Calculating Hazard Function of Survival Model by Bayesian Approach using Linex and General Entropy Loss Function with Jeffrey’s Prior

S. W Rizki* and E Sulistianingsih
Mathematics Department, Tanjungpura University, Pontianak, 78124, Indonesia
*E-mail: setyo.wirarizki@math.untan.ac.id

Abstract. This research observes about the cancer patients in a censored data. The purpose is to work out parameter estimation of hazard function of a survival model. By using Bayesian and giving a prior, posterior formula is constructed from Exponential - Jeffrey’s and Likelihood function. The formula is used to predict the parameter \( \lambda \) under Linex and General Entropy Loss Function (GELF) approach. The parameter estimation of hazard function \( \hat{h}_{BL} \), \( \hat{h}_{GELF} \) and survival function \( \hat{S}_{BL} \), \( \hat{S}_{GELF} \) are determined after \( \lambda_{BL} \) and \( \lambda_{GELF} \) found. The result shows that the value of \( \hat{h}_{BL} \) and \( \hat{h}_{GELF} \) are 1,058533880 and 0,008221809, respectively. Thus, we compare the result hazard value under both of Linex and GELF approximation with the real hazard value to find the best method for the research by finding smaller MSE. The MSE of hazard value under Linex and GELF are 1,103155 and 0, respectively. It means that the GELF approach is the best method.

1. Introduction
Survival model is a special random variable which presents degenerating probability value. It is often used in some cases which have high failure chance. The failure probability is expressed as hazard function. There are some models for representing the survival cases. In this observation, data describes the length of a censoring time of cancer patients and has an exponential distribution. The exponential distribution is a popular distribution which applied in some survival subjects. Let \( T \) denotes the failure time of an entity at time \( t = 0 \) and it is usually called as the failure time random variable, [1]. It is defined by its Survival Distribution Function (SDF) and Probability Distribution Function (PDF) as (1) and (2) equations,

\[
S(t) = e^{-\lambda t}, \quad t \geq 0, \quad \lambda \geq 0 \tag{1}
\]
\[
f(t) = -\frac{d}{dt}S(t) = \lambda e^{-\lambda t} \tag{2}
\]

Using (1) and (2) can be defined a Hazard Rate Function (HRF) which implies the conditional failure rate of survival time \( T \) as,

\[
\lambda(t) = \frac{f(t)}{S(t)} = \lambda \tag{3}
\]

The research investigates about the hazard value of cancer patients after gaining a treatment using Linex and General Entropy Loss Function approach in Bayesian inference. The data is taken from R 3.3.0 which presents the censoring time in patients of lung cancer observation. Cancer, one of dangerous diseases, has a high hazard value which presents the life opportunity of patients is decreasing and
converging to zero during less than three years. The results of survival probability and hazard value were calculated by Bayesian Linex Loss Function approach with gamma prior, see [2].

2. Formulating A Posterior Distribution on Censored Data
Censoring is a way to manage the partially known data in observation and there are three types of censoring, [3]. The censoring is conducted to overcome some problems for analysing data like the researcher needs some more time and costs to find complete data until the object gets an event.

The survival study can predict probability of survival and hazard rate by using some methods. There are some estimation methods in statistics. Bayesian is one of them which need likelihood function and prior distribution to find posterior distribution. The Bayesian approach allows direct probability statements about the parameter. Bayes theorem gives the way to find the predictive distribution of future observations. This is not always easily done in frequentist way, [4]. On censored data for observation \( (t_i, \delta_i), i = 1, 2, \ldots, n \), the likelihood function of exponential distribution can be constructed by

\[
L(t_i; \lambda, \delta) = \prod_{i=1}^{n} f(t_i; \lambda)^{\delta_i} [S(t_i; \lambda)]^{1-\delta_i}
\]

\[
= \prod_{i=1}^{n} [\lambda e^{-\lambda t_i}]^{\delta_i} [e^{-\lambda t_i}]^{1-\delta_i}
\]

\[
= \lambda^{\sum_{i=1}^{n} \delta_i} e^{-\lambda \sum_{i=1}^{n} t_i}
\]

(4)

Jeffrey’s prior is pointed as an informative prior to exponential distribution. The new extension of Jeffrey prior was developed and used by [5]. It is defined as following,

\[
f(\lambda) \propto \frac{1}{\lambda} e^{-\lambda t_i} e^{-\lambda \sum_{i=1}^{n} t_i}
\]

(5)

Then,

\[
f(\lambda) = \frac{kn^c}{2^c} ; \lambda, c > 0, k \text{ is a constant}
\]

Let \( \sim \text{Exponential} (\lambda) \) and prior density function \( \lambda \sim \text{Jeffrey’s} \) then we can build the posterior distribution which can be written as a conditional function of \( \lambda \) and knowing \( t \),

\[
f(\lambda | t_i) = \frac{\int_{0}^{\infty} f(\lambda) f(t_i; \lambda) \, d\lambda}{\int_{0}^{\infty} f(t_i; \lambda) \, d\lambda}
\]

\[
= \frac{\frac{kn^c}{2^c} \lambda^{\sum_{i=1}^{n} \delta_i} e^{-\lambda \sum_{i=1}^{n} t_i}}{\int_{0}^{\infty} \frac{kn^c}{2^c} \lambda^{\sum_{i=1}^{n} \delta_i} e^{-\lambda \sum_{i=1}^{n} t_i} \, d\lambda}
\]

\[
= \frac{\frac{kn^c}{2^c} \lambda^{\sum_{i=1}^{n} \delta_i} e^{-\lambda \sum_{i=1}^{n} t_i}}{\int_{0}^{\infty} \frac{kn^c}{2^c} \lambda^{\sum_{i=1}^{n} \delta_i} e^{-\lambda \sum_{i=1}^{n} t_i} \, d\lambda}
\]

(6)

From above we get the posterior of exponential distribution using Jeffrey’s prior is a \( \text{Gamma}(\sum_{i=1}^{n} t_i, \sum_{i=1}^{n} \delta_i - 2c + 1) \) distribution with \( \lambda \) is variable and \( t_i \) is sample.

3. Bayesian Approach under Linex and General Entropy Loss Function
One of some popular estimation methods in statistics is Bayesian. It has three loss functions. Those are Square Error, Linear Exponential, General Entropy and Lindley Loss Function approximation. The posterior expectation of the Bayesian Linear Exponential and General Entropy Loss Function is according to Zellner in [6] are denoted by equation (7) and (8) as below,

\[
\hat{\lambda}_{BL} = -\frac{1}{c} \ln[E(e^{-c\lambda})]
\]

(7)
\[
\hat{\lambda}_{BG} = \left[ E_{\lambda}(\lambda)^{-k} \right]^{-\frac{1}{k}} \tag{8}
\]

### 3.1. Linear Exponential Loss Function

The posterior function in equation (6) is used to find \( \hat{\lambda}_{BL} \) under Bayesian Linex loss function as following,

\[
E[e^{-c\lambda}] = \int_{0}^{\infty} e^{-c\lambda} f(\lambda | t_i) \, d\lambda \\
= \int_{0}^{\infty} e^{-c\lambda} \left[ \sum_{i=1}^{n} t_i \delta_i - 2c + 1 \lambda \sum_{i=1}^{n} \delta_i - 2c \, e^{-\lambda \sum_{i=1}^{n} t_i} \right] \, d\lambda \\
= \left( \sum_{i=1}^{n} t_i \right) \sum_{i=1}^{n} \delta_i - 2c + 1 \left( c + \sum_{i=1}^{n} t_i \right) \\
\hat{\lambda}_{BL} = -\frac{1}{c} \ln \left[ \left( \sum_{i=1}^{n} t_i \right) \sum_{i=1}^{n} \delta_i - 2c + 1 \left( c + \sum_{i=1}^{n} t_i \right) \right] \\
\quad = -\frac{1}{c} \left[ \left( \sum_{i=1}^{n} \delta_i - 2c + 1 \right) \ln \left( \sum_{i=1}^{n} t_i \right) + \left( 2c - \sum_{i=1}^{n} \delta_i - 1 \right) \ln \left( c + \sum_{i=1}^{n} t_i \right) \right] \tag{9}
\]

Equation (7) is estimation of parameter \( \lambda \) using Linex loss function approximation.

\[
\hat{S}_{BL}(t_i; \hat{\lambda}) = e^{\frac{1}{c} \left( \sum_{i=1}^{n} \delta_i - 2c + 1 \right) \ln \left( \sum_{i=1}^{n} t_i \right) + \left( 2c - \sum_{i=1}^{n} \delta_i - 1 \right) \ln \left( c + \sum_{i=1}^{n} t_i \right)} \tag{10}
\]

\[
\hat{h}_{BL}(t_i; \hat{\lambda}) = -\frac{1}{c} \ln \left[ \left( \sum_{i=1}^{n} t_i \right) \sum_{i=1}^{n} \delta_i - 2c + 1 \left( c + \sum_{i=1}^{n} t_i \right) \right] \\
\quad = -\frac{1}{c} \left[ \left( \sum_{i=1}^{n} \delta_i - 2c + 1 \right) \ln \left( \sum_{i=1}^{n} t_i \right) + \left( 2c - \sum_{i=1}^{n} \delta_i - 1 \right) \ln \left( c + \sum_{i=1}^{n} t_i \right) \right] \tag{11}
\]

The equations (10) and (11) are the estimation of survival and hazard function under Bayesian Linex Loss Function.

### 3.2. General Entropy Loss Function

The posterior function in equation (6) is used to find \( \hat{\lambda}_{BL} \) under Bayesian General Entropy Loss Function as following,

\[
E_{\lambda}(\lambda)^{-k} = \int_{0}^{\infty} \lambda^{-k} \left( \sum_{i=1}^{n} t_i \right)^{\sum_{i=1}^{n} \delta_i - 2c + 1} \lambda^{\sum_{i=1}^{n} \delta_i - 2c} \, e^{-\lambda \sum_{i=1}^{n} t_i} \, d\lambda \\
= \left( \sum_{i=1}^{n} t_i \right)^{k} \Gamma \left( \sum_{i=1}^{n} \delta_i - 2c + 1 \right) \\
\hat{\lambda}_{BG} = \left[ \left( \sum_{i=1}^{n} t_i \right)^{k} \Gamma \left( \sum_{i=1}^{n} \delta_i - 2c + 1 \right) \right]^{-\frac{1}{k}} \tag{12}
\]

Equation (12) is estimation of parameter \( \lambda \) using General Entropy Loss Function approximation.

\[
\hat{S}_{BG}(t_i; \hat{\lambda}) = e^{-\frac{1}{k} \left( \sum_{i=1}^{n} t_i \right)^{k} \Gamma \left( \sum_{i=1}^{n} \delta_i - 2c + 1 \right) \Gamma \left( \sum_{i=1}^{n} \delta_i - 2c - k + 1 \right)} t_i \tag{13}
\]
Equations (13) and (14) are the estimation of survival and hazard function under Bayesian General Entropy Loss Function.

4. Result and Discussion
Simulation uses data which is taken from R 3.3.0 in cancer studies. The data describes about the length of time of the lung cancer patients in observation. The measurement of censoring time is in days. The censoring time Interval is between 1 and 999 days. This research observes survival and hazard probability of cancer patients in censored data.

Table 1. Result of Survival Probability Estimation by LINEX and GELF Approaches

| The Length of Censoring Time | Survival Probability | Survival Probability Estimation under LINEX | Survival Probability Estimation under GELF |
|-----------------------------|----------------------|---------------------------------------------|---------------------------------------------|
| 1                           | 0.991811898          | 0.346964129                                 | 0.991811898                                 |
| 2                           | 0.983690841          | 0.120384107                                 | 0.983690841                                 |
| 3                           | 0.975636280          | 0.041768967                                 | 0.975636280                                 |
| 4                           | 0.967647670          | 0.014492333                                 | 0.967647670                                 |
| 7                           | 0.944072173          | 0.000605330                                 | 0.944072173                                 |
| 8                           | 0.936342013          | 0.000210028                                 | 0.936342013                                 |
| 10                          | 0.921071062          | 0.000025284                                 | 0.921071062                                 |
| 11                          | 0.913529238          | 8.77264E-06                                 | 0.913529238                                 |
| 12                          | 0.90649168           | 3.04379E-06                                 | 0.90649168                                 |
| 13                          | 0.898630344          | 1.05609E-06                                 | 0.898630344                                 |
| 15                          | 0.883974439          | 1.27136E-07                                 | 0.883974439                                 |
| 16                          | 0.876736366          | 4.41116E-08                                 | 0.876736366                                 |
| 18                          | 0.862437533          | 5.31034E-09                                 | 0.862437533                                 |
| 19                          | 0.855375806          | 1.84250E-09                                 | 0.855375806                                 |
| 20                          | 0.848371902          | 6.39281E-10                                 | 0.848371902                                 |
| 21                          | 0.841425346          | 2.21807E-10                                 | 0.841425346                                 |
| 22                          | 0.834535670          | 7.69592E-11                                 | 0.834535670                                 |
| 24                          | 0.820925094          | 9.26467E-12                                 | 0.820925094                                 |
| 25                          | 0.814203275          | 3.21451E-12                                 | 0.814203275                                 |
| 27                          | 0.800924304          | 3.86976E-13                                 | 0.800924304                                 |
| 29                          | 0.787861902          | 4.65857E-14                                 | 0.787861902                                 |
| 30                          | 0.781410809          | 1.61636E-14                                 | 0.781410809                                 |
| 31                          | 0.775012537          | 5.60818E-15                                 | 0.775012537                                 |
| 33                          | 0.762372734          | 6.75136E-16                                 | 0.762372734                                 |
| 999                         | 0.000270947          | 0                                          | 0.000270947                                 |

Table 1 above shows that the real survival probability has same value with survival probability estimation under Bayesian GELF approach. It is contras with value of survival probability estimation under Bayesian Linex approach. There is a much difference between the real’s and Linex’s. The survival probability value under Linex is lowest.
### Table 2. Result of Hazard Value Estimation by LINEX and GELF Approaches

| Hazard Value | Hazard Value Estimation under Linex (c=0.1) | Hazard Value Estimation under GELF (c=0.1, k=0.5) |
|--------------|------------------------------------------|------------------------------------------|
| 0.008221809  | 1.058533880                              | 0.008221809                              |

### Table 3. MSE of Survival and Hazard Value Estimation by LINEX and GELF Approaches

| MSE          | using Linex Approach | using GELF Approach |
|--------------|----------------------|---------------------|
| Survival Probability | 0.366379           | 0                   |
| Hazard Value          | 1.103155             | 0                   |

The table 2 presents the failure rate estimation under Linex and General Entropy Loss Function. From the table 2, the hazard value of the real and using GELF approach are 0.008221809. It is too far with the hazard value under Linex Loss Function is 1.058533880. It means that using Linex Loss Function approach, the life probability of patients is lower than using GELF approach.

The result shows that MSE value of hazard and survival under Bayesian Linex are 1.103155 and 0.366379, respectively. Furthermore, MSE values of hazard and survival using Bayesian GELF are zero. It concludes that the Bayesian General Entropy Loss Function approach is better than Linex Loss Function approach.

### Acknowledgements

Authors wishing to thank the Chief of LPPKM and Dean of Mathematics and Natural Science Faculty UNTAN who fund and support this research. By finishing and publishing this research, we can enrich and disclosure the knowledge of Bayesian methodology in Survival Studies.

### References

[1] London D 1988 *Survival Models and Their Estimation* Third ed. Connecticut: Actex Publications.
[2] Rizki SW, Mara MN, and Sulistianingsih E 2017 Survival Bayesian Estimation of Exponential-Gamma Under Linex Loss Function *Journal of Physics: Conference Series* **855**.
[3] Lee ET and Wang JW 2003 *Survival Methods for Survival Data Analysis* New Jersey: John Wiley & Sons
[4] Bolstad WM. 2007 *Introduction to Bayesian Statistics* Second ed. New Jersey: John Wiley & Sons, Inc.
[5] Al Kutubi HS and Ibrahim NA 2009 Bayes Estimator for Exponential Distribution with Extension of Jeffreys's Prior Information *Malaysian Journal of Mathematical Science* **3** (2) 297-313.
[6] Giure CB and Ibrahim NA 2012 Bayesian Analysis of the Survival Function and Failure Rate of Weibull Distribution with Censored Data *Mathematical Problems in Engineering*. 