A Study on Comparative Evaluation of Software Reliability Model Applying Modified Exponential Distribution

Hee-Cheul Kim1*

1Associate Professor, Department of Industrial & Management Engineering, Namseoul University, South Korea.
*Corresponding Author
ORCID: 0000-0002-4786-5672

Abstract
This paper, the reliability of software of lifetime distributions based on non-Homogeneous Poisson process were discussed. The lifetime distributions was used the exponential distribution widely used in the field of software reliability and the inverse-exponential distribution which can be commonly applied in the ecological field, Burr-Hatke-exponential distribution which following decrease hazard function. In the present study, the mean square error of the exponential distribution model is smaller than that of the inverse-exponential distribution model and Burr-Hatke-exponential distribution. Also, in the mean function pattern comparison, the exponential distribution model can be seen that the width of the prediction difference is the smallest in the comparison with the true value. Therefore, the exponential distribution model can be regarded as an efficient model in terms of accuracy. However, the estimated value of determination coefficient for the Burr-Hatke-exponential distribution model is larger than the exponential distribution model and the inverse-exponential distribution model. So, the Burr-Hatke-exponential distribution model can be regarded as an efficient model in terms of goodness-of-fit. However, all of the proposed models have an estimated value of more than 90%, so all models are judged as efficient models. Through this study, it is concluded that software operators using the mean square error, the mean value function and the pattern of the hazard function as a preliminary information for detect the software failure phenomenon can be developed the failure type of the lifetime distribution.

Keywords: Software Reliability Model, Mean Squared Error, Burr-Hatke-Exponential distribution, NHPP, Box-plot Test

I. INTRODUCTION
Software systems can be regarded as basic elements of important intellectual factors of industrial management. Such a system can provide high-quality service to software users if it can achieve correctness and reliability. Therefore, software maintenance is to maintain the accuracy and reliability of the software system. The software reliability growth models, which is an engineering study related to such maintenance, have been proposed. These studies have developed a software reliability model that follows the non-homogeneous Poisson process (NHPP), which predicts the failure intensity function and the mean value function using the reliability attribute factors such as the number of failures remaining and the failure characteristics [1]. Software reliability models are used to predict and test software reliability, number of failures remaining, failure intensity, and total cost of software development. Under these circumstances, Yamada and Osaki [2] emphasized that the results of the mean value function can be predicted by using the maximum likelihood estimation method. Also, and the reliability characteristics of the mean value function was presented by explaining the graph of the confidence interval of the mean value function [1, 3]. The defect detection rate developed using the exponential distribution, which is the basic model in this field, consists of the intensity function which is a constant (hazard function) [4]. Kim [5] also studied the reliability of life distribution using Burr-XII and Type-2 Gumbel distributions. In this study with reference to these prior studies, was compared the reliability of the software reliability model by applying the inverse-exponential distribution and the Burr-Hatke-exponential distribution.

II. INFINITE NHPP SOFTWARE RELIABILITY MODEL

II.1 Exponential distribution model
Among the models widely used in the field of software reliability, the most studied model is the Goel-Okumoto model whose life distribution follows the exponential distribution. In this NHPP model, the intensity function and man value function for this model are as follows [4].

\[ \lambda(t | \theta, \beta) = \theta f(t) = \theta \beta e^{-\beta t} \] (1)
\[ m(t | \theta, \beta) = \theta F(t) = \theta (1 - e^{-\beta t}) \] (2)

Note that \( t \in (0, \infty) \) , \( \beta > 0 \) is the shape parameter. In finite failure NHPP model, \( \theta \) was specified the expected value of faults that would be discovered observing time \( (0,t) \). In equations (1) and (2), \( f(t) \) is the probability density function and \( F(t) \) is the cumulative distribution function.
Using equations (1) and (2) and \( \Theta=(\theta, \beta) \) denote the parameter space, the likelihood function of the finite fault NHPP model is as follows [5, 6],

\[
L_{NHPP}(\Theta | \chi) = \prod_{i=1}^{n} \lambda(x_i) \exp[-m(x_i)] \\
= \left( \prod_{i=1}^{n} \theta \beta e^{-\beta x_i} \right) \exp[-\theta(1-e^{-\beta x_i})]
\]  
(3)

Note that \( \chi=(x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n) \). Therefore, using equation (3), the log-likelihood function [5] for using the maximum likelihood estimation is derived as follows.

\[
\ln L_{NHPP}(\Theta | \chi) = -m(x_n) + \sum_{i=1}^{n} \ln \lambda(x_i) \\
= -\theta(1-e^{-\beta x_i}) + n \ln \theta + n \ln \beta - \beta \sum_{i=1}^{n} x_i
\]  
(4)

Thus, maximum likelihood estimators [6] \( \hat{\theta}_{MLE} \) and \( \hat{\beta}_{MLE} \) must be satisfied following conditional expression, using the partial differential equations for parameters \( \theta \) and \( \beta \) in equation (4),

\[
\frac{\partial \ln L_{NHPP}(\Theta | \chi)}{\partial \theta} = \frac{n}{\theta} - 1 + e^{-\beta x_1} = 0
\]  
(5)

\[
\frac{\partial \ln L_{NHPP}(\Theta | \chi)}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^{n} x_i - \theta - \beta x_n e^{-\beta x_1} = 0
\]  
(6)

II. II Inverse exponential distribution model

A widely applicable distribution in medicine, reliability and ecology is the Inverse Weibull distribution [7, 8]. The cumulative distribution function for this distribution is as follows.

\[
F(t) = \exp[-(\beta t)^{-\gamma}]
\]  
(7)

In equation (7), the case of the shape parameter \( \gamma = 1 \) is the inverse exponential distribution. Therefore, the cumulative distribution function and probability density function of the inverse-exponential distribution are as follows [8],

\[
F(t) = \exp[-(\beta t)^{-1}]
\]  
(8)

\[
f(t) = \beta^{-1} t^{-2} \exp[-(\beta t)^{-1}]
\]  
(9)

Note that \( t \in (0, \infty) \), \( \beta > 0 \) is the shape parameter. In finite failure NHPP model, \( \theta \) was specified the expected value of faults that would be discovered observing time \( (0, t) \). Thus, in the finite-failure NHPP model, the intensity function and the mean value function of the NHPP in the inverse exponential distribution model using the equations (8) and (9) are expressed as next forms [8],

\[
\lambda(t | \theta, a, b) = \theta f(t) = \theta^{-1} t^{-2} e^{-\beta t}
\]  
(10)

\[
m(t | \theta, a, b) = \theta F(t) = \theta e^{-\beta t}^{-1}
\]  
(11)

Therefore, the log-likelihood function for using the maximum likelihood estimation is derived as follows.

\[
\ln L_{NHPP}(\Theta | \chi) = -m(x_n) + \sum_{i=1}^{n} \ln \lambda(x_i)
\]

\[
= -\theta e^{-\beta x_1} + n \ln \theta - n \ln \beta - 2 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i (\beta x_i)^{-1}
\]  
(12)

Note that \( \chi=(x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n) \). Therefore, the estimator \( \hat{\theta}_{MLE} \) and \( \hat{\beta}_{MLE} \) must be meted the following condition for the maximum likelihood estimation about each parameter using equation (12).

\[
\frac{\partial \ln L_{NHPP}(\Theta | \chi)}{\partial \theta} = n \frac{n}{\theta} - \beta x_n e^{-\beta x_1} = 0
\]  
(13)

\[
\frac{\partial \ln L_{NHPP}(\Theta | \chi)}{\partial \beta} = \frac{n}{\beta} - n \beta x_n e^{-\beta x_1} = 0
\]  
(14)

II. III Burr-Hatke-exponential distribution model

The hazard function of basic exponential distribution, which characterises the rate of failure per unit time, has a constant independent of the failure time. But Burr-Hatke-exponential distribution in which these patterns of hazard functions display an increasing pattern or a decreasing pattern [9, 10]. The probability density function and cumulative distribution function of this distribution are defined as follows.

\[
F(t | \lambda) = 1 - \frac{e^{-\lambda t}}{1 + \lambda t}, 
\]  
(15)

\[
m(t | \lambda, \beta) = \lambda e^{-\lambda t} \frac{2 + \lambda t}{(1 + \lambda t)^2}
\]  
(16)

\[
\lambda(t | \theta, \lambda) = \theta f(t) = \theta e^{-\lambda t} \frac{2 + \lambda t}{(1 + \lambda t)^2}
\]  
(17)

In equation (16) and equation (17), \( x_n \) is replaced with the last failure time point and the likelihood function is known as follows [8],

\[
L_{NHPP}(\Theta | \chi) = \prod_{i=1}^{n} \lambda(x_i) \exp[-m(x_i)]
\]  
(18)

Note that \( \chi=(x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n) \) and \( \Theta = \{ \theta, \lambda \} \) is the parameter space.
The log-likelihood function by means of the equation (19) can be detailed ensuing relation [2].

\[
L_{NHPP}(\Theta | x) = -m(x_0) + \sum_{i=1}^{n} \ln \lambda (x_i)
\]

\[
= -\theta \left(1 - e^{-\lambda x_0}\right) + n \ln \theta + n \ln \lambda - \lambda \sum_{i=1}^{n} x_i
\]

\[\text{(19)}\]

The estimator \(\hat{\theta}_{MLE}\) and \(\hat{\lambda}_{MLE}\) must be assessed the following construction for the maximum likelihood estimation about all parameter using the equation (19).

\[
\frac{\partial \ln L_{NHPP}(\Theta | x)}{\partial \theta} = \frac{n}{\theta} - \left(1 - e^{-\lambda x_0}\right) = 0
\]

\[\text{(20)}\]

\[
\frac{\partial \ln L_{NHPP}(\Theta | x)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{x_i}{2 + \lambda x_i}
\]

\[\text{(21)}\]

\[\lambda_{MLE} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} \frac{x_i}{2 + \lambda x_i}} \left(1 + \lambda x_i\right)^2 = 0\]

The reliability structures of the software reliability model were studied using the software failure time data [10, 11]. The failure time data is revealed in Table 1. Furthermore, a trend test should be headed in order to assure reliability of data. In this study, the trend analysis was used the Box-plot test [10, 12]. Therefore, in the outcome of Figure 1, three data (28th, 29th, 30th) were excluded from the parameter estimation due to the occurrence of an abnormal value (extreme value). The maximum likelihood estimation so as to the parameter estimation calculation was applied. The results are summarized in Table 2. So as to enable the parameter estimation, in this research, a mathematical transformation data (Failure Time = 1, 0, ... ) was used and calculation method of nonlinear equations was used bisection method. These calculations solve the root exactly, since the initial values were specified 0.0001 and 1.000 and the tolerance value for the measurement of interval (10^-6) were specified, with an accomplished reiteration of 100 times using C-language checking satisfactory convergent.

### Table 1. Failure time data

<table>
<thead>
<tr>
<th>Failure Number</th>
<th>Failure Time (hours)</th>
<th>Failure Time × 10^-1</th>
<th>Failure Number</th>
<th>Failure Time (hours)</th>
<th>Failure Time × 10^-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.02</td>
<td>0.3002</td>
<td>16</td>
<td>151.78</td>
<td>1.5178</td>
</tr>
<tr>
<td>2</td>
<td>31.46</td>
<td>0.3146</td>
<td>17</td>
<td>177.5</td>
<td>1.775</td>
</tr>
<tr>
<td>3</td>
<td>53.93</td>
<td>0.5393</td>
<td>18</td>
<td>180.29</td>
<td>1.8029</td>
</tr>
<tr>
<td>4</td>
<td>55.29</td>
<td>0.5529</td>
<td>19</td>
<td>182.21</td>
<td>1.8221</td>
</tr>
<tr>
<td>5</td>
<td>58.72</td>
<td>0.5872</td>
<td>20</td>
<td>186.34</td>
<td>1.8634</td>
</tr>
<tr>
<td>6</td>
<td>71.92</td>
<td>0.7192</td>
<td>21</td>
<td>256.81</td>
<td>2.5681</td>
</tr>
<tr>
<td>7</td>
<td>77.07</td>
<td>0.7707</td>
<td>22</td>
<td>273.88</td>
<td>2.7388</td>
</tr>
<tr>
<td>8</td>
<td>80.9</td>
<td>0.809</td>
<td>23</td>
<td>277.87</td>
<td>2.7787</td>
</tr>
<tr>
<td>9</td>
<td>101.9</td>
<td>1.019</td>
<td>24</td>
<td>453.93</td>
<td>4.5393</td>
</tr>
<tr>
<td>10</td>
<td>114.87</td>
<td>1.1487</td>
<td>25</td>
<td>535</td>
<td>5.35</td>
</tr>
<tr>
<td>11</td>
<td>115.34</td>
<td>1.1534</td>
<td>26</td>
<td>537.27</td>
<td>5.3727</td>
</tr>
<tr>
<td>12</td>
<td>121.57</td>
<td>1.2157</td>
<td>27</td>
<td>552.9</td>
<td>5.529</td>
</tr>
<tr>
<td>13</td>
<td>124.97</td>
<td>1.2497</td>
<td>28</td>
<td>673.68</td>
<td>6.7368</td>
</tr>
<tr>
<td>14</td>
<td>134.67</td>
<td>1.3467</td>
<td>29</td>
<td>704.49</td>
<td>7.0449</td>
</tr>
<tr>
<td>15</td>
<td>136.25</td>
<td>1.3625</td>
<td>30</td>
<td>738.68</td>
<td>7.3868</td>
</tr>
</tbody>
</table>

![Fig. 1. Box plot test](image-url)
In Figure 2, trend of the hazard function for the exponential distribution model follows the constant form, while the inverse exponential distribution model and the Burr–Hatke-exponential distribution model show decreasing trends.

In Figure 3, all the comparison models are non-increase form in the mean value function pattern comparison, but the exponential distribution model can be seen that the width of the prediction difference is the smallest in the comparison with the true value. In addition, the statistics of the mean square error ($MSE$) [14, 15], which represent a measure of the difference between the actual value and the predicted value, are as follows.

$$MSE = \frac{\sum_{i=1}^{n} [m(x_i) - \hat{m}(x_i)]^2}{n - k} \quad (26)$$

Note that $m(x_i)$ is the cumulated true number of the faults detected in $(0, x_i]$, $n$ states the number of realizing values and $k$ is the number of the parameter.

In Table 2, because the mean square error of exponential distribution model is smaller than the inverse-exponential and Burr–Hatke-exponential models, thus the exponential distribution model than the other distribution models is efficient. Specifically, the squared error for each failure point can be derived as follows using equation (26) [14].

$$SE = (m(x_i) - \hat{m}(x_i))^2, \quad i = 1, 2 \ldots 27 \quad (27)$$

Therefore, the square error for each time point using (27) is shown in Figure 4. In this figure, the exponential distribution model shows a smaller square error value than the inverse-exponential model and the Burr-Hatke-exponential distribution model at each failure time. However, as the time of failure increases, the width of the gap decreases gradually.

In addition, the coefficient of determination ($R^2$) is defined as an explanatory tools to explain the number of failures as the forecasting value. Thus the model with a large coefficient of determination is stared as an efficient model in terms of goodness-of-fit [5, 12].

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (m(x_i) - \hat{m}(x_i))^2}{\sum_{i=1}^{n} \left( m(x_i) - \frac{\sum_{j=1}^{n} m(x_j)}{n} \right)^2} \quad (28)$$

Thus, Burr–Hatke-exponential model can be regarded as an efficient model in terms of goodness-of-fit because the estimated value for coefficient of determination in Table 2 has a larger estimated value than the exponential model and the inverse-exponential model. However, because the proposed model also has 95% or more, all models are regarded as efficient models [13, 14].

In the NHPP model, the software failure happens at last test failure time $x_{27} = 5.529$ and reliability which is the probability
that the software failure does not occur between 5.529 and 5.529+t (where t is the mission time) can be specified using the following construction [13, 14].

\[
\hat{R}(t \mid x_{27} = 5.529) = e^{-\int_{5.529}^{5.529+t} \lambda(q) \, dq} = \exp \left[ -\left\{ m(t + 5.529) - m(5.529) \right\} \right]
\] (29)

In the form of the reliability function in Figure 5 using the equation (29), gradually seems as a non-increasing pattern as the mission time elapses. In terms of reliability, the inverse-exponential distribution model is higher than the Burr–Hatke-index distribution model and exponential distribution model shows.

![Fig. 5. Transition of reliability pattern](image)

**IV. CONCLUSION**

Software systems can be regarded as basic elements of important intellectual factors of industrial management. Such a system can provide high-quality service to software users if it can achieve correctness and reliability. Therefore, software maintenance is to maintain the accuracy and reliability of the software system. The software reliability growth model, which is an engineering study related to such maintenance, has been proposed. By quantitatively modelling the trend of failure during the software development process or the actual software operation process, it is possible to evaluate the efficiency by comparing and analysing software accuracy and reliability. In this study, it was applied exponential distribution widely used in the field of software reliability, inverse-exponential distribution widely applicable in the field of ecology and Burr–Hatke-exponential distribution which shows decreasing pattern of the hazard function. This models was compared and discussed the reliability properties of the software using the non-homogeneous Poisson process. The results of this study can be summarized as follows.

First, the hazard function of the exponential distribution model follows the constant form, while the inverse-index distribution model and the Burr–Hatke-index distribution model show a decreasing trend. Second, in comparison of mean function patterns, all the comparison models are non-increase form in the mean function pattern comparison, but the exponential distribution model can be seen that the width of the prediction difference is the smallest in the comparison with the true value.

Third, because the mean squared error is smaller than the inverse-exponential and Burr–Hatke-exponential distribution models, the exponential distribution model can be regarded as an efficient model with higher accuracy than other distribution models.

Fourth, Burr–Hatke-exponential model can be regarded as an efficient model because the estimated value for coefficient of determination has a larger estimated value than the exponential model and the inverse-exponential model in terms of goodness-of-fit. However, because the proposed model also has 95% or more, all models are regarded as efficient models. Fifth, in the form of the reliability function, as the mission time elapses, the non-increasing pattern gradually appears, and the inverse-exponential distribution model is higher than the Burr–Hatke-exponential distribution model and the exponential distribution model.

Through this study, software operators can use the information on the mean square error, mean value, and risk function to identify the software failure type using the characteristics of the life distribution of the modified exponential distribution.

**ACKNOWLEDGEMENTS**

Funding for this paper was provided by Namseoul University

**REFERENCES**


