Comparing the Bayesian Estimates of Hazard Rate of Mixed Distribution and Hazard Rates by the MLE Method

Suneung Ahn (sunahn@hanyang.ac.kr), Hyunmook Kim(mook@isl.hanyang.ac.kr)
Department of Industrial Engineering, Hanyang University, Ansan 425-791, South Korea

Abstract

This paper is intended to compare between the Bayesian estimates of hazard rate and the hazard rates of mixed distributions. In estimating hazard rates, especially when the MLE method is used, such difficulties as a lack of data and the existence of censored data make it difficult to estimate the rates. For this reason, the estimates of hazard rate based on the Bayesian approach are introduced. For the simplicity, the exponential and gamma distributions are adopted as a sampling distribution and its natural conjugate prior distribution, respectively.

1. Introduction

The hazard rate which is the instantaneous failure rate does frequently involve unknown parameters [1]. These parameters are usually estimated from observed data and then the hazard rate is expressed in terms of these estimated parameters. This inference on the hazard rate is fine as long as there are sufficient data. On the other hand, if few data are available, it is important to allow for parameters to have their own probability distributions. We do this approach by adopting the Bayesian inference [2, 5].

In the Bayesian approach, the hazard rate may be obtained by two methods. The first method is to derive the hazard rate, according to the definition of hazard rate, from a mixed distribution. The mixed distribution is resulted from the mixture (or marginalization) of a sampling distribution and a distribution of the parameters of interest. The second method is to estimate the hazard rate, which is called the Bayesian estimate of the hazard rate, by taking expectation of the hazard rate of a sampling distribution over the parameters having their own distribution. It is common that the second method is performed for a squared-error loss function (SELF) so that the Bayesian estimates includes the "posterior" means of the parameters given data [3, 5]. In this paper, the exponential distribution and the gamma distribution are employed as a sampling distribution and its natural conjugate distribution of the parameter of interest, respectively.

2. Hazard rates from the prior distribution

The sampling distribution of an outcome t which follows $\exp(\theta)$, given parameter θ, is

$$p(t|\theta) = \theta\, e^{-\theta t}. \tag{1}$$

The conjugate prior distribution [4] for the exponential parameter θ is gamma $(\theta \mid \alpha, \beta)$, which is of form

$$\pi(\theta \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \tag{2}$$

where $\alpha > 0$ and $\beta > 0$ are parameters characterizing the distribution of the parameter θ, which are called "hyperparameters". Denote (2) by $\pi(\theta)$. These hyperparameters will be updated by using the observed data afterwards.

In order to derive the hazard rate from the first method, we calculate the corresponding mixed exponential distribution using the "law of total probability",

$$p(t) = \alpha\left(\frac{\beta}{\beta+t}\right)\left(\frac{1}{\beta+t}\right) \tag{3}$$

The corresponding survival distribution, $S(t)$, is

$(\beta/(\beta+t))^{\alpha}$. Following the definition of hazard rate, or $h(t)=p(t)/S(t)$, we have

$$h(t)=\alpha/(\beta+t) \tag{4}$$

as the hazard rate of the mixed distribution.

Since the gamma distribution is a conjugate distribution for the exponential sampling distribution [6] and using a Bayes' formula, we have,

$$\pi(\theta \mid t)=\frac{(\beta+t)^{\alpha+1}}{\Gamma(\alpha+1)}\theta^{\alpha}e^{-(\beta+t)\theta}, \tag{5}$$

the gamma distribution with the updated shape and scale hyperparameters $\alpha+1$ and $\beta+t$ respectively.

Now, we derive the Bayesian estimate of the hazard rate under the E[$h(t\mid\theta)$] which is the expectation of the hazard rate of the sampling distribution $p(t\mid\theta)$ over the random parameter θ. We have the following Bayesian estimate of the hazard rate

$$\mathrm{E}[h(t\mid\theta)\mid t]=\int_0^{\infty}\theta\frac{(\beta+t)^{\alpha+1}}{\Gamma(\alpha+1)}\theta^{\alpha}e^{-(\beta+t)\theta}d\theta$$

$$=(\alpha+1)/(\beta+t). \tag{6}$$

We have two different hazard rates (4) and (6). Note that both of the hazard rates are functions of t which is the unobserved future observation. The hazard rates (4) and (6) are decreasing in t. This is due to the fact that a mixture of exponential distributions, each of which is DFR (decreasing failure rate), is itself DFR [1].

3. Hazard rates from the posterior distribution

We now consider the case that we have the observed data t and calculate the hazard rates with respect to the unobserved future observation $\tilde{t}$. By using (5) and the fact that $\tilde{t}$ and t are independently conditioned on the parameter θ we have the corresponding mixed distribution,

$$p(\tilde{t}\mid t)=(\alpha+1)\left(\frac{\beta+t}{\beta+t+\tilde{t}}\right)^{\alpha+1}\left(\frac{1}{\beta+t+\tilde{t}}\right). \tag{7}$$

Its survival distribution, $S(\tilde{t}\mid t)$, is $\left(\frac{\beta+t}{\beta+t+\tilde{t}}\right)^{\alpha+1}$. By following the definition of hazard rate, or $h(\tilde{t}\mid t)=p(\tilde{t}\mid t)/S(\tilde{t}\mid t)$, we have

$$h(\tilde{t}\mid t)=\frac{\alpha+1}{\beta+t+\tilde{t}} \tag{8}$$

as the hazard rate of the mixed distribution. If there are n additional data $\underset{\sim}{t}=\{t_1,\ldots,t_n\}$, we have

$$h(\tilde{t}\mid \underset{\sim}{t})=\frac{\alpha+n}{\beta+t_1+\ldots+t_n+\tilde{t}}. \tag{9}$$

We now derive the Bayesian estimate of the hazard rate. We take expectation over θ, with the posterior gamma distribution updated from (5). Thus, we have

$$\mathrm{E}[h(\tilde{t}\mid t,\theta)\mid\tilde{t}]=\int_0^{\infty}\theta\cdot\frac{(\beta+t+\tilde{t})^{\alpha+2}}{\Gamma(\alpha+2)}\theta^{\alpha+1}e^{-(\beta+t+\tilde{t})\theta}d\theta$$

$$=\frac{\alpha+2}{\beta+t+\tilde{t}}. \tag{10}$$

Similarly in the case of n observed data we have

$$\mathrm{E}[h(\tilde{t}\mid \underset{\sim}{t},\theta)\mid\tilde{t}]=\frac{\alpha+n+1}{\beta+t_1+\ldots+t_n+\tilde{t}}. \tag{11}$$

We have again two different hazard rates. Two hazard rates are also decreasing in time $\tilde{t}$ due to the fact that a mixture of exponential distributions is DFR.

4. Comparing resulting hazard rates with MLE

In classical approach, unknown parameters in stochastic models are usually estimated from observed data by adopting the maximum likelihood estimate (MLE) method. The MLE is given by $\hat{\theta}$ such that

$$p(t\mid\hat{\theta})\geq p(t\mid\theta) \quad \text{for all } \theta\neq\hat{\theta}. \tag{12}$$

The MLE of (1) is obtained by differentiating (1) with respect to θ and equating 0. It is

$$\hat{\theta}=1/t. \tag{13}$$

Since the MLE (13) is a function of time t whose value has not yet been observed, it is not a constant hazard rate. In order for the hazard rate of the exponential distribution to be constant, one way is to adopt the mean value of a prior distribution for the parameter θ. For example when the gamma distribution (2) is used for the prior distribution, the mean value of θ, if adopted as an estimate of the hazard rate, becomes α / β which is constant in time. When we have observed t and we estimate the hazard rate for the unobserved future observation $\tilde{t}$, the MLE of θ is updated by considering the joint distribution $p(t, \tilde{t} \mid \theta)$ and finding $\hat{\theta}$ maximizing it. The resulting MLE is

$$\hat{\theta} = 2 / (t + \tilde{t}), \qquad (14)$$

likewise, if there are n observed data $t_1, \ldots, t_n$, then we have

$$\hat{\theta} = n / (t_1 + \ldots + t_n + \tilde{t}). \qquad (15)$$

The estimate (14) is still a function of time, in which t has been observed and $\tilde{t}$ has not been observed yet. This estimate does not guarantee a constant hazard rate either. The corresponding mean value of the posterior distribution of (2), once it is adopted as an estimate, is $(\alpha + 1)/(\beta + t)$, which is constant since t has been observed to be known. As an estimate, the mean value of the prior or posterior distribution for the random parameter θ guarantees that the exponential sampling distribution has the constant hazard rate in time.

When there are sufficient data, the MLE of hazard rate of (1), which is evaluated by using only the collection of observed data, will work fine. We, however, often face with the situation in which few data are available due to various reasons. In this situation, the Bayesian approach allows one advantage that a subjective prior distribution can be used to incorporate expert judgement before observing data.

There is also a respect discriminating between the MLE method and the Bayesian approaches. In the MLE method, whenever new data are available, the estimates of the hazard rate are calculated by using the joint distribution of all the new and historical data. For example, the estimate of hazard rate (13) has become (14) from the joint distribution $p(t, \tilde{t} \mid \theta)$. On the other hand, only the new data are used for updating the hazard rates in the Bayesian approach. From this phenomenon we can say that the Bayesian approach includes a learning process.

Comparing between two hazard rates (4) and (6), it can be seen that the Bayesian estimate (6) is the estimate of the true hazard rate (4) and it is simply the posterior mean of the parameter conditioned on the future observation t which is assumed to be known. Both kinds of hazard rate can be used for the reliability analysis related to the unobserved future observation since they are functions of the historical data and the unobserved future observation themselves. It can be asserted that a true hazard rate will work better than its estimated value in performing the reliability analysis.

5. Numerical simulation

Followings are pseudo random data from exp(0.01) generated by random number generator:

83 22 75 34 185 195 144 219 53 45.

Let us assume that failure occurs independently according to the exponential distribution, that is $t \mid \theta \sim \exp(\theta)$. We will estimate the hazard rate from above data by foregoing estimate methods and compare those estimated hazard rates.

For a MLE, from (15) the estimated hazard rate is $\hat{\theta} \cong 0.0095$. In the case of applying the Bayesian approach, first we should determine a prior distribution. There are many kinds of method to make or select a prior distribution [6], but in this paper we use the gamma distribution as described in Section 2 whose hyperparameters are $\alpha = 5$, and $\beta = 467.3576$. Hence, from (8) and (11), the estimated hazard rates by the Bayesian approach are 0.0099 (from mixed distributions) and 0.0105 (from Bayesian estimate). Therefore the hazard rate by using the mixed distribution is the most exact.

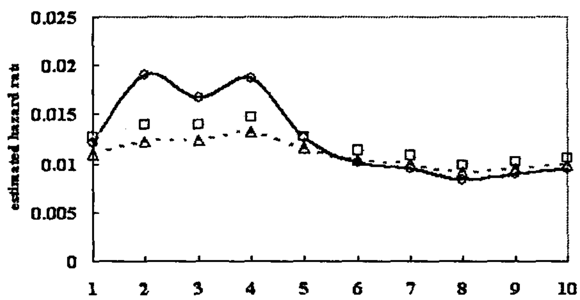

Fig.1. Trend of estimated hazard rate from MLE (line with circles), mixed distribution (line with rectangles) and Bayesian estimate (line with triangles)

Besides Fig.1 shows the trend of three estimated hazard rates corresponding to the addition of data and from Fig.1 we can see that our estimated value can be largely changed unless the sufficient data are obtained. However variation of estimated hazard rates using Bayesian approach is less changeable than that of MLE.

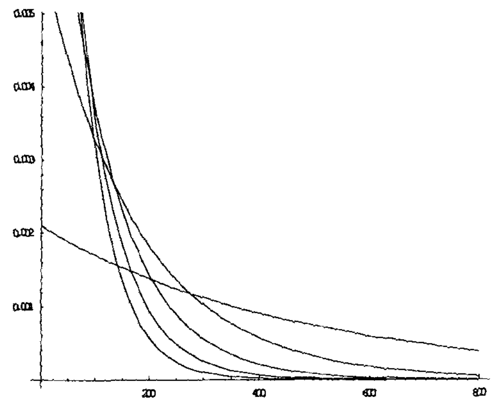

Fig.2. Approximate failure-time distribution corresponding to $\theta=\hat{\theta}\pm 2S.E.(\hat{\theta})$, $\theta=\hat{\theta}\pm S.E.(\hat{\theta})$, and $\theta=\hat{\theta}$

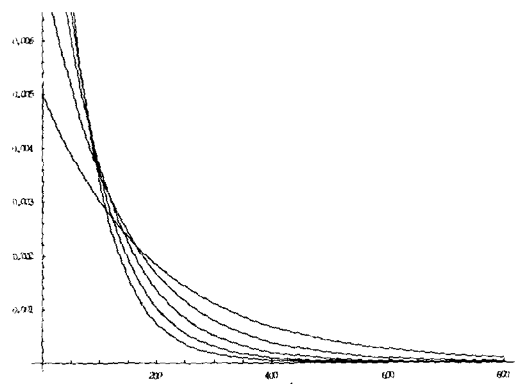

Fig.3. Approximate failure-time distribution from mixed distribution corresponding to $\theta=E[\theta]$, $\theta=E[\theta]\pm S.E.(\theta)$, $\theta=E[\theta]\pm 2S.E.(\theta)$

Parameter θ can be examined by the following consideration. The distribution of the MLE $\hat{\theta}$ is inverse gamma with mean, $E(\hat{\theta})=n\theta/(n-1)$ and variance, $Var(\hat{\theta})=(n\theta)^2/\{(n-1)^2(n-2)\}$. Also, for a Bayesian approach, since θ follows gamma distribution and hazard rate of $\exp(\theta)$ is parameter θ itself, we can show that the hazard rate θ after observing 10 data becomes gamma($\theta|$ 15, 1522.3576). Fig. 2 and Fig.3 show the $f(t|\theta)$ corresponding to the MLE and Mixed distribution. The corresponding mean lifetimes given by $\mu=1/\theta$ vary from 59 to 476 (MLE) hours and 67 to 204 (Mixed distribution). Clearly, decisions based upon assuming, $\theta=\hat{\theta}$ that is MLE, might not be reliable.

Finally Fig.4 illustrates each estimated hazard rate.

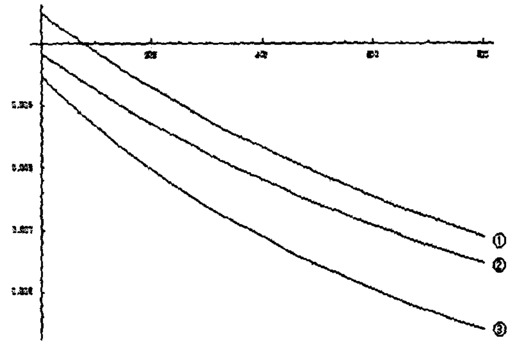

Fig.4. estimated hazard rate function for each method (① Bayesian estimate, ② Mixed distribution, ③ MLE)

6. Conclusions

In many real problems, data must be considered as very important and valuable things to analyze the problems. However analysis only using data is unreliable when there is no sufficient data to perform effective analysis. In order to resolve this situation, we introduced Bayesian approach in this paper. Recently Bayesian approach is considered as a good method to perform in probabilistic analysis. The most important properties of Bayesian approach are two things are learning process and fact that we can use both data and experts' acknowledge. These properties may make Bayesian approach more effective than the traditional statistics in many real engineering field.

References

[1] Barlow RE, Barlow, Proschan F. Statistical Theory of Reliability and Life Testing. MD Silver Spring: To begin with, 1981.

[2] Bernardo JM, Smith AFM. Bayesian Theory. Chichester: Wiley, 1993.

[3] Chauhan RK. Bayesian analysis of reliability and hazard rate function of a mixture model. Microelectronics Reliability 1997; 37: 6: 935-941.

[4] Gelman AB, Carlin JS, Stern HS, Rubin, DB. Bayesian Data Analysis. New York: Chapman & Hall/CRC, 2000.

[5] Martz HF, Waller RA. Bayesian Reliability Analysis. New York: John Wiley & Sons, Inc., 1982.

[6] Percy DF. Bayesian enhanced strategic decision making for reliability. European Journal of Operational Research, 2002; 139: 133-145.