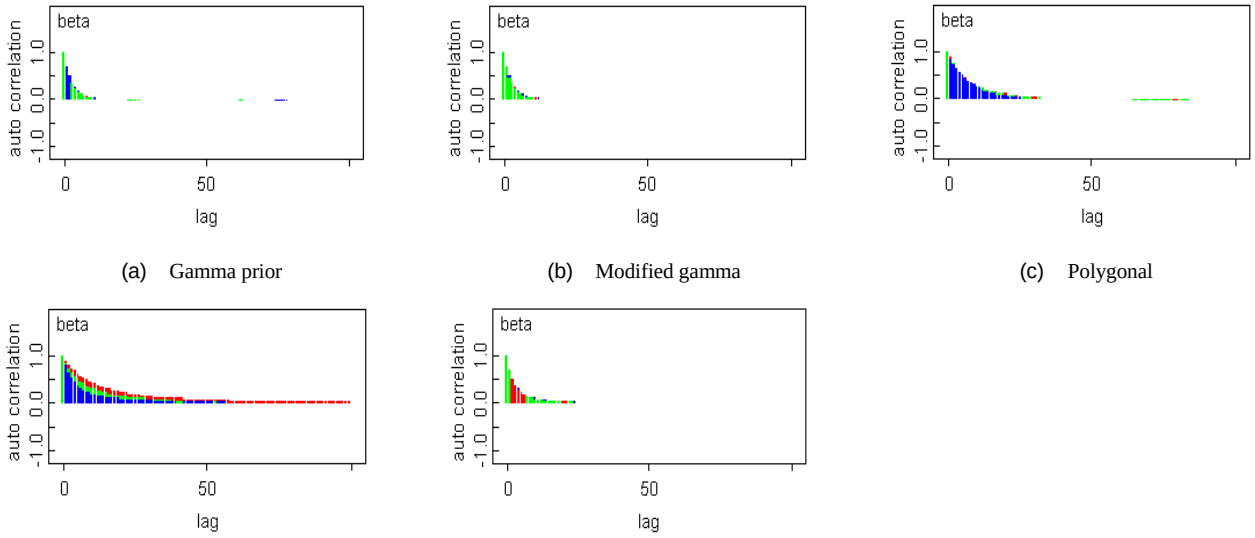
Figure 1 Different baseline hazard functions - estimated predictive history plots associated with the coefficient of the covariate



(a) Gamma prior (b) Modified gamma (c) Polygonal

(d) Gamma polygonal (e) Piecewise gamma

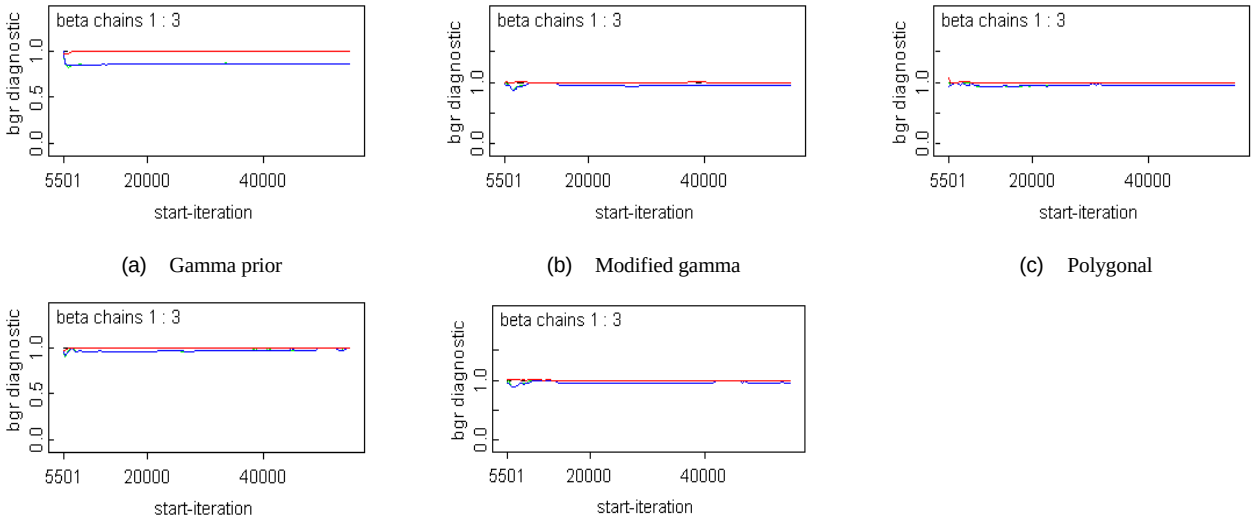
Figure 2 Different baseline hazard functions - estimated predictive density plots associated with the coefficient of the covariate.



(a) Gamma prior (b) Modified gamma (c) Polygonal

(d) Gamma polygonal (e) Piecewise gamma

Figure 3 Different baseline hazard functions - estimated predictive auto correlation plots associated with the coefficient of the covariate



(a) Gamma prior (b) Modified gamma (c) Polygonal

(d) Gamma polygonal (e) Piecewise gamma

Figure 4 Different baseline hazard functions-estimated predictive Brook-Gelman-Rubin diagnostic graphs associated with the coefficient of the covariate

The BGR convergence diagnostic graphs in Figure 4(a) to (e) show the line converted into one for stability indicating the convergence of the algorithm. This shows that our proposed model can be a good model as the existing models in the analysis of survival data. Apart from that, our model is eligible to become an alternative model that has great potential in applications involving survival data in the future.

4.0 CONCLUSION

Bayesian inference has several advantages particularly in the flexibility of model building for complex data over the frequentist approaches. The Bayesian approach enables us to make exact inference for any sample size based on the posterior distribution. OpenBUGS is a tool for analyzing survival data in a Bayesian framework using MCMC and provides the summaries of inferences and convergence in a table and graph.

We proposed a multiplicative piecewise gamma model using Bayesian approach to fit more flexible survival models for non-informative censored data using a Bayesian approach. Using OpenBUGS, we can see the performances of the proposed multiplicative piecewise gamma intensity models. We used the MCMC method in computing the Bayesian estimator on Leukemia data. The results obtained show that in analyzing paired survival data, the proposed model is as good as the existing multiplicative model. The proposed model shows a flexibility survival model for non-informative censored data and also can be a good model as existing multiplicative models.

Acknowledgement

This paper was supported in part by Research University Grant (GUP) grant 05J39. The authors would like to thank the Malaysian Ministry of Higher Education and Universiti Teknologi Malaysia for their financial funding through GUP grant.

References

[1] Kaplan, E. L., and Meier, P. 1958. Non-parametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*. 53: 457–481.
[2] Cox, D.R. 1972. Regression Models and Life Tables. *Journal of the Royal Statistical Society*. 34: 187–220.
[3] Chen, Y. Q., and Jewell, N. P. 2001. On a General Class of Semiparametric Hazards Regression Models. *Biometrika*. 88: 687–702.
[4] Arjas, E., and Gasbarra, D. 1994. Nonparametric Bayesian Inference from Right Censored Survival Data Using Gibbs Sampler. *Statistica Sinica*. 4: 505–524.
[5] Sinha, D., and Dey, D. K. 1997. Semiparametric Bayesian Analysis of Survival Data. *Journal of the American Statistical Association*. 92: 1195–1212.
[6] Ibrahim, J. G., and Chen, M. H. 1998. Prior Distributions and Bayesian Computation for Proportional Hazards Models. *The Indian Journal of Statistics*. 60: 48–64.
[7] Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. 1996. *BUGS 0.5: Examples Volume 1*, MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.
[8] Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. 2003. *WinBUGS User Manual, Version 1.4*, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine, UK. Available at: http://www.mrc-bsu.cam.ac.uk/bugs.
[9] Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. 2011. *OpenBUGS User Manual, Version 3.2.1*. MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine, UK. Available at: http://www.mrc-bsu.cam.ac.uk/bugs.
[10] Kalbfleisch, J. D. 1978. Nonparametric Bayesian Analysis of Survival Time Data. *Journal of the Royal Statistical Society*. 40: 214 – 221.
[11] Ismail, N. M., Khalid, Z. M., and Ahmad, N. 2012. Estimating Proportional Hazards Model Using Frequentist and Bayesian Approaches. *Journal of Fundamental & Applied Sciences*. 8(2): 73–82.
[12] Ayman, A. M., and Anis, B. G. 2011. Using WINBUGS to Cox Model with Changing from the Baseline Hazard Function. *Journal Applied Mathematical Sciences*. 5: 2217–2240.
[13] Ismail, N. M., Khalid, Z. M., and Ahmad, N. 2013. Survival Data Analysis using Additive and Multiplicative Gamma Polygonal Hazards Function. *Journal Matematika*. 29(1b): 117–127.
[14] Beamonte, E. and Bermudez, J. D. 2003. A Bayesian Semiparametric Analysis for Additive Hazard Model with Censored Observations. *Sociedad de Estadustica e Investigacion Operativa*. 12: 347–363.
[15] E.O. Freireich, *et al*. 1963. The Effect of 6-Mercaptopmine on the Duration of Steroid Induced Remission in Acute Leukemia. *Blood*. 21: 699–716.