IOP Conf. Series: Journal of Physics: Conf. Series **855** (2017) 012036 doi:10.1088/1742-6596/855/1/012036

Survival Bayesian Estimation of Exponential-Gamma Under Linex Loss Function

S. W Rizki, M. N Mara, and E Sulistianingsih
Mathematics Department of Mathematics and Natural Sciences Faculty of Tanjungpura University. Indonesia

E-mail: setyo.wirarizki@math.untan.ac.id

ABSTRACT. This paper elaborates a research of the cancer patients after receiving a treatment in cencored data using Bayesian estimation under Linex Loss function for Survival Model which is assumed as an exponential distribution. By giving Gamma distribution as prior and likelihood function produces a gamma distribution as posterior distribution. The posterior distribution is used to find estimatior $\hat{\lambda}_{BL}$ by using Linex approximation. After getting $\hat{\lambda}_{BL}$, the estimators of hazard function $\hat{h}_{BL}$ and survival function $\hat{S}_{BL}$ can be found. Finally, we compare the result of Maximum Likelihood Estimation (MLE) and Linex approximation to find the best method for this observation by finding smaller MSE. The result shows that MSE of hazard and survival under MLE are 2.91728E-07 and 0.000309004 and by using Bayesian Linex worths 2.8727E-07 and 0.000304131, respectively. It concludes that the Bayesian Linex is better than MLE.

1. Introduction

The Exponential distribution is a known distribution which can be used in many studies like biology, economics and demography. It is a very popular one-parameter distribution which is often used in survival model study, see [1]. Let T is defined to be the time of failure of the entity known to exist at time $t = 0$, and is therefore frequently called the failure time random variable. Now, if T is the time to failure, then the probability of still functioning at time t is the same as the probability that the failure is later (mathematically greater) than the value of t. The survival density function (SDF), probability density function (PDF) and hazard rate function (HRF) are defined as below,

$$S(t) = e^{-\lambda t} \quad . \quad t \geq 0. \ \lambda \geq 0 \tag{1}$$

$$f(t) = -\frac{d}{dt}S(t) = \lambda e^{-\lambda t} \tag{2}$$

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

Besides, using Maximum Likelihood to estimate the parameters, Bayesian method is required in this research. It builds a posterior distribution by formulating a likelihood function of the exponential distribution and a prior. The Gamma distribution is an appropriate prior for it because if we switch the parameter α with 1, it will change to be an Exponential distribution [2]. This research will observe survival probability of cancer patients after receiving the treatment in a cencored data. The data is taken from R versi 3.3.0 which has an exponential distribution. Cencoring is a feature that recurrent in lifetime and reliability data analysis, it occurs when exact lifetimes or run-outs can only be collected for a portion of the inspection units [3]. The survival study can predict probability of survival and hazard rate by some methods. There are some estimation methods in statistics. Bayesian is one of them which use likelihood function and prior distribution to find posterior distribution. The data is an Exponential distributed which will be composed with a Gamma distribution as its prior to construct a posterior distribution. It gives the relative weights to each parameter value after analizing the data [4]. The Bayesian inference has some approximations such as Generalised Non-informative Prior. Linear Exponential Loss Function, Lindley Approximation, General Entropy Loss Function and Squared Error Loss Function.

2. Maximum Likelihood Estimation on Cencored Data

Cencoring is a way to handle an uncomplete data which is caused some events like death. loss or out from observation. According to [5], variables $T_1.\dots.T_n$ represent n individual lifetimes. A time t_i is the lifetime or a cencoring time. The variable $\delta_i = 1$ if $T_i = t_i$ and 0 if $T_i > t_i$ is called the cencoring or status indicator for t_1. Value t_1 is obtained from $\min(T_i, C_i), i = 1,2,3,\dots,n$ where T_i is the duration of their remission measured from time of entry to study and C_i is the time between their date of entry and the end of study. The likelihood function of cencored data for observation (t_i, δ_i) $i = 1,2,\dots,n$ can be calculated is defined as,

$$L(t_i;\lambda,\delta) = \prod_{i=1}^{n}[f(t_i;\lambda)]^{\delta_i}[S(t_i;\lambda)]^{1-\delta_i} \tag{4}$$

The likelihood function of exponential distribution for observation (t_i, δ_i) $i = 1,2,\dots,n$ can be calculated by

$$\begin{aligned} L(t_i;\lambda,\delta) &= \prod_{i=1}^{n}\left[\lambda e^{-\lambda t_i}\right]^{\delta_i}\left[e^{-\lambda t_i}\right]^{1-\delta_i} \\ &= \left[\lambda^{\delta_1}\cdot\lambda^{\delta_2}\cdots\lambda^{\delta_n}\right]\left[e^{-\lambda t_1}\cdot e^{-\lambda t_2}\cdots e^{-\lambda t_n}\right] \\ &= \lambda^{\sum_{i=1}^{n}\delta_i} e^{-\lambda\left(\sum_{i=1}^{n} t_i\right)} \end{aligned} \tag{5}$$

then we find a natural logaritme of likelihood function above,

$$l = \ln L(t_i;\lambda,\delta) = \left(\sum_{i=1}^{n}\delta_i\right)\ln\lambda - \left(\sum_{i=1}^{n} t_i\right)\lambda$$

by deriving l to parameter λ. we obtain,

$$\frac{dl}{d\lambda} = 0$$

$$\frac{d}{d\lambda}\left[\left(\sum_{i=1}^{n}\delta_i\right)\ln\lambda - \left(\sum_{i=1}^{n} t_i\right)\lambda\right] = 0$$

$$\frac{\sum_{i=1}^{n}\delta_i}{\lambda} - \sum_{i=1}^{n} t_i = 0$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n} t_i} \tag{6}$$

We get $\hat{\lambda}$ is a Maximum Likelihood Estimation of λ. Later, composing $\hat{\lambda}$ into both of survival model and hazard function are

$$\hat{S}_{ML}(t_i;\hat{\lambda}) = e^{-\hat{\lambda}t_i} = e^{-\left(\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n} t_i}\right)t_i} \tag{7}$$

$$\hat{h}_{ML}(t_i;\hat{\lambda}) = \hat{\lambda} = \frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n} t_i} \tag{8}$$

$\hat{S}_{ML}(t_i;\hat{\lambda})$ and $\hat{h}_{ML}(t_i;\hat{\lambda})$ are Maximum Likelihood Estimation of survival model and hazard function.

3. Formulating Posterior Distribution

In this case, gamma distribution is assigned as a conjugate prior distribution to exponential distribution with parameter λ where $0 < \lambda < \infty$. The exponential distribution is a special form of gamma distribution. The gamma distribution is defined by

$$\Gamma(\alpha) = \int_0^\infty \lambda^{\alpha-1} e^{-\lambda}\, d\lambda \tag{9}$$

where $\alpha > 0$ and $0 < \lambda < \infty$
Selecting gamma distribution with parameter (α, β) as prior distribution for exponential distribution is based on closed form of gamma distribution. Let T is a continous random varible of gamma distribution with parameter α and β, we can describe the probability density function as below.

$$f(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} \;\; ; t > 0, \alpha > 0, \beta > 0 \tag{10}$$

The posterior distribution is a formula of likelihood function and prior distribution. It is defined by

$$f(\lambda \mid t_i) = \frac{f(\lambda) f(t_i; \lambda)}{\int_0^\infty f(\lambda) f(t_i; \lambda)\, d\lambda} \tag{11}$$

The density function $f(\lambda \mid t_i)$, $f(\lambda)$ and $f(t_i; \lambda)$ show the posterior distribution, the prior distribution and likelihood function, respectively. Let $T \sim Exponential\,(\lambda)$ and prior density function $\lambda \sim Gamma(\alpha, \beta)$ then we can formulate the posterior distribution which can be stated as a conditional function of λ and knowing t as,

$$\begin{aligned} f(\lambda \mid t_i) &= \frac{f(\lambda; t_i)}{f(t_i)} \\ &= \frac{f(\lambda) f(t_i; \lambda)}{\int_0^\infty f(\lambda) f(t_i; \lambda)\, d\lambda} \\ &= \frac{\left(\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}\right)\left(\lambda^{\sum_{i=1}^n \delta_i} e^{-\lambda(\sum_{i=1}^n t_i)}\right)}{\int_0^\infty \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}\right)\left(\lambda^{\sum_{i=1}^n \delta_i} e^{-\lambda(\sum_{i=1}^n t_i)}\right) d\lambda} \\ &= \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\sum_{i=1}^n \delta_i + \alpha - 1} e^{-\lambda(\sum_{i=1}^n t_i + \beta)}}{\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\sum_{i=1}^n \delta_i + \alpha - 1} e^{-\lambda(\sum_{i=1}^n t_i + \beta)}\, d\lambda} \\ &= \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\sum_{i=1}^n \delta_i + \alpha - 1} e^{-\lambda(\sum_{i=1}^n t_i + \beta)}}{\frac{\beta^\alpha}{\Gamma(\alpha)} \Gamma\left(\sum_{i=1}^n \delta_i + \alpha\right)\left(\frac{1}{\sum_{i=1}^n t_i + \beta}\right)^{\sum_{i=1}^n \delta_i + \alpha}} \\ &= \frac{\lambda^{\sum_{i=1}^n \delta_i + \alpha - 1} e^{-\lambda(\sum_{i=1}^n t_i + \beta)}}{\Gamma\left(\sum_{i=1}^n \delta_i + \alpha\right)\left(\frac{1}{\sum_{i=1}^n t_i + \beta}\right)^{\sum_{i=1}^n \delta_i + \alpha}} \\ &= \frac{\left(\sum_{i=1}^n t_i + \beta\right)^{\sum_{i=1}^n \delta_i + \alpha}}{\Gamma\left(\sum_{i=1}^n \delta_i + \alpha\right)} \lambda^{\sum_{i=1}^n \delta_i + \alpha - 1} e^{-\lambda(\sum_{i=1}^n t_i + \beta)} \end{aligned} \tag{12}$$

From the result above we obtain the posterior distribution is a Gamma distribution or it can be expressed by $Gamma(\sum_{i=1}^n \delta_i + \alpha\, , \sum_{i=1}^n t_i + \beta)$ where λ is a variable and t_i is sample.

4. Bayesian Linex Loss Function

Bayesian Method is a well known estimation method in whole statistical methodology studies. There are three loss functions of the Bayes estimator. Linex is one of loss function in Bayesian estimation. The posterior expectation of the Linex loss function is according to Zellner in [3]. The Linex loss function is one of bayesian approaches. Using result of Zellner (1986) in [5], the parameter estimator of λ which is denoted by $\hat{\lambda}_{lin}$ under Linex loss function is defined by,

$$\hat{\lambda}_{lin} = -\frac{1}{c}\ln\left[E\left(e^{-c\lambda}\right)\right] \tag{13}$$

We found the posterior formula in (12) that can be used to find a parameter estimator $\hat{\lambda}_{BL}$ under Bayes Linex loss function,

$$\begin{aligned}
E\left[e^{-c\lambda}\right] &= \int_0^\infty e^{-c\lambda} f(\lambda \mid t_i)\, d\lambda \\
&= \int_0^\infty e^{-c\lambda}\left[\frac{\left(\sum_{i=1}^n t_i+\beta\right)^{\sum_{i=1}^n \delta_i+\alpha} \lambda^{\sum_{i=1}^n \delta_i+\alpha-1} e^{-\lambda\left(\sum_{i=1}^n t_i+\beta\right)}}{\Gamma\left(\sum_{i=1}^n \delta_i+\alpha\right)}\right] d\lambda \\
&= \frac{\left(\sum_{i=1}^n t_i+\beta\right)^{\sum_{i=1}^n \delta_i+\alpha}}{\Gamma\left(\sum_{i=1}^n \delta_i+\alpha\right)} \int_0^\infty \lambda^{\sum_{i=1}^n \delta_i+\alpha-1} e^{-\lambda\left(\sum_{i=1}^n t_i+\beta+c\right)}\, d\lambda \\
&= \frac{\left(\sum_{i=1}^n t_i+\beta\right)^{\sum_{i=1}^n \delta_i+\alpha}\Gamma\left(\sum_{i=1}^n \delta_i+\alpha\right)}{\left(\sum_{i=1}^n t_i+\beta+c\right)^{\sum_{i=1}^n \delta_i+\alpha}\Gamma\left(\sum_{i=1}^n \delta_i+\alpha\right)} \\
&= \left(\frac{\sum_{i=1}^n t_i+\beta}{\sum_{i=1}^n t_i+\beta+c}\right)^{\sum_{i=1}^n \delta_i+\alpha}
\end{aligned} \tag{14}$$

from (14) we get the parameter estimation under Linex loss function as following

$$\begin{aligned}
\hat{\lambda}_{BL} &= -\frac{1}{c}\ln\left[E\left(e^{-c\lambda}\right)\right] \\
&= -\frac{1}{c}\ln\left[\left(\frac{\sum_{i=1}^n t_i+\beta}{\sum_{i=1}^n t_i+\beta+c}\right)^{\sum_{i=1}^n \delta_i+\alpha}\right]
\end{aligned} \tag{15}$$

Futhermore, the survival and hazard function under Linex loss function are expressed by equation (16) and (17),

$$\begin{aligned}
\hat{S}_{BL}\left(t_i;\hat{\lambda}\right) &= e^{-\hat{\lambda} t_i} \\
&= e^{-\left(-\frac{1}{c}\ln\left[\left(\frac{\sum_{i=1}^n t_i+\beta}{\sum_{i=1}^n t_i+\beta+c}\right)^{\sum_{i=1}^n \delta_i+\alpha}\right]\right) t_i}
\end{aligned} \tag{16}$$

$$\begin{aligned}
\hat{h}_{BL}\left(t_i;\hat{\lambda}\right) &= \hat{\lambda} \\
&= -\frac{1}{c}\ln\left[\frac{\left(\sum_{i=1}^n t_i+\beta\right)^{\sum_{i=1}^n \delta_i+\alpha}}{\left(\sum_{i=1}^n t_i+\beta+c\right)^{\sum_{i=1}^n \delta_i+\alpha}}\right]
\end{aligned} \tag{17}$$

5. Result

There are many diseases around our life. Some of them are dangerous and making a high death risk like diabetes mellitous, cancer, stroke and heart attack. Researchers are interested in study about cancer. In the last decades, cancer patients are growing bigger in the world. This disease has suffered people in all ages. Science makes some contributions to detect the risk which is caused by the diseases especially in Mathematics and Statistics. Survival study is derived from both mathematics and statistics. It relates to Actuarial Science which has developed in some countries. The data is taken from R versi 3.3.0. It describes about patients of lung cancer which has an exponential distribution.
The result of survival value under MLE and Bayes Linex is presented as below,

Table 1. Calculation of Survival Probability under MLE and Bayes Linex Loss Function.

The lenghth of Cencoring Time	Survival Probability	Estimation of Survival Probability under MLE	Estimation of Survival Probability under Bayesian Linex
1	0.991811898	0.992347739	0.992343628
1	0.991811898	0.992347739	0.992343628
2	0.983690841	0.984754035	0.984745875
3	0.975636280	0.977218440	0.977206294
4	0.967647670	0969740509	0.969724439
7	0.944072173	0.947648307	0.947620826
8	0.936342013	0.940396655	0.940365488
10	0.921071062	0.926059399	0.926021035
11	0.913529238	0.918972951	0.918931073
12	0.906049167	0.911940730	0.911895395
13	0.898630344	0.904962322	0.904913584
15	0.883974439	0.891165298	0.891109920
16	0.876736366	0.884345868	0.884287250
18	0.862437533	0.870863161	0.870798222
19	0.855375806	0.864199089	0.864131067
20	0.848371902	0.857586012	0.857514958
21	0.841425346	0.851023540	0.850949504
22	0.834535670	0.844511285	0.844434318
24	0.820925094	0.831635896	0.831553211
25	0.814203275	0.825272001	0.825186530
27	0.800924304	0.812689932	0.812599032
29	0.787861902	0.800299690	0.800203545
30	0.781410809	0.794175588	0.794076889
31	0.775012537	0.788098348	0.787997140
33	0.762372734	0.776083028	0.775976934
35	0.749939076	0.764250893	0.764140085
36	0.743798498	0.758402646	0.758289544
42	0.707996613	0.724241065	0.724115058
44	0.696449783	0.713199311	0.713069317
45	0.690747181	0.707741724	0.707609793
48	0.673918010	0.691618263	0.691480743
51	0.657498860	0.675862120	0.675719335
52	0.652115192	0.670690246	0.670545776
53	0.646775606	0.665557949	0.665411828
54	0.641479742	0.660464926	0.660317187
56	0.631017746	0.650395501	0.650244626
59	0.615643806	0.635578476	0.635423141
61	0.605603173	0.625888469	0.625730318
63	0.595726294	0.616346195	0.616185350
72	0.553236206	0.575174574	0.575003032
73	0.548706251	0.570773187	0.570600595
.	.	.	.
.	.	.	.
.	.	.	.
999	0.000270947	0.000465751	0.00046575

The table 1 presents the survival probability of cancer patients after getting treatment under MLE and Bayesian Linex Loss function. Both of results give the value are greater than the real value but the estimation value under Bayesian Linex is closer to the survival value than the MLE's. The result shows that the all values are inversely proportional with the time. The survival probability of patients is getting smaller and converges to zero after less 3 years later. It means that the effect of treatment will be fading by the time. Decreasing point of survival probability is around 10% in about 10 days of different range. The degenerate rate is associated to hazard function of survival study. The hazard value estimation under both MLE and Bayesian Linex Loss function is given in table 2 as following,

Table 2. The Hazard Values of MLE and Bayes Linex Loss Function Estimation.

Hazard	Estimation of Hazard Value under MLE	Estimation of Hazard Value under Bayesian Linex
0.008221809	0.007681689	0.007685833

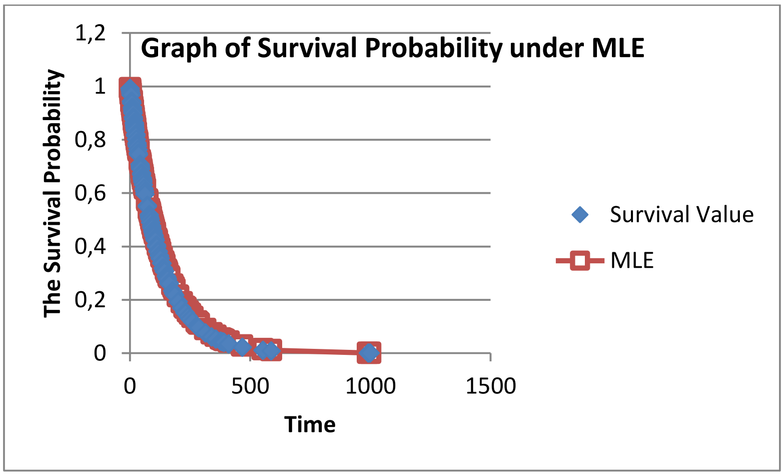
The hazard value serves a failure rate in reliability. From table 2, the failure rate of the real value, using MLE and Bayesian Linex are 0.8221809%, 0.7681689% and 0.7685833%, respectively. It means the Bayesian's is nearer to the real value than MLE's. Calculation of Mean Squared Error (MSE) is one way to find the best method by viewing the smallest MSE of both results. Those are shown by table 3 as below,

TABLE 3. MSE Values of Survival and Hazard under MLE and Bayes Linex Loss Function.

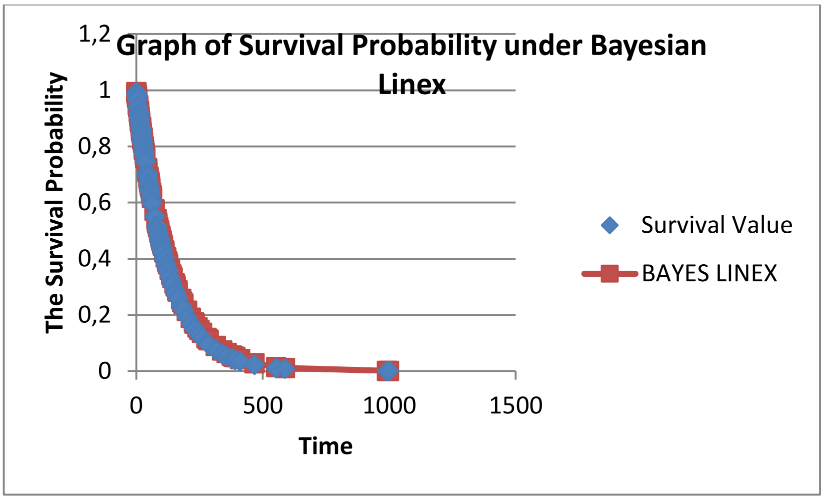
MSE	Hazard	Survival
MLE	2.91728E-07	0.000309004
Bayesian Linex Loss Function	2.8727E-07	0.000304131

The result shows that MSE value of Hazard and survival under Bayes Linex Loss Function is less than MSE value of Hazard and survival under MLE.

Graph 1. The Comparison Plot of The Real Survival Probability and Using MLE.



Graph 2. The Comparison Plot of The Real Survival Probability and Using Bayesian Linex.



Although, both of graph 1 and graph 2 show that they have a slightly different value with the real value of survival probability, the line of graph 2 has a more rapid distant between the real's and Bayesian's.

6. Conclusion

To sum up, the research tells some points. First, the cancer is a high risk disease because the survival opportunity is decreasing rapidly by the time. It is not longer than 3 years the survival probability limits to zero. It means the cancer is almost impossible to be reaped. Second, the hazard rate and length of post treatment observation have a particular role to survival value. Those are inversely proportional with the survival probability. The last, the MSE shows that the Bayesian Linex Loss function is better than MLE.

7. Acknowledgements

We wish to thank the Chief of LPPKM and Dean of Mathematics and Natural Science Faculty Tanjungpura University who fund and support this research. By doing the research, we can improve the creativity and enrich knowledge in our field study.

IOP Conf. Series: Journal of Physics: Conf. Series **855** (2017) 012036 doi:10.1088/1742-6596/855/1/012036

REFERENCES

[1] London. D,. 1988. *Survival Models and Their Estimation.* Third Edition, Connecticut: Actex Publications.

[2] Fitria. S,. Helmi & Rizki. S. W.. 2016. Survival Model Parameter Estimation Exponential Distribution of Data Censored by the Method of Maximum Likelihood and Bayesian SELF. *Bimaster,* 5(03), pp. 213-220.

[3] Guure. C. B. & Ibrahim. N. A,. 2012. Bayesian Analysis of the Survival Function and Failure Rate of Weibull Distribution with Censored Data. *Mathematical Problems in Engineering.*

[4] Bolstad. W. M,. 2007. *Introduction to Bayesian Statistics.* Second Edition, New Jersey: John Wiley & Sons, Inc.

[5] Metiri. F.. Zeghdoudi. H. & Remita. M. R,. 2016. On Bayes Estimatesof Lindley Disitribution under Linex Loss Function: Informative and Non Informative Priors. *Global Journal of Pure and Applied Mathematics,* 12(1), pp. 391-400.