

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

Hee-Cheul Kim^{1*}

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].

$$\lambda(t | \theta, \beta) = \theta f(t) = \theta \beta e^{-\beta t} \quad (1)$$

$$m(t | \theta, \beta) = \theta F(t) = \theta (1 - e^{-\beta t}) \quad (2)$$

Note that $t \in (0, \infty]$, $\beta > 0$ is the shape parameter. In finite failure NHPP model, θ 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].

$$L_{NHPP}(\Theta | \underline{x}) = \prod_{i=1}^n \lambda(x_i) \exp[-m(x_n)] \quad (3)$$

$$= \left(\prod_{i=1}^n \theta \beta e^{-\beta x_i} \right) \times \exp[-\theta(1 - e^{-\beta x_n})]$$

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

$$\ln L_{NHPP}(\Theta | \underline{x}) = -m(x_n) + \sum_{i=1}^n \ln \lambda(x_i)$$

$$= -\theta(1 - e^{-\beta x_n}) + n \ln \theta + n \ln \beta - \beta \sum_{i=1}^n x_i \quad (4)$$

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 θ and β in equation (4),

$$\frac{\partial \ln L_{NHPP}(\Theta | \underline{x})}{\partial \theta} = \frac{n}{\theta} - 1 + e^{-\beta x_n} = 0 \quad (5)$$

$$\frac{\partial \ln L_{NHPP}(\Theta | \underline{x})}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^n x_i - \theta \beta x_n e^{-\beta x_n} = 0 \quad (6)$$

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.

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

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].

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

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

Note that $t \in (0, \infty]$, $\beta > 0$ is the shape parameter. In finite failure NHPP model, θ 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].

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

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

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

$$\ln L_{NHPP}(\Theta | \underline{x}) = -m(x_n) + \sum_{i=1}^n \ln \lambda(x_i)$$

$$= -\theta e^{-(\beta t)^{-1}} + n \ln \theta - n \ln \beta - 2 \sum_{i=1}^n x_i - \sum_{i=1}^n x_i (\beta x_i)^{-1} \quad (12)$$

Note that $\underline{x} = (x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n)$. 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).

$$\frac{\partial \ln L_{NHPP}(\Theta | \underline{x})}{\partial \theta} = \frac{n}{\theta} - e^{-(\beta t)^{-1}} = 0 \quad (13)$$

$$\frac{\partial \ln L_{NHPP}(\Theta | \underline{x})}{\partial \beta} = -\frac{n}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^n x_i - \theta \frac{1}{\beta^2 x_n} e^{-(\beta x_n)^{-1}} = 0 \quad (14)$$

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.

$$F(t | \lambda) = 1 - \frac{e^{-\lambda t}}{1 + \lambda t}, \quad f(t | \lambda) = \lambda e^{-\lambda t} \frac{2 + \lambda t}{(1 + \lambda t)^2} \quad (15)$$

Note that $t \in (0, \infty]$ and $\lambda > 0$ is the shape parameter. In finite failure NHPP model, θ was specified the expected value of faults that would be discovered observing time $(0, t]$. Thus the intensity function and the mean value function of NHPP are known as follows [5].

$$\lambda(t | \theta, \lambda) = \theta f(t) = \theta \lambda e^{-\lambda t} \frac{2 + \lambda t}{(1 + \lambda t)^2} \quad (16)$$

$$m(t | \theta, \lambda) = \theta F(t) = \theta \left[1 - \frac{e^{-\lambda t}}{1 + \lambda t} \right] \quad (17)$$

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].

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

$$= \left[\prod_{i=1}^n \theta \lambda e^{-\lambda x_i} \frac{2 + \lambda x_i}{(1 + \lambda x_i)^2} \right]$$

$$\times \exp \left[-\theta \left(1 - \frac{e^{-\lambda x_n}}{1 + \lambda x_n} \right) \right] \quad (18)$$

Note that $\underline{x} = (x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n)$ and $\Theta = \{\theta, \lambda\}$ is the parameter space.

The log-likelihood function by means of the equation (19) can be detailed ensuing relation [2].

$$\begin{aligned} \ln L_{NHPP}(\Theta | \underline{x}) &= -m(x_n) + \sum_{i=1}^n \ln \lambda(x_i) \\ &= -\theta \left(1 - \frac{e^{-\lambda x_n}}{1 + \lambda x_n} \right) + n \ln \theta + n \ln \lambda - \lambda \sum_{i=1}^n x_i \\ &+ \sum_{i=1}^n \ln(2 + \lambda x_i) - 2 \sum_{i=1}^n \ln(1 + \lambda x_i) \end{aligned} \quad (19)$$

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).

$$\frac{\partial \ln L_{NHPP}(\Theta | \underline{x})}{\partial \theta} = \frac{n}{\theta} - \left(1 - \frac{e^{-\lambda x_n}}{1 + \lambda x_n} \right) = 0 \quad (20)$$

$$\begin{aligned} \frac{\partial \ln L_{NHPP}(\Theta | \underline{x})}{\partial \lambda} &= \frac{n}{\lambda} - \sum_{i=1}^n x_n + \sum_{i=1}^n \frac{x_i}{2 + \lambda x_i} \\ -2 \sum_{i=1}^n \frac{x_i}{1 + \lambda x_i} - \theta x_n e^{-\lambda x_n} \frac{(2 + \lambda x_i)}{(1 + \lambda x_i)^2} &= 0 \end{aligned} \quad (21)$$

III. SOFTWARE FAILURE TIME RELIABILITY ANALYSIS USING MODIFIED EXPONENTIAL LIFE DISTRIBUTION

Table 1. Failure time data

Failure Number	Failure Time (hours)	Failure Time $\times 10^{-2}$	Failure Number	Failure Time (hours)	Failure Time $\times 10^{-2}$
1	30.02	0.3002	16	151.78	1.5178
2	31.46	0.3146	17	177.5	1.775
3	53.93	0.5393	18	180.29	1.8029
4	55.29	0.5529	19	182.21	1.8221
5	58.72	0.5872	20	186.34	1.8634
6	71.92	0.7192	21	256.81	2.5681
7	77.07	0.7707	22	273.88	2.7388
8	80.9	0.809	23	277.87	2.7787
9	101.9	1.019	24	453.93	4.5393
10	114.87	1.1487	25	535	5.35
11	115.34	1.1534	26	537.27	5.3727
12	121.57	1.2157	27	552.9	5.529
13	124.97	1.2497	28	673.68	6.7368
14	134.07	1.3407	29	704.49	7.0449
15	136.25	1.3625	30	738.68	7.3868

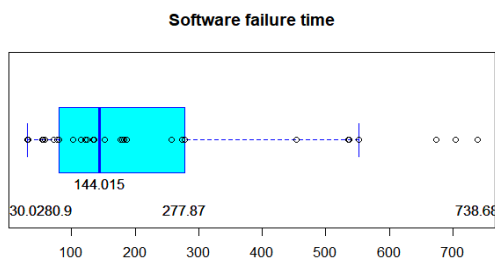


Fig. 1. Box plot test

In this section, 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 was 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 ($Failure\ Time \times 10^{-2}$) 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^{-5}) were specified, with an accomplished reiteration of 100 times using C-language checking satisfactory convergent.

Table 2. MLE, MSE and R^2 for the each model

Model	MLE	Model Comparison	
		MSE	R^2
Exponential distribution	$\hat{\beta}_{MLE} = 0.4809$ $\hat{\theta}_{MLE} = 29.0332$	3.5739	0.9655
Inverse exponential distribution	$\hat{\beta}_{MLE} = 1.6984$ $\hat{\theta}_{MLE} = 30.3914$	29.0184	0.9369
Burr-Hatke-exponential distribution	$\hat{\lambda}_{MLE} = 0.2991$ $\hat{\theta}_{MLE} = 29.0996$	6.4349	0.9749

Note. MLE : Maximum likelihood estimation. MSE : Mean square error. R^2 : Coefficient of determination.

The hazard function, which means the instantaneous failure rate for the specified failure time, is defined by the following pattern [5, 13].

$$h(t) = \frac{f(t)}{1 - F(t)} \quad (22)$$

Note that probability distribution function is $f(t)$ and $F(t)$ means distribution function. Using equation (22), the hazard function for the exponential distribution model, Inverse exponential distribution and Burr-Hatke-exponential distribution is defined as follows.

$$h_{\text{Exponential}}(t) = \beta \quad (23)$$

$$h_{\text{Inverse exponential}}(t) = \frac{\beta^{-1} t^{-2} e^{-(\beta t)^{-1}}}{1 - e^{-(\beta t)^{-1}}} \quad (24)$$

$$h_{\text{Burr-Hatke-exponential}}(t) = \frac{\lambda(2 + \lambda t)}{1 + \lambda t} \quad (25)$$

Therefore, the hazard functions using the parameter estimates in equations (23), (24), (25) and Table 2 are shown in Figure 2.

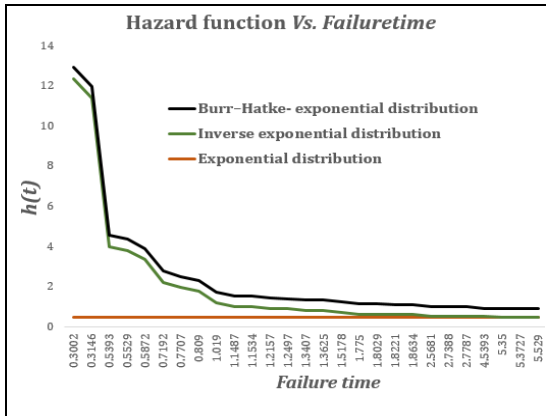


Fig. 2. Trend of hazard function

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.

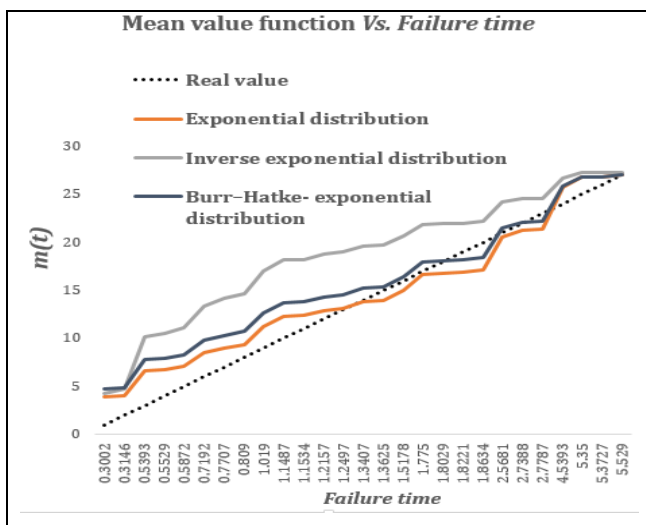


Fig. 3. Pattern of mean value function for each model

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 perceived in $(0, x_i]$ and $\hat{m}(x_i)$ estimated 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].

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

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.

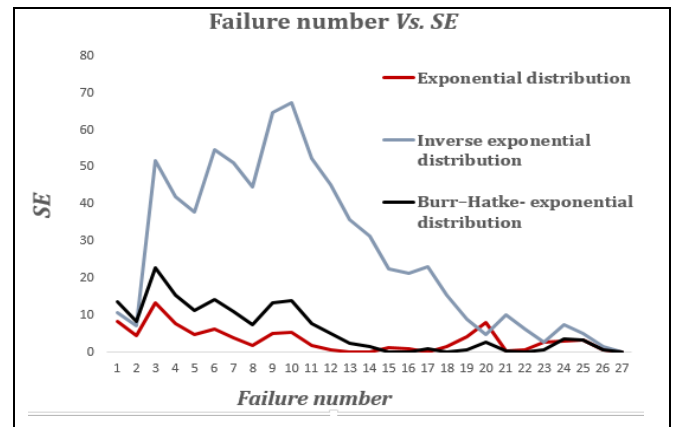


Fig. 4. Estimation of square error for each time

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].

$$\begin{aligned} \hat{R}(t | x_{27} = 5.529) &= e^{-\int_{5.529}^{5.529+t} \lambda(\eta) d\eta} \\ &= \exp[-\{m(t + 5.529) - m(5.529)\}] \end{aligned} \quad (29)$$

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.

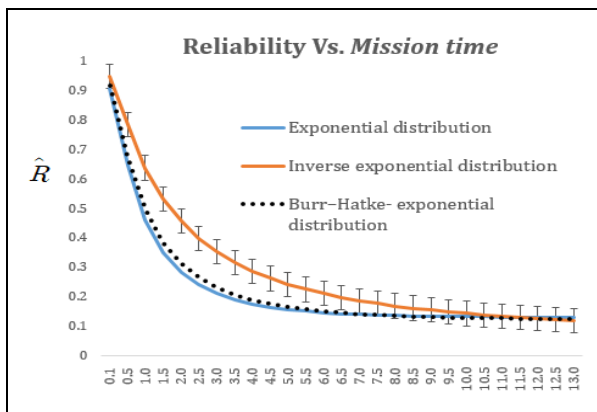


Fig. 5. Transition of reliability pattern

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, 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

- [1] 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*. 2017; 7(983): 1-16.
- [2] S. Yamada, S. Osaki, Software reliability growth modelling: models and applications. *IEEE Transactions on Software Engineering*. 1985; 11(12): 1431-1437.
- [3] X. Teng, H. Pham, A new methodology for predicting software reliability in the random field environments. *IEEE Transactions on Reliability*. 2006; 55(3): 458-468.
- [4] Goel AL, Okumoto K. Time dependent error - detection rate model for software reliability and other performance measure. *IEEE Trans Reliability*. 1979; R-28(3): 206–211.
- [5] Hee-Cheul Kim. The 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*. 2019; 12(1): 73-78.
- [6] S. S. Gokhale, K. S. Trivedi, A time/structure based software reliability model. *Annals of Software Engineering*. 1999; 8(85): 85-121.
- [7] Kersey, J. X. Weighted Inverse Weibull and Beta-Inverse Weibull Distribution. *Electronic Theses & Dissertations Paper 661*. 2010; 1-53.
- [8] Hee-Cheul Kim. A Comparative Study on the Cost of Software Development Model Based on Inverse Exponential Distribution. *International Journal of Applied Business and Economic Research*. 2017; 1-7.

- [9] Yadav, A. S. Altun, E, Yousof, H. M. Burr–Hatke Exponential Distribution: A Decreasing Failure Rate Model, *Statistical Inference and Applications. Annals of Data Science.* 2019; 1-20, 2019.
- [10] Hee-Cheul Kim. A Comparative Study on the Cost of Software Development Model Based on Burr–Hatke-Exponential Distribution. *International Journal of Engineering Research and Technology.* 2019; 12(11): 2036-2040.
- [11] Satya Prasad R, Rao KRH, Kantha RRL., Software reliability measuring using modified maximum likelihood estimation and SPC, *International Journal of Computer Applications.* 2011; 21(7):1–5.
- [12] Hee-Cheul Kim, A Comparative Study on Software Reliability Models with Shape Parameter of Type-2 Gumbel Life Distribution, *International Journal of Soft Computing,* 2017; 12(5-6):351-354.
- [13] Hee-Cheul Kim. A Comparative Study on the Finite Failure Software Reliability Model with Modified Lindley Type Lifetime Distribution. *International Journal of Engineering Research and Technology.* 2019; 12(6): 760-764.
- [14] Kim K-S, Kim H-C, The Performance Analysis of the Software Reliability NHPP Log-linear Model Depend on Viewpoint of the Learning Effects, *Indian Journal of Science and Technology.* 2016; 9(37):1-5.
- [15] Kim, H.C., The Property of Learning effect based on Delayed Software S-Shaped Reliability Model using Finite NHPP Software Cost Model. *Indian Journal of Science and Technology,* 2015; 8(34): 1-7