Confidence Intervals for the Failure Intensity and Number of Remaining Errors Functions Based on Non-Homogeneous Poisson Process Model

Lutfiah Ismail Al turk*

Statistics Department, Faculty of Sciences, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

*Corresponding author; Email: lturk@kau.edu.sa

Manuscript submitted September 29, 2018; accepted November 2, 2018.
doi: 10.17706/jsw.13.12.675-686

Abstract: The most important attribute of software quality is software reliability. Over the past years, many reliability models have been constructed for measuring and improving the growth of reliability. In reliability engineering, less attention has been paid to find confidence intervals around the estimated parameters and measures of reliability although its importance at producing precise prediction results. In this paper, we focus on the construction of confidence intervals based on a reliability model that b elongs to the family of Non-Homogeneous Poisson Process (NHPP) which is created using the Half-Logistic Distribution (HLD). In more detail, Asymptotic Confidence Intervals (ACIs) are derived using the Fisher information matrix for the constructed model's parameters as well as its intensity and number of remaining errors functions. Analysis of the DACS software reliability dataset to examine the accuracy of the point and interval estimation are provided.

Keywords: Non-homogeneous Poisson process, half-logistic distribution, intensity function, number of remaining errors, asymptotic confidence interval, Fisher information matrix.

1. Introduction

Software engineers predominantly aims to produce high quality software for real world applications. Error detection and correction process yields the software reliability growth which probabilistic models aim to describe and develop. These probabilistic models can be categorized into several classes for examples: models represent the failure rate of an equipment; models based on curve fitting; fault seeding models; reliability growth models; Markov structure models; models constructed by using Non-Homogenous Poisson Process (NHPP) etc. The NHPP model based on rich assumptions to describe the process of fault discovery, its simplicity, convenience, and compatibility has attracted many researchers; some among them are: (Goel and Okumoto [1]; Yamada and Osaki [2]; Musa et al. [3]; Hossain and Dahiya [4]; Pham et al. [5]; Stringfellow and Andrews [6]; Almering et al. [7]).

Half-Logistic Distribution (HLD) belongs to the logistic distribution family, it was initially created by Balakrishnan [8] through the use of the absolute transformation of the logistic distribution, its hazard rate increases monotonically for all the parameter values as seen in Fig. 1 and the behavior of its tail thickness is between the half-normal and half-Cauchy distributions hence it has a rather flexible functional form. This distribution is greatly considered by many researchers in modeling datasets of numerous areas, for examples: (Mbah and Tsokos [9]; Giles [10]; Saran and Pande [11]). Because of the importance of the HLD as a life testing model, constructing a NHPP model based on it may produce repairable system model which may has more ability to suit different projects. The probability density function (pdf) and cumulative distribution function (cdf) of the HLD are, respectively, defined as follows:

$$f(\mathrm{t_i}; \mathrm{a}, \xi) = \frac{2e^{-\frac{t_i}{\xi}}}{\xi\left(1+e^{-\frac{t_i}{\xi}}\right)^2} \quad (1)$$

$$\mathrm{F}(\mathrm{t_i}; \mathrm{a}, \xi) = \left(\frac{1-e^{-\frac{t_i}{\xi}}}{1+e^{-\frac{t_i}{\xi}}}\right) \quad (2)$$

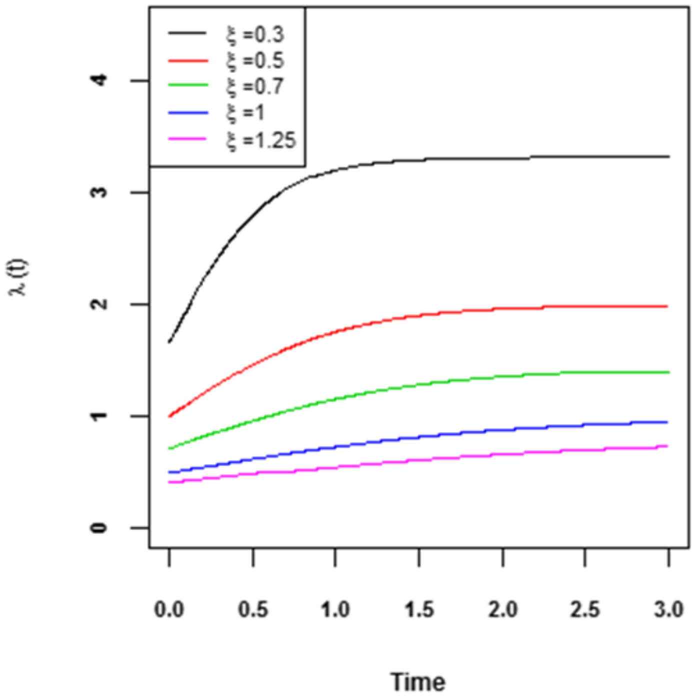

Fig. 1. Hazard rate function of the HLD for some indicated values of ξ.

In statistics, there are two types of parameter estimation comprises the use of failure data collected during software testing phase, namely point and interval estimation. A point estimation assesses a specific value for the actual parameter which is practically clear to be imprecise (see; Koch and Link [12]) while the interval estimation gives several values with identified probability of attaining the actual parameter, Neyman [13] was the first who proposed the theory of confidence interval estimation. In reliability engineering, the estimation of the models' parameters is regularly performed by point estimation method, hence the prediction of the measures of reliability of models needs to be investigated more by using the confidence interval approach so that we can precisely trust the predictive results and reasonably make a release decision of a software product.

This paper is arranged as follows. Section 2 gives a review of the finite failure NHPP models, followed by the construction of a NHPP model based on the HLD (NHPP HLD model), then the formulation of some of the constructed model's characteristics. In Section 3, the Asymptotic Confidence Intervals (ACIs) are derived for the suggested model's parameters, its intensity and number of remaining errors functions. The accuracy of these ACIs is assessed in Section 4 using the DACS failure data. Finally, the concluding remarks are given in Section 5.

2. Finite NHPP Model

The Finite failure NHPP class of modeling is a class of time domain models which describes the behavior of failure detection process by a NHPP. Let $N(t_i)$ represent the cumulative number of the errors detected by time t_i; $i = (1,2,\dots,n)$ and was identified to follow a Poisson distribution with parameter $m(t_i;\Lambda)$, that is:

$$P(N(t_i) = n) = \frac{[m(t_i;\Lambda)]^n}{n!} e^{-m(t_i;\Lambda)} \quad ,\text{where } n = 0,1,\dots \tag{3}$$

The mean value function of the finite failure NHPP models, $m(t_i;\Lambda) = E[N(t_i)]$, and the failure intensity function $\lambda(t_i;\Lambda)$ can be derived, respectively, as follows:

$$m(t_i;\Lambda) = \int_0^{t_i} \lambda(s;\Lambda)ds, \tag{4}$$

$$\lambda(t_i;\Lambda) = \frac{dm(t_i;\Lambda)}{dt}. \tag{5}$$

Then, $m(t_i;\Lambda)$ can also be written as: (see; Lyu [14]);

$$m(t_i;\Lambda) = a\,F(t_i;\Lambda) \quad ,\text{where} \tag{6}$$

a: symbolize the expected number of errors that would be determined by a NHPP model.
Λ: Unknown parameters of a NHPP model.
$F(t_i;\Lambda)$: is a cdf of the time to failure. Moreover, the expected number of errors remaining in the software at time t, can be given by:

$$n(t_i;\Lambda) = a - m(t_i;\Lambda). \tag{7}$$

2.1. Model Formulation and Characteristics

Using Eqs. (2), (5), and (6) the formulas for the mean value and failure intensity functions of the Non-Homogeneous Poisson Process Half Logistic Distribution (NHPP HLD) model can be, respectively, expressed as follows:

$$m(t_i;a,\xi) = a\left(\frac{1-e^{-\frac{t_i}{\xi}}}{1+e^{-\frac{t_i}{\xi}}}\right), \tag{8}$$

$$\lambda(t_i;a,\xi) = \frac{2ae^{-\frac{t_i}{\xi}}}{\xi\left(1+e^{-\frac{t_i}{\xi}}\right)^2}, \tag{9}$$

where

a: the number of initially found errors in the software,
ξ: is the scale parameter.
$\frac{m(t_i;a,\xi)}{a}$: is the cdf of HLD.
According to Eqs. (7) and (8) the number of remaining errors of the NHPP HLD model can be written as follows:

$$n(t_i;a,\xi) = \frac{2ae^{-\frac{t}{\xi}}}{1+e^{-\frac{t}{\xi}}}. \tag{10}$$

The intensity and number of remaining errors functions are represented graphically for different values of parameters in Fig. 2. As shown, the two characteristic functions depend on the number of starting errors in the software a and the shape parameter ξ. The intensity function starts declining from a higher value as the parameter a increases, and in all cases the curve stabilizes at a value near to zero as time increases, the starting value of the curve decreases as ξ increases. Whereas, all the curves of the number of remaining errors function have the same starting point, which represents the initial number of errors in a software, the sharpness of the curves decreases as ξ deceases.

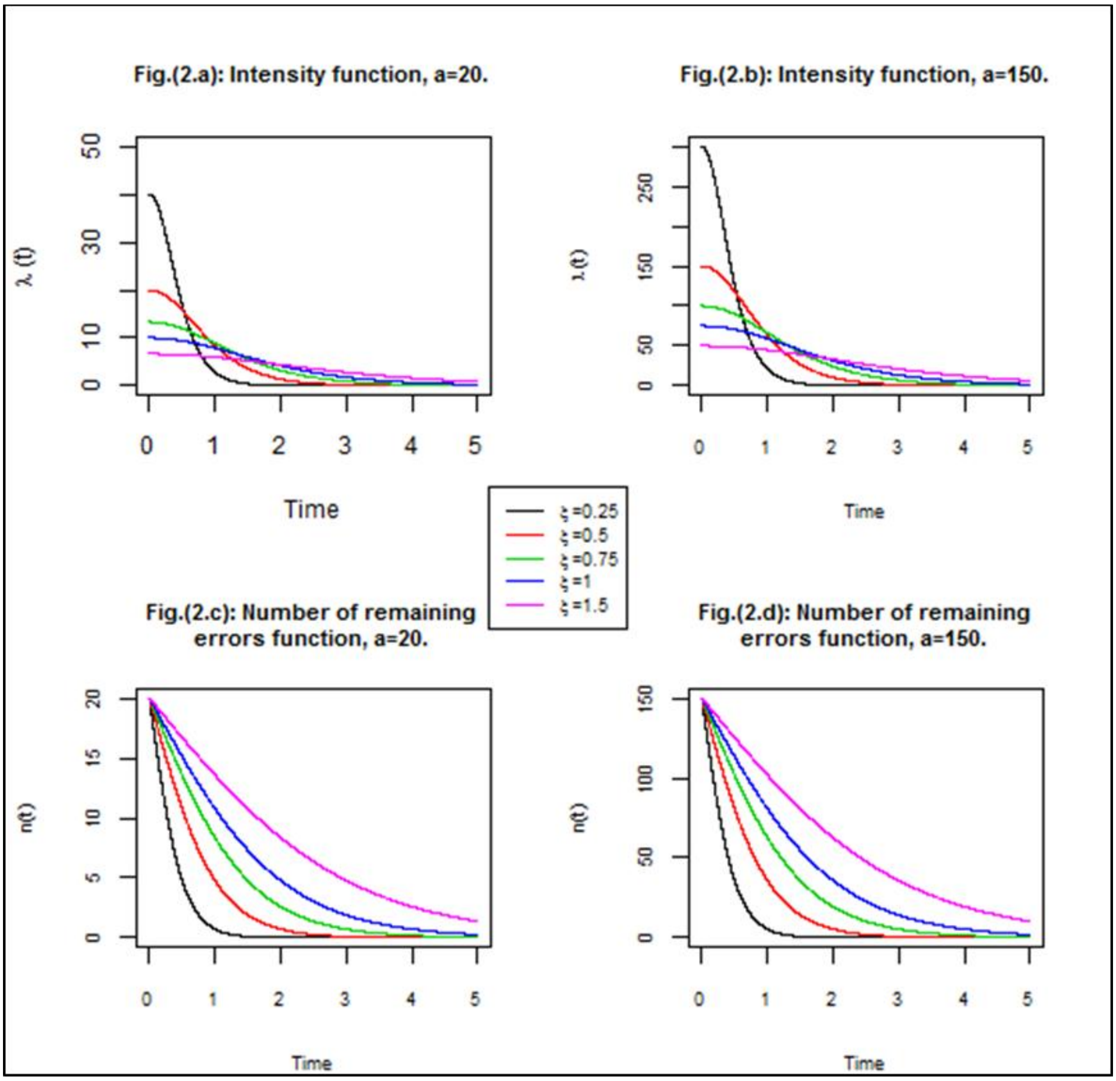

Fig. 2. Some characteristics of the NHPP HLD model for some indicated values of ξ.

3. Parameters Estimation of the NHPP Model

Usually by analyzing the failure data that observed during the testing phase a point estimate is obtained and used to assess software reliability. Two obstacles usually faced when using point estimation: different estimates to model's parameters are obtained when using different estimation methods, and the unavailability of large sample sizes that needed to give precise estimates. Thus, to solve these obstacles interval estimation approach can be used to assess software reliability. As a first step, the Fisher information matrix is calculated to obtain the confidence bounds for parameters Λ of a NHPP model. The Fisher information is an approach of quantifying the amount of information that observable software failure times S_i $(i = 1, 2, \dots, n)$ contains about the unknown parameters Λ of a NHPP model that models S_i. This approach uses the matrix of negative second partial derivatives of the natural logarithm of likelihood function. The inverse of the Fisher information matrix gives the asymptotic variance and covariance of the

Maximum Likelihood (ML) estimates of the parameters Λ of a NHPP model. Thus, the $100(1-\alpha)\%$ asymptotic confidence interval for the parameters Λ of a NHPP model are then obtained by:

$$\left(\widehat{\Lambda} - Z_{\alpha/2}\left[var(\widehat{\Lambda})\right]^{1l2}, \widehat{\Lambda} + Z_{\alpha/2}[var(\widehat{\Lambda})]^{1l2}\right), \tag{11}$$

where $\widehat{\Lambda}$ is the ML estimator of Λ. $var(\widehat{\Lambda})$ is the estimated variance of $\widehat{\Lambda}$, and $Z_{\alpha/2}$ is the percentile of standard normal distribution with right-tail probability $\alpha/2$ (Zhang et al. [15]).

Now, let T_i denote a random variable indicating the time between (i-1)st and ith failure (i = 1, 2, ..., n), Then, $S_i = \sum_{i=1}^{i} T_i$ is a random variable indicating the ith failure occurrence time where $T_i = S_i - S_{i-1}$ ($i = 1, 2, \dots, n$; $S_0 = 0$) the joint density function of the unknown parameters of a NHPP model with $m(s_n;\Lambda)$ is given by:

$$L(S|\Lambda) = e^{-m(s_n;\Lambda)} \prod_{i=1}^{n} \lambda(s_i;\Lambda). \tag{12}$$

Taking the natural logarithm of Eq. (12) yields:

$$\ln L(S|\Lambda) = -\,m(s_n;\Lambda) + (\ln \prod_{i=1}^{n} \lambda(s_i;\Lambda))$$

$$= -m(s_n;\Lambda) + (\sum_{i=1}^{n} \ln\lambda(s_i;\Lambda)). \tag{13}$$

By taking the partial derivatives of Eq. (13) with respect to each of the NHPP model parameter then setting the resulted equations equal to zero and solving the obtained equations simultaneously, the parameters of the NHPP model can be estimated by the ML estimation method.

3.1. Parameters Estimation of the NHPP HLD Model

This section covers the use of the ML estimation method for estimating the parameters of the NHPP HLD model. Asymptotic confidence intervals for the model's parameters (a and ξ), intensity function ($\lambda(t_i;a,\xi)$) and number of remaining errors function ($n(t_i;a,\xi)$) will be obtained.

3.1.1. Asymptotic Confidence Intervals for the Parameters a and ξ

To calculate the estimates of the parameters a and ξ, the ln likelihood function is obtained by substituting Eqs. (8) and (9) in Eq. (13) as follows:

$$\ln L(t_i;a,\xi|\underline{S}) = -a\left(\frac{1-e^{-\frac{s_n}{\xi}}}{1+e^{-\frac{s_n}{\xi}}}\right) + \sum_{i=1}^{n} \ln\left(\frac{2ae^{-\frac{s_i}{\xi}}}{\xi\left(1+e^{-\frac{s_i}{\xi}}\right)^2}\right)$$

$$= -a\left(\frac{1-e^{-\frac{s_n}{\xi}}}{1+e^{-\frac{s_n}{\xi}}}\right) + n\ln a + n\ln 2 - \frac{\sum_{i=1}^{n} s_i}{\xi} - n\ln\xi - 2\sum_{i=1}^{n}\ln\left(1+e^{-\frac{s_i}{\xi}}\right). \tag{14}$$

Then, the derivatives of Eq. (14) with respect to the parameters a and ξ are found as follows:

$$\begin{cases} \dfrac{\partial \ln L(t_i;a,\xi|\underline{S})}{\partial a} = -\left(\dfrac{1-e^{-\frac{s_n}{\xi}}}{1+e^{-\frac{s_n}{\xi}}}\right) + \dfrac{n}{a} \\ \dfrac{\partial \ln L(t_i;a,\xi|\underline{S})}{\partial \xi} = \dfrac{2as_n e^{-\frac{s_n}{\xi}}}{\xi^2\left(1+e^{-\frac{s_n}{\xi}}\right)^2} + \dfrac{\sum_{i=1}^{n} s_i}{\xi^2} - \dfrac{n}{\xi} + \dfrac{2}{\xi^2}\sum_{i=1}^{n}\dfrac{s_i e^{-\frac{s_i}{\xi}}}{1+e^{-\frac{s_i}{\xi}}} \end{cases} \tag{15}$$

Setting these derivatives to zero and solving the obtained equations for the parameters a and ξ, the following equation are found:

$$\begin{cases} \hat{a} = n\left(\frac{1+e^{-\frac{s_n}{\hat{\xi}}}}{1-e^{-\frac{s_n}{\hat{\xi}}}}\right) \\ \frac{2ns_n e^{-\frac{s_n}{\hat{\xi}}}}{\hat{\xi}^2\left(1+e^{-\frac{s_n}{\hat{\xi}}}\right)(1-e\text{^}(-s\_n/\xi))} + \frac{\sum_{i=1}^{n} s_i}{\hat{\xi}^2} - \frac{n}{\hat{\xi}} + \frac{2}{\hat{\xi}^2}\sum_{i=1}^{n}\frac{s_i e^{-\frac{s_i}{\hat{\xi}}}}{1+e^{-\frac{s_i}{\hat{\xi}}}} = 0 \end{cases} \tag{16}$$

The second expression of Eq. (16) is non-linear and need to be solved numerically to obtain the estimate of the parameter ξ, then the estimate of the parameter a can be obtained by substituting $\hat{\xi}$ in the first expression of Eq. (16). Thus, to find the confidence bounds for the parameters a and ξ of the NHPP HLD model, the Fisher information matrix is calculated as follows:

$$\mathrm{I}(\Lambda) = \begin{bmatrix} -\frac{\partial^2 \ln L(\mathrm{t_i};\mathrm{a},\xi|\underline{\mathrm{S}})}{\partial a^2} & -\frac{\partial^2 \ln L(\mathrm{t_i};\mathrm{a},\xi|\underline{\mathrm{S}})}{\partial a \partial \xi} \\ -\frac{\partial^2 \ln L(\mathrm{t_i};\mathrm{a},\xi|\underline{\mathrm{S}})}{\partial \xi \partial a} & -\frac{\partial^2 \ln L(\mathrm{t_i};\mathrm{a},\xi|\underline{\mathrm{S}})}{\partial \xi^2} \end{bmatrix} \tag{17}$$

After that, the asymptotic variance and covariance matrix is found as follows:

$$\mathrm{I}^{-1}(\Lambda) = \begin{bmatrix} Var(a) & Cov(\mathrm{a},\xi) \\ Cov(\mathrm{a},\xi) & Var(\xi) \end{bmatrix} = \frac{1}{\left(\frac{\partial^2 \ln L(t_i;a,\xi|\underline{S})}{\partial \xi^2}\right)\left(\frac{\partial^2 \ln L(t_i;a,\xi|\underline{S})}{\partial a^2}\right) - \left(\frac{\partial^2 \ln L(t_i;a,\xi|\underline{S})}{\partial \xi \partial a}\right)^2} \times \begin{bmatrix} -\frac{\partial^2 \ln L(t_i;a,\xi|\underline{S})}{\partial \xi^2} & \frac{\partial^2 \ln L(t_i;a,\xi|\underline{S})}{\partial a \partial \xi} \\ \frac{\partial^2 \ln L(t_i;a,\xi|\underline{S})}{\partial \xi \partial a} & -\frac{\partial^2 \ln L(t_i;a,\xi|\underline{S})}{\partial a^2} \end{bmatrix}, \tag{18}$$

where,

$$\begin{cases} \frac{\partial^2 \ln L(\mathrm{t_i};\mathrm{a},\xi|\underline{\mathrm{S}})}{\partial a^2} = -\frac{n}{a^2} \\ \frac{\partial^2 \ln L(\mathrm{t_i};\mathrm{a},\xi|\underline{\mathrm{S}})}{\partial a \partial \xi} = \frac{\partial^2 \ln L(\mathrm{t_i};\mathrm{a},\xi|\underline{\mathrm{S}})}{\partial \xi \partial a} = \frac{-2s_n e^{-\frac{s_n}{\xi}}}{\xi^2\left(1+e^{-\frac{s_n}{\xi}}\right)^2} \\ \frac{\partial^2 \ln L(\mathrm{t_i};\mathrm{a},\xi|\underline{\mathrm{S}})}{\partial \xi^2} = \frac{2as_n e^{-\frac{s_n}{\xi}}}{\xi^4\left(1+e^{-\frac{s_n}{\xi}}\right)^3}\left(s_n\left(1-e^{-\frac{s_n}{\xi}}\right) - 2\xi\left(1+e^{-\frac{s_n}{\xi}}\right)\right) \end{cases}. \tag{19}$$

So, the 100(1− α)% asymptotic confidence interval for the parameters a and ξ of the NHPP HLD model are given, respectively, by:

$$\left(\hat{\mathrm{a}} \pm Z_{\alpha/2}[Var(\hat{\mathrm{a}})]^{1/2}\right), \tag{20}$$

$$\left(\hat{\xi} \pm Z_{\alpha/2}[Var(\hat{\xi})]^{1/2}\right), \tag{21}$$

where, $Z_{\alpha/2}$ is the percentile of standard normal distribution with right-tail probability $\alpha/2$, $Var(\hat{\mathrm{a}})$ and $Var(\hat{\xi})$ are, respectively, the diagonal elements of the asymptotic variance and covariance matrix given by Eq. (18).

3.1.2. Asymptotic Confidence Intervals for the Functions $\lambda(t_i; a, \xi)$ and $n(t_i; a, \xi)$

According to the invariance property of the ML estimators, the estimate of the intensity function of the NHPP HLD model is obtained by:

$$\hat{\lambda}(t_i; \mathrm{a}, \xi) = \frac{2\hat{a}e^{-\frac{t_i}{\hat{\xi}}}}{\hat{\xi}\left(1+e^{-\frac{t_i}{\hat{\xi}}}\right)^2}, \tag{22}$$

while its variance is defined as:

$$V(\hat{\lambda}) = \left(\frac{\partial\lambda}{\partial a}\right)^2\Big|_{a=\hat{a}} V(\hat{a}) + \left(\frac{\partial\lambda}{\partial\xi}\right)^2\Big|_{\xi=\hat{\xi}} V(\hat{\xi}) + 2\left(\frac{\partial\lambda}{\partial a}\right)\left(\frac{\partial\lambda}{\partial\xi}\right)\Big|_{a=\hat{a},\xi=\hat{\xi}} \mathrm{Cov}\,(\hat{a},\hat{\xi})$$

$$= \frac{4e^{-\frac{2s_i}{\hat{\xi}}}}{\hat{\xi}^2\left(1+e^{-\frac{s_i}{\hat{\xi}}}\right)^4} V(\hat{a}) + \frac{4\hat{a}^2 e^{-\frac{2t_i}{\hat{\xi}}}\left(t_i-\hat{\xi}-(t_i+\hat{\xi})e^{-\frac{s_i}{\hat{\xi}}}\right)^2}{\hat{\xi}^6\left(1+e^{-\frac{t_i}{\hat{\xi}}}\right)^6} V(\hat{\xi}) + \frac{16\hat{a}e^{-\frac{2t_i}{\hat{\xi}}}\left(t_i-1-(t_i+\hat{\xi})e^{-\frac{t_i}{\hat{\xi}}}\right)}{\hat{\xi}^4\left(1+e^{-\frac{t_i}{\hat{\xi}}}\right)^5} Cov(\hat{a},\hat{\xi}) +$$

$$\frac{16\hat{a}e^{-\frac{2t_i}{\hat{\xi}}}\left(t_i-1-(t_i+\hat{\xi})e^{-\frac{t_i}{\hat{\xi}}}\right)}{\hat{\xi}^4\left(1+e^{-\frac{t_i}{\hat{\xi}}}\right)^5} Cov(\hat{a},\hat{\xi}) \tag{23}$$

Correspondingly, the estimate of the number of remaining errors function of the NHPP HLD model is obtained by:

$$\hat{n}(t_i; \mathrm{a}, \xi) = \frac{2\hat{a}e^{-\frac{t_i}{\hat{\xi}}}}{1+e^{-\frac{t_i}{\hat{\xi}}}}, \tag{24}$$

and its variance is defined as:

$$V(\hat{n}) = \left(\frac{\partial n}{\partial a}\right)^2\Big|_{a=\hat{a}} V(\hat{a}) + \left(\frac{\partial n}{\partial\xi}\right)^2\Big|_{\xi=\hat{\xi}} V(\hat{\xi}) + 2\left(\frac{\partial n}{\partial a}\right)\left(\frac{\partial n}{\partial\xi}\right)\Big|_{a=\hat{a},\xi=\hat{\xi}} \mathrm{Cov}\,(\hat{a},\hat{\xi})$$

$$= \frac{4e^{\frac{-2t_i}{\hat{\xi}}}}{\left(1+e^{-\frac{t_i}{\hat{\xi}}}\right)^2} V(\hat{a}) + \frac{4\hat{a}^2 t_i^2 e^{-\frac{2t_i}{\hat{\xi}}}}{\hat{\xi}^4\left(1+e^{-\frac{t_i}{\hat{\xi}}}\right)^4} V(\hat{\xi}) + \frac{8\hat{a}^2 t_i e^{-\frac{2t_i}{\hat{\xi}}}}{\hat{\xi}^2\left(1+e^{-\frac{t_i}{\hat{\xi}}}\right)^3} \mathrm{Cov}\,(\hat{a},\hat{\xi}). \tag{25}$$

We employ the central limit theorem of $\left(\lambda(\mathrm{t_i}; \mathrm{a}, \xi), n(\mathrm{t_i}; \mathrm{a}, \xi)\right)$ and gets the asymptotic 100(1− α)% confidence bounds for the actual values as follows: Lower and upper confidence bounds for the intensity function $\lambda(\mathrm{t_i}; \mathrm{a}, \xi)$:

$$\left(\hat{\lambda} \pm Z_{\alpha/2}[Var(\hat{\lambda})]^{1/2}\right), \tag{26}$$

where, $Z_{\alpha/2}$ is the percentile of standard normal distribution with right-tail probability $\alpha/2$, $\hat{\lambda}$ is obtained from Eq.(22), and $Var(\hat{\lambda})$ is defined by Eq.(23). Whereas lower and upper confidence bounds for the number of remaining error $n(\mathrm{t_i}; \mathrm{a}, \xi)$ is:

$$\left(\hat{n} \pm Z_{\alpha/2}[Var(\hat{n})]^{1/2}\right), \tag{27}$$

where, $Z_{\alpha/2}$ is the percentile of standard normal distribution with right-tail probability $\alpha/2$, $\hat{n}$ is obtained from Eq.(24), and $Var(\hat{n})$ is defined by Eq.(25).

4. Illustrative Application

An application is conducted here to study the accuracy of point and interval estimation of the NHPP HLD model. Time between-failure-data from Data & Analysis Center for Software (DACS) (Musa et al. [3]) is used to do the calculations in this illustrative application. The number of failures found in the testing phase is 136. The data set is displayed in Table 1 and represented graphically in Fig. 3.

Table 1. DACS Software Reliability Dataset.

Failure Number	Failure Interval Length (in CPU seconds)	Failure Number	Failure Interval Length (in CPU seconds)	Failure Number	Failure Interval Length (in CPU seconds)	Failure Number	Failure Interval Length (in CPU seconds)
1	3	35	227	69	529	103	108
2	30	36	65	70	379	104	0
3	113	37	176	71	44	105	3110
4	81	38	58	72	129	106	1247
5	115	39	457	73	810	107	943
6	9	40	300	74	290	108	700
7	2	41	97	75	300	109	875
8	91	42	263	76	529	110	245
9	112	43	452	77	281	111	729
10	15	44	255	78	160	112	1897
11	138	45	197	79	828	113	447
12	50	46	193	80	1011	114	386
13	77	47	6	81	445	115	446
14	24	48	79	82	296	116	122
15	108	49	816	83	1755	117	990
16	88	50	1351	84	1064	118	948
17	670	51	148	85	1783	119	1082
18	120	52	21	86	860	120	22
19	26	53	233	87	983	121	75
20	114	54	134	88	707	122	482
21	325	55	357	89	33	123	5509
22	55	56	193	90	868	124	100
23	242	57	236	91	724	125	10
24	68	58	31	92	2323	126	1071
25	422	59	369	93	2930	127	371
26	180	60	748	94	1461	128	790
27	10	61	0	95	843	129	6150
28	1146	62	232	96	12	130	3321
29	600	63	330	97	261	131	1045
30	15	64	365	98	1800	132	648
31	36	65	1222	99	865	133	5485
32	4	66	543	100	1435	134	1160

33	0	67	10	101	30	135	1864
34	8	68	16	102	143	136	4116

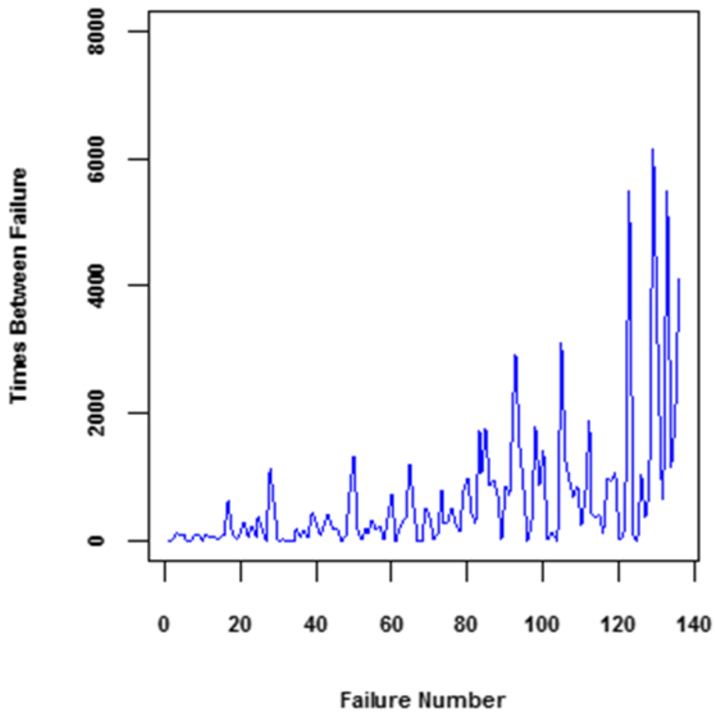

Fig. 3. DACS software reliability dataset: 136 failures.

The MLE method discussed in Section 3 is used to estimate the proposed model's parameters and the results for the last four failure indices (133, 134, 135, 136) are listed in Table 2, the results include the ML estimates of the number of initial errors in the software (a), the ML estimates of the shape parameter(ξ), and 95% ACIs for these parameters along with their observed lengths.

Table 2. Estimated Parameter Values for the NHPP HLD Model and 95% Confidence Intervals.

Failure Number	Failure Time (Hours) t	Prediction	ACI Lower limit $\hat{a}$ ACI Upper limit	ACI Lower limit $\hat{\xi}$ ACI Upper limit	Observed Interval Length $\hat{a}$	Observed Interval Length $\hat{\xi}$
Using 133 failures	5485	133.000002	135.1894 137.2356 139.2817	19281.39 19621.36 19961.34	4.092276	679.9488
Using 134 failures	1160	134.000006	136.3641 138.408 140.4518	19748.69 20054.4 20360.12	4.087709	611.4321
Using 135 failures	1864	134.999988	137.3169 139.3548 141.3928	20141.76 20411.13 20680.5	4.075910	538.7386
Using 136 failures	4116	135.999997	137.661 139.6835 141.7059	20299.74 20550.12 20800.51	4.044821	500.7703

Table 3 contains the predicted and 95% ACIs for the intensity and number of remaining errors functions and the observed length of their ACIs. The ACIs are compared via their lengths which are calculated respectively as follows: $2Z_{\alpha/2}[Var(\hat{a})]^{1/2}$, $2Z_{\alpha/2}[Var(\hat{\xi}))]^{1/2}$, $2Z_{\alpha/2}[Var(\hat{\lambda})]^{1/2}$, and $2Z_{\alpha/2}[Var(\hat{n})]^{1/2}$. Regarding the lengths of the ACIs presented in Table 2 and Table 3, it can be seen that as the number of detected failures increases narrower interval for the parameters' estimates, failure intensity prediction and number of remaining errors prediction are obtained.

Table 3. 95% Confidence Intervals for the Intensity and Number of Remaining Errors Functions Based on the NHPP HLD Model.

Error number	ACI Lower limit $\hat{\lambda}$ ACI Upper limit	ACI Lower limit $\hat{n}$ ACI Upper limit	Observed Interval Length $(\hat{\lambda})$	Observed Interval Length $(\hat{n})$
Using 133 failures	0.000200 0.000213 0.000225	3.910245 4.235567 4.560888	2.40886e-05	0.6506429
Using 134 failures	0.000206 0.000216 0.000227	4.110456 4.407989 4.705522	2.15619e-05	0.595066
Using 135 failures	0.000201 0.000210 0.000219	4.097494 4.354854 4.612213	1.84783e-05	0.5147191
Using 136 failures	0.000169 0.000177 0.000184	3.475239 3.683461 3.891684	1.51731e-05	0.4164452

Fig. 4 shows the difference between the actual and predicted values of the NHPP HLD model based on 136 observed failures, while Fig. 5 shows the relative errors for the prediction obtained by the NHPP HLD model, according to these figures we can see that the NHPP HLD model give a good representation to the DACS data. The 95% ACIs for the intensity and number of remaining errors functions are represented graphically in Fig. 6, these graphical representation shows improved prediction are obtained with the progress of the testing time.

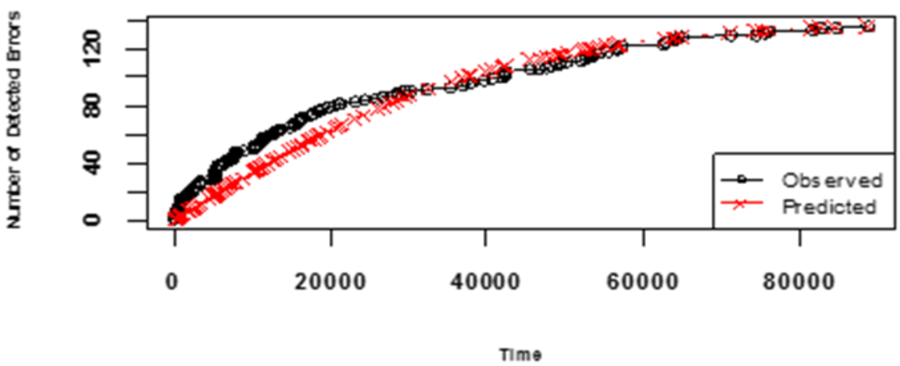

Fig. 4. The actual date and prediction results of the NHPP HLD model.

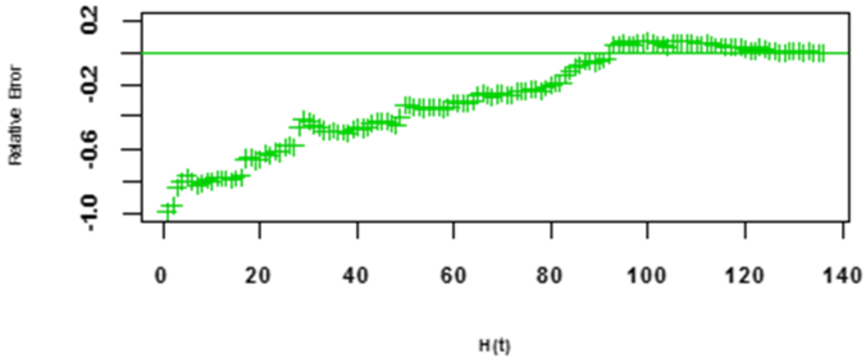

Fig. 5. Predictive relative error curve of the NHPP HLD model.

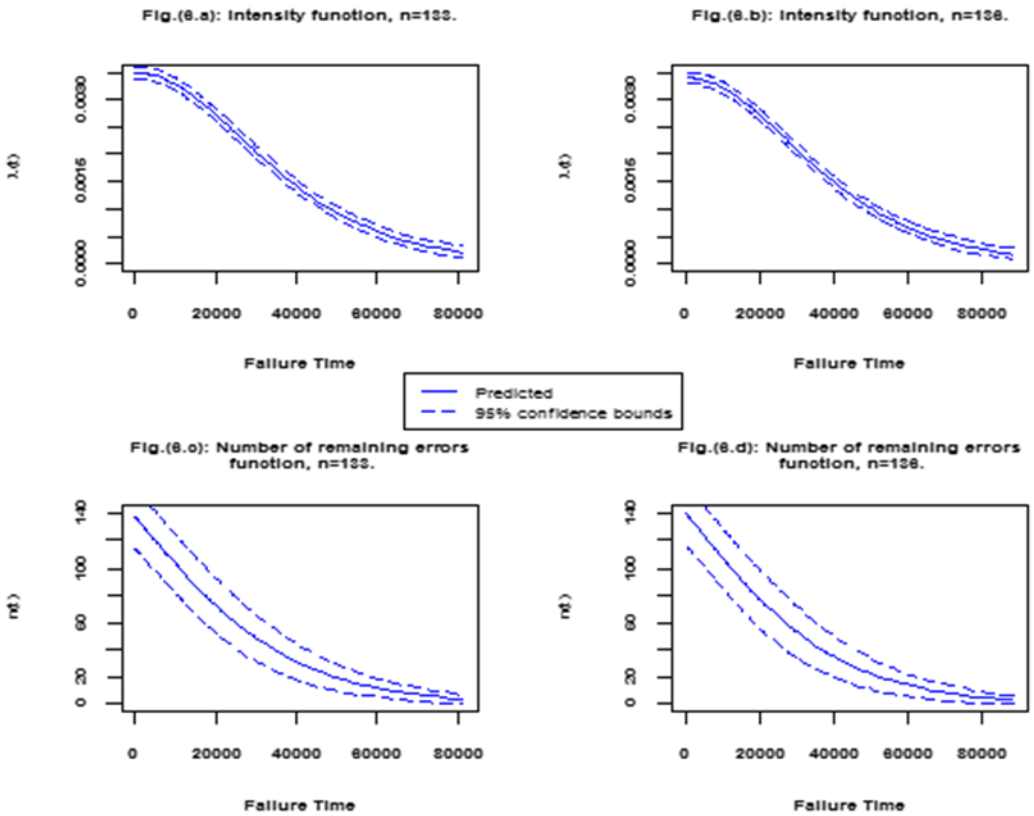

Fig. 6. Confidence intervals of the NHPP HLD model with a confidence coefficient of 95%.

5. Conclusion

In this paper, point and interval estimation approaches of the NHPP HLD model have been discussed. Specifically, confidence limits for the NHPP HLD model's parameters, intensity function and number of remaining errors function have been constructed. We have analyzed the point and interval estimation approaches on the DACS dataset, the application results show that the NHPP HLD model provide accurate prediction results and demonstrate the usefulness of interval estimation approach in practice. The approach can be extended to many other reliability models.

References

[1] Goel, A.L. and Okumoto, K. (1979). "Time Dependent Error Detection Rate Model for Software Reliability and Other Performance Measure", IEEE Trans. Reliab. R-28(3), pp. 206-211.

[2] Yamada, S. and Osaki, S. (1985). Software Reliability Growth Modeling: Models and Applications. IEEE

Trans. on Software Engineering, SE-11(12):1431–1437.
[3] Musa, J. D., Iannino, A. and Okumoto, K. (1987). Software Reliability: Measurement Prediction Application. MC Graw Hill, New York.
[4] Hossain, S.A. and Dahiya, R.C. (1993). Estimating the parameters of a nonhomogeneous Poisson-process model for software reliability. IEEE Trans. Reliab. 42(4), 604–612.
[5] Pham, H., Nordmann, L. and Zhang, X.M. (1999). A general imperfect software-debugging model with S-shaped fault-detection rate. IEEE Transactions on Reliability 48 (2), 169–175.
[6] Stringfellow, C. and Andrews, A.A. (2002). An Empirical Method for Selecting Software Reliability Growth Models, Empirical Software Engineering, 7, 319-343.
[7] Almering, V., von Genuchten, M., Cloudt, G. and Sonnemans, P.J.M. (2007). Using Software Reliability Growth Models in Practice, IEEE Software, 24, 82-88.
[8] Balakrishnan, N. (1985). Order statistics from the half logistic distribution, Journal of statistical computation and simulation, 20, 287-309.
[9] Mbah, A. K. and Tsokos, C. P. (2008). Record values from half logistics and inverse Weibull probability distribution functions. Neural, Parallel and Scientific Computations 16: 73 - 92.
[10] Giles, D.E. (2012). Bias reduction for the maximum likelihood estimators of the parameters in the half-logistic distribution. Communications in Statistics-Theory and Methods 41(2): 212-222.
[11] Saran, J. and Pande, V. (2012). Recurrence relation for moments of progressively type-II right censored order statistics from half logistic distribution. J. Statistical Theory and Applications, 11, 87 - 96.
[12] Koch, G.S. JR. and Link, R.F. (2002). Statistical Analysis of Geological Data, New York: Dover Publication, Inc.
[13] Neyman, J. (1937). "Outline of a theory of statistical estimation based on the classical theory of probability," Philosophical Transactions of the Royal Society Series A, vol. 236, pp. 333–380.
[14] Lyu, M. R. (2002). Software Reliability Theory. Encyclopedia of Software Engineering, Wiley, 1611-1630.
[15] Zhang, X., Teng, X. and Pham, H. (2003)."Considering Fault Removal Efficiency in Software Reliability Assessment", IEEE Trans. on Systems, Man, & Cybernetics – Part A, vol. 33, no.1, pp. 114-120.

Lutfiah Ismail Al turk is currently working as associate professor of mathematical statistics in Statistics Department at Faculty of Sciences, King AbdulAziz University, Jeddah, Kingdom of Saudi Arabia. Lutfiah Ismail Al turk obtained her B.Sc degree in statistics and computer science from Faculty of Sciences, King AbdulAziz University in 1993 and M.Sc (mathematical statistics) degree from Statistics Department, Faculty of Sciences, King AbdulAziz University in 1999. She received her Ph.D in mathematical statistics from university of Surrey, UK in 2007. Her current research interests include software reliability modeling and statistical machine learning.