A Comparative Study on the Property of NHPP Software Reliability Model Based on Shape Parameters of Exponential Family Lifetime Distribution

Seung Kyu Park

Namseoul University, Department of Electronic Engineering, Chungnam, 31020, Korea.

Abstract

In this study, after applying the exponential family distribution (Goel-Okumoto, Rayleigh, Erlang) which are widely used in the field of reliability to the finite failure NHPP software reliability model, we compared and analyzed the reliability property based on shape parameters of the lifetime distribution. Software failure time data was used to identify software failure phenomena, and parametric estimation was applied to the maximum likelihood estimation method. As a result, in terms of the intensity function, the Rayleigh model was more efficient than the other models because the intensity function significantly decreased as the failure time passed. In the pattern of the mean value function, the Rayleigh model showed a slightly overestimated pattern compared to the true value, but it was more efficient than the Erlang model because of the smaller error. Also, as a result of comparing reliability by applying future mission time, the Rayleigh model was high and stable together with the Erlang model, but the Goel-Okumoto model showed a decreasing tendency. In conclusion, we found that the Rayleigh model is an efficient model with the best performance among the proposed models. Through this study, we have newly analyzed the property of software reliability model with the exponential family lifetime distribution without existing research case, and it was able to present new research information that software developers could use as basic guidelines.

Keywords: Erlang Distribution, Exponential Family, Goel-Okumoto Basic Model, NHPP Software Reliability, Rayleigh Distribution, Shape Parameter.

1. INTRODUCTION

With the rapid development of the software technology industry, the scale of the software system is constantly expanding and complicated. For this reason, the need for software reliability research that can process large amounts of information accurately without failures is also rapidly increasing. The software reliability, defined as the probability that the software will function normally without failure for a specified amount of time in a particular environment, is the most important software quality standard [1]. Until now, the software reliability models based on the non-homogeneous Poisson process (NHPP) have been proposed to predict software reliability. In particular, to estimate the reliability attributes, the software reliability models based on finite failure NHPP model using the intensity function and the mean value function were developed and proposed [2]. Yamada and Osaki [3] stated that the results of the mean value function can be estimated using the maximum likelihood estimation method, while Huang [4] presented and explained a graph

showing the confidence interval of the mean value function. Pham and Zhang [5] proposed a software reliability cost model based on the software reliability model and software failure time. Also, Yang [6] analyzed the attributes of the finite failure NHPP software reliability model based on Weibull lifetime distribution. Kim [7] compared and analyzed the attributes of the software reliability model using Burr-XⅡ and Type-2 Gumbel lifetime distribution.

Therefore, in this study, after applying the exponential family distribution which is widely used in the field of reliability to the finite failure NHPP software reliability model, we were newly analyzed the reliability property of the proposed models and will present the optimal model through this results.

2. RELATED RESEARCH

2.1 NHPP Software Reliability Model

The NHPP model contains property about mean value m(t) and intensity pattern $\lambda(t)$. N(t) is the cumulative number of failures of the software detected up to time t, m(t) is a mean value function when $\lambda(t)$ is expressed by an intensity function, the cumulative failure number N(t) follows a Poisson probability density function having a parameter m(t). The NHPP software reliability model is a model that measures the reliability using the mean failure rate function by the number of failures generated per unit time. That is

$$P\{N(t) = n\} = \frac{[m(t)]^n \cdot e^{-m(t)}}{n!} \qquad (1)$$

Note. $n = 0,1,2, \cdots \infty$.

The mean value function m(t) and the intensity function $\lambda(t)$ of the NHPP model are as follows.

$$m(t) = \int_0^t \lambda(s)ds \qquad (2)$$

$$\frac{dm(t)}{d(t)} = \lambda(t) \qquad (3)$$

In terms of software reliability, the mean value function represents a software failure occurrence expected value, the intensity function is the failure rate function and means the failure occurrence rate per defect. Also, the time domain NHPP models are classified into a finite failure that the failure does not occur at the time of repairing the failure, and an infinite failure that the failure occurs at the time of repairing

failure. In this study, we will analyze the software reliability property based on finite failure cases. That is, in the finite-failure NHPP model, if the expected value of the failure that can be found up to time [0, t] is θ, then the mean value function and the intensity function are as follows.

$$m(t|\theta, b) = \theta F(t) \tag{4}$$

$$\lambda(t|\theta, b) = \theta F(t)' = \theta f(t) \tag{5}$$

Considering Eq. 4 and Eq. 5, the likelihood function of the finite-failure NHPP model is derived as follows.

$$L_{NHPP}(\Theta|\underline{x}) = \left(\prod_{i=1}^{n} \lambda(x_i)\right) \exp[-m(x_n)] \tag{6}$$

Note. $\underline{x} = (x_1, x_2, x_3 \cdots x_n)$.

2.2 Goel-Okumoto basic model : NHPP model

The Goel-Okumoto is a well-known basic model in the software reliability field. This model assumes the exponential distribution as the lifetime distribution per fault. Therefore, the rate of occurrence of faults is constant. Let f(t) and F(t) for the Goel-Okumoto model be a probability density function and a cumulative density function, respectively.

Assuming that the expected value of the number of failures of the observation point [0, t] is θ, the mean value function and the intensity function of the finite fault NHPP Goel-Okumoto model are as follows [8].

$$m(t|\theta, b) = \theta F(t) = \theta(1 - e^{-bt}) \tag{7}$$

$$\lambda(t|\theta, b) = \theta f(t) = \theta b e^{-bt} \tag{8}$$

Note. $\theta > 0, \ b > 0$.

Substituting Eq. 7 and Eq. 8 into Eq. 6, the likelihood function of the finite-failure NHPP model is derived as follows.

$$L(\theta, b|\underline{x}) = \left(\prod_{i=1}^{n} \theta b e^{-bx_i}\right) \exp[-\theta(1 - e^{-bx_n})] \tag{9}$$

Note. $\underline{x} = (0 \le x_1 \le x_2 \le \cdots \le x_n)$.

The log-likelihood function, using Eq. 9, is simplified to the following log conditional expression.

$$\ln L_{NHPP}(\Theta|\underline{x}) =$$

$$n\ln\theta + n\ln b - b\sum_{k=1}^{n} x_k - \theta(1 - e^{-bx_n}) \tag{10}$$

Note. $\underline{x} = (0 \le x_1 \le x_2 \le \cdots \le x_n, \ \Theta$ is parameter space.

Therefore, in the Eq. 11 and Eq. 12, the maximum likelihood estimator $\hat{\theta}_{MLE}$ and $\hat{b}_{MLE}$ can be estimated by a numerical method (bisection method).

$$\frac{\partial \ln L_{NHPP}(\Theta|\underline{x})}{\partial \theta} = \frac{n}{\hat{\theta}} - 1 + \exp(-\hat{b}x_n) = 0 \tag{11}$$

$$\frac{\partial \ln L_{NHPP}(\Theta|\underline{x})}{\partial b} = \frac{n}{\hat{b}} - \sum_{i=1}^{n} x_i - \hat{\theta}x_n \exp(-\hat{b}x_n) \tag{12}$$

Note. $x = (x_1, x_2, \ x_3 \cdots x_n)$

2.3 Rayleigh distribution : NHPP model

The Rayleigh distribution is the lifetime distribution of the exponential family widely used in the reliability field. The probability density function and the cumulative distribution function considering the shape parameter(α) are as follows.

$$f(t) = \frac{t^{\alpha-1}}{\beta^2} e^{-\frac{t^\alpha}{2\beta^2}} \tag{13}$$

$$F(t) = 1 - e^{-\frac{t^\alpha}{2\beta^2}} \tag{14}$$

Note. $\beta > 0, \ t \in [0, \ \infty]$.

In order to simplify Eq. 13 and Eq. 14, if substitution by the equation $\frac{1}{2\beta^2} = b$ is as follows.

$$f(t) = 2bt^{a-1}e^{-bt^\alpha} \tag{15}$$

$$F(t) = 1 - e^{-bt^\alpha} \tag{16}$$

Note. $b > 0, \ t \in [0, \ \infty]$.

In the Eq. 15 and Eq. 16, a Rayleigh distribution is obtained when the shape parameter(α) is 2. Therefore, the mean value function and the intensity function of the finite fault NHPP Rayleigh model are as follows [9].

$$m(t|\theta, b) = \theta F(t) = \theta(1 - e^{-bt^2}) \tag{17}$$

$$\lambda(t|\theta, b) = \theta f(t) = 2\theta b t e^{-bt^2} \tag{18}$$

Note. $\theta > 0, \ b > 0$.

The log-likelihood function to maximum likelihood estimation by using Eq. 17 and Eq. 18 is derived as follows.

$$\ln L_{NHPP}(\Theta|\underline{x}) = n\ln 2 + n\ln\theta + n\ln b + \sum_{i=1}^{n} \ln x_i$$

$$- b\sum_{i=1}^{n} x_i^2 - \theta\left(1 - e^{-bx_n^2}\right) \tag{19}$$

Note. $\underline{x} = (0 \le x_1 \le x_2 \le \cdots \le x_n, \ \Theta$ is parameter space.

Therefore, in the Eq. 20 and Eq. 21, the maximum likelihood estimator $\hat{\theta}_{MLE}$ and $\hat{b}_{MLE}$ can be estimated by a numerical method (bisection method).

$$\frac{\partial \ln L_{NHPP}(\Theta|\underline{x})}{\partial \theta} = \frac{n}{\hat{\theta}} - 1 + \exp(-\hat{b}x_n^2) = 0 \tag{20}$$

$$\frac{\partial \ln L_{NHPP}(\Theta|\underline{x})}{\partial b} = \frac{n}{\hat{b}} - \sum_{i=1}^{n} x_i^2 - \hat{\theta}x_n^2 \exp(-\hat{b}x_n^2) = 0 \quad (21)$$

Note. $x = (x_1, x_2, x_3 \cdots x_n)$.

2.4 Erlang distribution : NHPP model

The Erlang distribution is the lifetime distribution of the exponential family widely used in the reliability field. The probability density function and the cumulative density function considering the shape parameter (a) and the scale parameter (b) are as follows.

$$f(t) = \frac{b^a}{\Gamma(a)} t^{a-1} e^{-bt} \quad (22)$$

$$F(t) = \left(1 - e^{-bt} \sum_{i=0}^{a-1} \frac{(bt)^i}{i!}\right) \quad (23)$$

Note. $a, b > 0, \ a = 1,2,3 \cdots, \ t \in [0, \infty]$

In this study, we consider the case of the shape parameter $a = 2$. The log-likelihood function to maximum likelihood estimation by using Eq. 22 and Eq. 23 is derived as follows.

$$\ln L_{NHPP}(\Theta|\underline{x}) =$$

$$nln\theta - nln\Gamma(a) + naln b + (a-1)\sum_{i=1}^{n} lnx_i - b\sum_{i=1}^{n} x_i$$

$$-\theta + \theta e^{-bx_n}\left(\sum_{i=0}^{a-1} \frac{(bx_n)^i}{i!}\right) \quad (24)$$

Note. $\underline{x} = (0 \le x_1 \le x_2 \le \cdots \le x_n, \ \Theta$ is parameter space.

Therefore, in the Eq. 25 and Eq. 26, the maximum likelihood estimator $\hat{\theta}_{MLE}$ and $\hat{b}_{MLE}$ can be estimated by a numerical method (bisection method).

$$\frac{\partial \ln L_{NHPP}(\Theta|\underline{x})}{\partial \theta} = \frac{n}{\theta} - 1 + e^{-bx_n}(1 + bx_n) = 0 \quad (25)$$

$$\frac{\partial \ln L_{NHPP}(\Theta|\underline{x})}{\partial b} = \frac{2n}{b} - \sum_{i=1}^{n} x_i - \theta bx_n^2 e^{-bx_n} = 0 \quad (26)$$

Note. $x = (x_1, x_2, x_3 \cdots x_n)$.

3. SOFTWARE FAILURE TIME ANALYSIS AND SOLUTIONS

Let compare and analyze the performance property of the proposed models using the software failure time data [10] as shown in Table 1. This software failure time is the data that

was occurred 30 times in 738.68 unit time.

Table 1. Software Failure Time Data

Failure Number	Failure Time (hours)	Failure Number	Failure Time (hours)
1	30.02	16	151.78
2	31.46	17	177.50
3	53.93	18	180.29
4	55.290	19	182.21
5	58.720	20	186.34
6	71.920	21	256.81
7	77.070	22	273.88
8	80.900	23	277.87
9	101.90	24	453.93
10	114.87	25	535.00
11	115.34	26	537.27
12	121.57	27	552.90
13	124.97	28	673.68
14	134.07	29	704.49
15	136.25	30	738.68

Laplace trend test was used to verify the reliability of the software failure time data as shown in Figure 1.
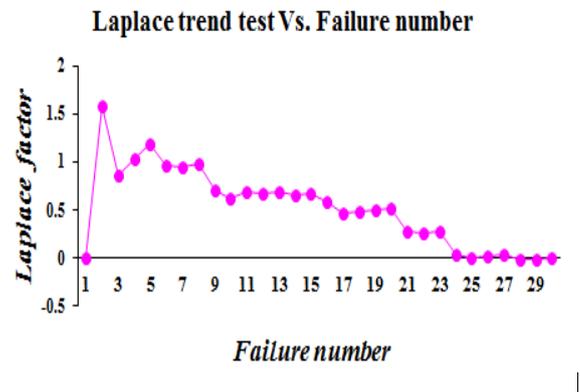

Fig 1. Estimation Results of Laplace Trend Test

In general, if the Laplace factor estimates are distributed between -2 and 2, the data are reliable because the extreme values do not exist and are stable. The estimated value of the Laplace factor was distributed between 0 and 2, as shown in Figure 1. Therefore, it is possible to apply this data because there is no extreme value [11].

In this study, the maximum likelihood estimation (MLE) was used to perform parameter estimation. And, numerical conversion data (Failure time[hours] $\times 10^{-2}$) to facilitate the

parameter estimation was used, and the results are shown in Table 2.

Table 2. Parameter Estimation of Each Model

Model	MLE		Model comparison	
			MSE	R^2
Goel-Okumoto	$\hat{\theta} = 33.4092$	$\hat{b} = 0.3090$	5.8424	0.9814
Rayleigh	$\hat{\theta} = 24.0116$	$\hat{b} = 0.3707$	8.6587	0.9725
Erlang	$\hat{\theta} = 30.5978$	$\hat{b} = 0.0079$	53.8545	0.8293

Explanatory notes. $MLE : Maximum\ Likelihood\ Estimation,$

$MSE : Mean\ Square\ Error.\ R^2 : Coefficient\ of\ Determination.$

As the basis for determining the efficient model, the mean square error(MSE) is defined as follows [12].

$$MSE = \frac{\sum_{i=1}^{n} \left(m(x_i) - \hat{m}(x_i)\right)^2}{n - k} \qquad (27)$$

Note that $m(x_i)$ is the total accumulated number of errors observed within time is $(0, x_i)$, $\hat{m}(x_i)$ is the estimated cumulative number of errors at time x_i obtained from the fitting mean value function, n is the number of observations and k is the number of parameters to be estimated. When selecting an efficient model, the smaller the value of the mean square error, the more efficient the model.

The coefficient of determination (R^2) is a measuring value to the explanatory power of the difference between the target value and the observed value. When selecting an efficient model, the larger the value of the decision coefficient, the more efficient the model [13].

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

As shown in Table 2, we can see that the Rayleigh model is more efficient than the Erlang model. But, the Goel-Okumoto model has the largest coefficient of determination and the smallest mean square error is more efficient than the other models.

Figure 2 shows the transition of mean square error according to each failure number. That is, in this figure, the Goel-Okumoto model shows better estimates than the other model in the total range of failure numbers.
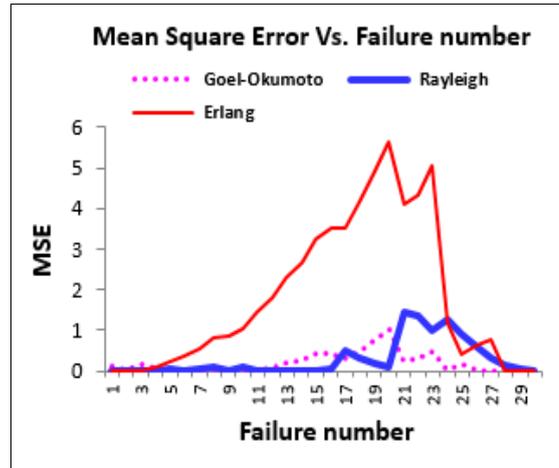

Fig 2. Transition of Mean Square Error

Also, the mean squared error of the Rayleigh model shows a trend of the comparative small error, which is more efficient than the Erlang model in terms of fitness.

Figure 3 shows trends in the strength function, which is the instantaneous failure rate. The Rayleigh model shows the greatest decreasing tendency as the failure time passes, indicating that it is an efficient model. But, only the Erlang model showed an increasing pattern.
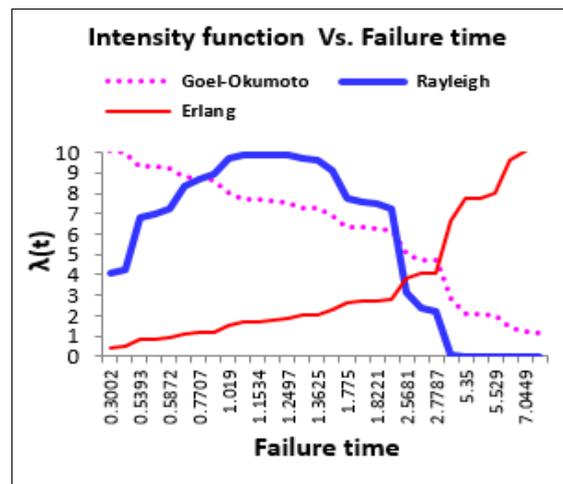

Fig 3. Transition of Intensity Function $\lambda(t)$

Figure 4 shows the pattern trend for the mean value function, which is the failure occurring expected value. In this figure, the Erlang model has underestimated significantly from the difference between the true values, but the Rayleigh model has a slightly overestimated pattern. That is, the Rayleigh model is more efficient than the Erlang model because the error value is estimated to be small.
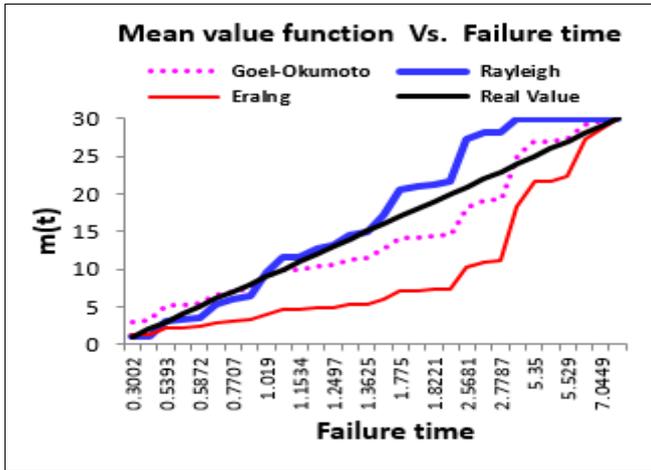
Fig 4. Pattern of Mean Value Function $m(t)$

Let analyze the reliability property of the proposed models for future mission time.

Here, reliability $(\hat{R})$ is the probability that a software failure will occur when testing at $x_n = 738.68 \times 10^{-2}$ and no software failure will occur between confidence intervals $[x_n, x_n + \tau]$ (where τ is the future mission time). Therefore, the reliability of future mission time is as follows [14].

$$\hat{R}(\tau|x_n) = e^{-\int_{x_n}^{x_{n+\tau}} \lambda(\tau)d\tau} \qquad (29)$$
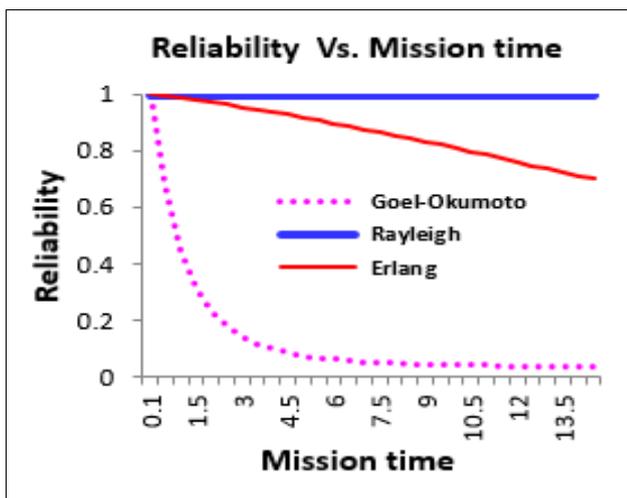$$= \exp[-\{m(\tau + x_n) - m(x_n)\}]$$

Fig 5. Transition of Reliability $\hat{R}(t)$

As shown in Fig 5, the Rayleigh model and the Erlang model show a higher and stable reliability trend than the Goel-Okumoto model having the lower reliability as the mission time passed. That is, in terms of reliability, the Rayleigh model is further reliable than the other models because the reliability is the highest.

4. CONCLUSION

It is possible to efficiently improve the software quality by analyzing the reliability after quantitatively modeling the occurrence of the failure in the software development process. In this study, based on the finite-fault NHPP model with software failure time data, we compared and analyzed the reliability property of the exponential family distribution considering shape parameters of failure lifetime distribution.

The results of this study can be summarized as follows.

First, in terms of intensity function, the Rayleigh model was effective because the occurrence rate of the failure decreased as the failure time passed. But, the Erlang model increases inversely as the failure time passed.

Second, in the pattern of the mean value function, the Erlang model underestimates the error from the true value, but the Rayleigh model shows a slightly overestimated pattern. Therefore, it is found that the Rayleigh model is more efficient than the Erlang model because of the small error width.

Third, in terms of reliability, the Rayleigh model and the Erlang model show stable and high-reliability trends. On the other hand, the Goel-Okumoto basic model shows that future reliability decreases as the mission time passed.

In conclusion, a comprehensive analysis of these simulations results shows that the Rayleigh model is the most efficient model among the proposed models.

As a result, through this study, together with a new analysis of the reliability performance property of the proposed model, we were able to provide the research information that software developers can use as a basic design guideline.

Also, future research will be needed to find the optimal model through the reliability performance analysis after applying the same type of software failure time data to various software lifetime distribution.

ACKNOWLEDGEMENT

REFERENCES

[1] Gokhale, S. S. and Trivedi, K. S. A, "time/structure based software reliability model," Annals of Software Engineering. 8, pp. 85-121, 1999.

[2] K. Y. Song, I.H. Chang, H. Pham, "A Software Reliability Model with a Weibull Fault Detection Rate Function Subject to Operating Environments," Applied Science, Vol. 7, No. 983, pp. 1-16, 2017.

[3] S. Yamada, S. Osaki, "Software reliability growth model -ing : models and applications," IEEE Transactions on Software Engineering, Vol. 11, No. 12, pp. 1431-1437, 1985.

[4] C. Y. Huang, "Performance analysis of softwarer reliabi- ity growth models with testing-effort and

change-point," Journal of Systems and Software, Vol. 76, No. 2, pp. 181- 194, 2005.

[5] Pham H, Zhang X., "NHPP software re aliability and cost models with testing coverage," Eur.J. Oper. Res, 145, pp. 445-454, 2003.

[6] Tae-Jin Yang, "A Study on the Reliability Performance Analysis of Finite Failure NHPP Software Reliability Model Based on Weibull Life Distribution," International Journal of Engineering Research and Technology, Vol. 12, No. 11, pp. 1890-1896, 2019.

[7] Hee-Cheul Kim, "A Comparison Analysis about Reliability Features of Software Reliability Model Using Burr-XⅡ and Type-2 Gumbel Lifetime Distribution," International Journal of Engineering Research and Technology, Vol. 12, No. 1, pp. 73-78, 2019.

[8] Goel A L, Okumoto K, "Time-dependent fault detection rate model for software and other performance measures," IEEE Transactions on Software Engineering, Vol. 28, pp. 206-211, 1978.

[9] Hyun-Dae Shin and Hee-Cheul Kim, "Assessing Rayleigh software reliability model based on NHPP using SPCC," J. Next Gen. Information Technology, Vol. 5, No. 77-84, 2014.

[10] R. Satya Prasad, K. R. H. Rao and R. R. L Kantham, "Software Reliability Measuring using Modified Maximum Like-lihood Estimation and SPC," International Journal of Computer Applications (0975–8887), Vol. 21, No. 7, pp. 1-5, 2011.

[11] Tae-Jin Yang, J.G. Park "A Comparative Study of the Software NHPP Based on Weibull Extension Distribution and Flexible Weibull Extension Distribution," International Journal of Soft Computing, Vol. 11, No. 4, pp. 276~281, 2016.

[12] Kuei-Chen, C., Yeu-Shiang, H., and Tzai-Zang, L., "A Study of Software Reliability growth from the perspective of learning effects," Reliability Engineering and System Safety 93, pp. 1410-1421, 2008

[13] Tae-Jin Yang, "The Analysis and Predict of Software Failure Time Based on Nonlinear Regression," Journal of Engineering and Applied Sciences, Vol. 13, Issue 12, pp. 4376-4380, 2018.

[14] Tae-Jin Yang, "A Comparative Study on Reliability Attributes of Software Reliability Model Based on Type-2 Gumbel and Erlang Life Distribution," Journal of Engineering and Applied Sciences, Vol. 14, Issue 10, pp. 3366-3370, 2019.