

Convergence of MLE to MVUE of reliability for exponential class software reliability models

B. Roopashri Tantri
Department of ISE, NCET
Bangalore, India
e-mail: roopa_tan@hotmail.com

Murulidhar N. N.
Department of MACS, NITK
Surathkal, India
e-mail: murulidhrann@gmail.com

Abstract— Software quality has become a major concern of all software manufacturers. One such measure of software quality is the reliability, which is the probability of failure-free operation of a software in a specified environment for a specified time. If T denotes the time to failure of any software, then, the reliability of this software, denoted by $R(t)$, is given by $R(t)=P(T>t)$. The reliability of any software can be estimated using various methods of estimation. The simplest among these methods, is the method of maximum likelihood estimation. Even though it satisfies most of the desirable properties of a good estimator, it is still not as efficient as the minimum variance unbiased estimator. In this paper, the minimum variance unbiased estimator of $R(t)$ for exponential class software reliability models, is obtained using a procedure called blackwellization. The error in the two estimators is obtained for a model belonging to the exponential class, viz, the Jelinski - Moranda model. It is found that as the failure time increases, the difference decreases exponentially and finally both approach zero for a very large failure time.

Keywords- Exponential class models, Maximum likelihood estimator, Minimum variance unbiased estimator, Software reliability, Software reliability models, Variance.

I. INTRODUCTION

The increasing pace of change in computing technology means that information systems become economically obsolete more rapidly. Software can become obsolete either as a result of new hardware or software technology. The most important software product characteristics are quality, cost and schedule. Quantitative measures exist for the latter two characteristics, but the quantification of quality has become more difficult. It is most important, however, because the absence of a concrete measure for software quality generally means that quality will suffer when it competes for attention against cost and schedule. One of the important measures of software quality is the reliability. Software reliability concerns itself with how well the software functions to meet the requirements of the customer. It is the probability that the software will work without failure for a specified period of time[1]. It relates to operation rather than design of the program. In general, reliability is the probability of failure free operation of software in a specified environment for a specified period of time[2]. The failure times are random in nature. If T denotes the time to failure of software, then, its reliability is given by $R(t) = P(T > t)$. For any given software, this reliability can be estimated using various methods of estimation. The most popular method is the method of maximum likelihood estimation (MLE). In case of exponential class models, where failure time is assumed to be exponential, the MLE of $R(t)$ is easy to obtain and hence is widely used[3]. But, usually such an MLE is biased and is not efficient. To obtain an unbiased and more efficient estimator of $R(t)$, the method of minimum variance unbiased estimation (MVUE) is used.

Some terminologies used:

Software reliability[1]: It is the probability of failure-free operation of a software in a specified environment for a specified time.

Software reliability models[1] and [4]: These describe the behavior of failure with time, by expressing failures as random processes in either times of failure or the number of failures, at fixed times.

Exponential class software reliability models[2] and [4]: This group consists of all finite failure models which have exponential failure time distribution.

Estimator: A function of the sample observations used to estimate the value of the parameter of any distribution is said to be an estimator.

Any estimator is expected to possess certain statistical properties. An estimator that satisfies most of such desirable properties is considered as the best estimator.

The statistical properties of good estimators are given below:

Statistical properties of good estimators [3]: A good estimator should be (i) Consistent (ii) Unbiased (iii) Sufficient (iv). Efficient.

(i) Consistency: An estimator T based on a sample of size n from a distribution with parameter θ is said to be consistent for θ , if T converges in probability to θ . i.e., if $P(|T-\theta| < \epsilon) \rightarrow 1$ as $n \rightarrow \infty$, for every $\epsilon > 0$.

(ii) Unbiasedness: An estimator T based on a sample of size n from a distribution with parameter θ , is said to be unbiased for θ , if $E(T)=\theta$.

(iii) Sufficiency: An estimator T is said to be a sufficient estimator for the parameter θ , if it contains all the information in the sample regarding that parameter. Or, in other words, an estimator T is sufficient for the parameter θ if the conditional distribution of the sample given T is independent of θ . Using the factorization theorem, a statistic T is sufficient for the parameter θ , if the likelihood function L can be expressed as $L=g(T, \theta).h(x)$, where in g , θ depends on x only through T and $h(x)$ is independent of θ .

(iv) Efficiency: If in a class of estimators, there exists one whose variance is less than that of any such estimator, then it is called an efficient estimator. Thus, if T_1 and T_2 are two estimators, then T_1 is more efficient than T_2 , if $V(T_1) < V(T_2)$.

Complete estimator: Let T be an estimator and let $h(T)$ be any function of T . Then, T is said to be the complete estimator, if $E[h(T)=0]$ implies $h(T)=0$.

Complete sufficient estimator: A sufficient estimator which is also complete is called the complete sufficient estimator.

II. METHODS OF ESTIMATION

In order to estimate the reliability, various methods of estimation are available. Some of these methods are – Method of maximum likelihood estimation (MLE), Method of least squares, Method of moments, Method of minimum variance unbiased estimation (MVUE) etc. All these estimators should satisfy the statistical properties described above. The one that satisfies most of these statistical properties is considered as the best estimator.

Two methods of estimation viz, the method of MLE and the method of MVUE are explained below.

A. Method of maximum likelihood estimation

(MLE)[3]: Let $x_1, x_2, x_3, \dots, x_n$ be a random sample of size n from a population with probability function $f(x; \theta)$. Then, the likelihood function is given by

$$L = \prod_{i=1}^n f(x_i, \theta)$$

We have to find an estimator that maximizes the likelihood function.

By differential calculus, this means

$$\frac{\partial L}{\partial \theta} = 0 \text{ with } \frac{\partial^2 L}{\partial \theta^2} < 0$$

Since $L > 0$, we can write the above as,

$$\frac{1}{L} \frac{\partial L}{\partial \theta} = 0 \text{ i. e., } \frac{\partial(\ln L)}{\partial \theta} = 0.$$

Thus, MLE is the solution of

$$\frac{\partial(\ln L)}{\partial \theta} = 0 \text{ with } \frac{\partial^2(\ln L)}{\partial \theta^2} < 0$$

B. Method of minimum variance unbiased estimation

(MVUE)[4]: If a statistic T based on a sample of size n is such that - T is unbiased for θ and has the smallest variance among the class of all unbiased estimators of θ , then, T is called the MVUE of θ . It is found that such an MVUE is always unique. To find this MVUE, we start with an unbiased estimator U and then improve upon it by defining a function $\Phi(t)$ of the sufficient statistic T . This procedure, called Blackwellization is explained below:

Let $U=U(x_1, x_2, \dots, x_n)$ be an unbiased estimator of parameter θ and let $T=T(x_1, x_2, \dots, x_n)$ be a sufficient statistic for θ . Consider the function $\Phi(t)$ of this sufficient statistic, defined as $\Phi(t)=E[U/T=t]$. Then, $E[\Phi(t)]=\theta$ and $V[\Phi(t)] \leq V(U)$. If in addition, T is also complete, then $\Phi(t)$ is also the unique estimator.

III. RELIABILITY ESTIMATION FOR EXPONENTIAL CLASS MODELS

MLE is a very simple and widely used method of estimation. MLEs are always consistent and sufficient. But they need not be unbiased and efficient. But, MVUEs are always unbiased and efficient. Thus they are considered as the best estimators among the class of all estimators [5].

Let us consider the problem of finding MLE and MVUE of $R(t)$ and compare their measures of dispersion, viz, the variance, to find the best among the two.

MLE of $R(t)$: The maximum likelihood estimators have a property that, if $\hat{\theta}$ is an MLE of θ , and if $g(\theta)$ is any function of θ , then $g(\hat{\theta})$ is the MLE of $g(\theta)$. This is called the invariance property. Using this property, we can obtain the MLE of $R(t)$ as follows:

In the exponential class models, it is assumed that the time to failure of a software is exponentially distributed with parameter Φ , the failure rate. Thus, if T denotes the time to failure, then $T \sim \exp(\Phi)$. Hence, its probability function is given by $f(t) = \Phi e^{-\Phi t}$. Using the above procedure of finding MLE, we get the MLE of Φ as $\hat{\Phi} = \frac{n}{\sum_{i=1}^n t_i}$ [6]. Thus, using the

invariance property of MLE, we get the MLE of $R(t)$ as

$$\hat{R}(t) = e^{-\hat{\Phi}t} = e^{-\frac{nt}{\sum_{i=1}^n t_i}} \text{-----(1)}$$

MVUE of $R(t)$: To find the MVUE of $R(t) = P(T \geq t)$, define a function

$U(t_1, t_2, \dots, t_n)$, such that

$$U(t_i) = \begin{cases} 1 & \text{if } t_i \geq t \\ 0 & \text{otherwise} \end{cases}$$

Thus, $E[U(t_i)] = 1 \cdot P(T_i \geq t) + 0 \cdot P(T_i \leq t) = P(T_i \geq t) = R(t)$.

This shows that $U(t_i)$ is unbiased for $R(t)$.

Also, the likelihood function is given by

$$L = \prod_{i=1}^n f(t_i) = \Phi^n e^{-\Phi \sum_{i=1}^n t_i}$$

Using factorization theorem, $\sum_{i=1}^n t_i$ is the complete sufficient

estimator for the parameter Φ .

Thus, MVUE of $R(t)$ [7] is obtained as

$$\tilde{R}(t) = E[U(t_i) | \sum_{i=1}^n t_i] = \int_t^\infty f(t_i | \sum_{i=1}^n t_i) dt_i \text{-----(2)}$$

$$\text{But, } f(t_i | \sum_{i=1}^n t_i) = \frac{f(t_i, \sum_{i=1}^n t_i)}{f(\sum_{i=1}^n t_i)} \text{-----(3)}$$

It can be shown that

$$f(\sum_{i=1}^n t_i) = \frac{\Phi^n}{\Gamma n} e^{-\Phi \sum_{i=1}^n t_i} (\sum_{i=1}^n t_i)^{n-1} \text{-----(4)}$$

To find the probability function $f(t_1, \sum_{i=1}^n t_i)$, we split the sample $(T_1, T_2, T_3, \dots, T_n)$ into two samples as $T_1 = t_1$ of size one and (T_2, T_3, \dots, T_n) of size $n-1$. Noting that T_i s are independent,

and are exponentially distributed, we get the joint probability

function of T_1 and $\sum_{i=2}^n T_i$ as

$$f(t_1, \sum_{i=2}^n t_i) = \frac{e^{-\Phi \sum_{i=1}^n t_i}}{\Gamma(n-1)} \Phi^n \left(\sum_{i=2}^n t_i \right)^{n-1}$$

Considering the transformation $\sum_{i=1}^n t_i = t_1 + \sum_{i=2}^n t_i$, the Jacobean of the inverse transformation and simplifying, we get the probability function in the numerator of (3) as

$$f(t_1, \sum_{i=1}^n t_i) = \frac{e^{-\Phi \sum_{i=1}^n t_i}}{\Gamma(n-1)} \Phi^n \left(\sum_{i=1}^n t_i - t_1 \right)^{n-1} \dots\dots\dots(5)$$

Now, using (4) and (5) in (3) and simplifying, we get the conditional probability function as

$$f(t_1 | \sum_{i=1}^n t_i) = \begin{cases} (n-1) \left(1 - \frac{t_1}{\sum_{i=1}^n t_i} \right)^{n-1} & \text{if } 0 < t_1 < \sum_{i=1}^n t_i \\ 0 & \text{otherwise} \end{cases}$$

Substituting the above probability function in (2), we get the estimate (MVUE) of reliability as

$$\tilde{R}(t) = \begin{cases} \left(\frac{n-1}{n} \right) \left(1 - \frac{t}{\sum_{i=1}^n t_i} \right)^{n-1} & t < \sum_{i=1}^n t_i \\ 0 & \text{otherwise} \end{cases} \dots\dots\dots(6)$$

IV. DISCUSSION

MLE and MVUE are the two estimators which are considered as the most efficient estimators among the class of all estimators. The difference however exists between the two. Estimated values of MVUE of R(t) are always found to be less than estimated values of MLE of R(t), as seen by the two case studies which give failure times of different failures observed by Lyu[1] for an exponential class model, viz, the Jelinski - Moranda model [8]. Let us find the error in R(t).

Case study I: TABLE I gives the failure time of 10 failures. For this data, the MLE of the failure rate Φ is given by $\hat{\Phi} = 0.01335$. The MLE and MVUE of the reliability function R(t) are given by

$$\hat{R}(t) = e^{-0.01335t} \quad \text{and} \quad \tilde{R}(t) = 0.9 \left(1 - \frac{t}{749} \right)^{10}$$

Calculations in TABLE II show the values of MLE and MVUE of R(t) and error in R(t) values.

TABLE I

Failure Number	Time of Failure
1	7
2	18
3	26
4	36
5	51
6	73
7	93
8	118
9	146
10	181

TABLE II

Failure time	R(t)-MLE ($\hat{R}(t)$)	R(t)-MVUE ($\tilde{R}(t)$)	Error
7	.91078356	.81933854	0.09144502
18	.78639191	.70566423	0.08072768
26	.70673465	.63213132	0.07460333
36	.61841223	.54994547	0.06846676
51	.50618655	.44460932	0.06157723
73	.37736213	.32276861	0.05459352
93	.28893602	.23903568	0.04990034
118	.20694546	.16207571	0.04486975
146	.14240218	.10294315	0.03945903
181	.08924677	.05661189	0.03263488

From the calculations of the above table, we can see that as the failure time goes on increasing, the error factor in the two estimates goes on decreasing exponentially. Thus for a very large failure time, the error factor almost reaches zero.

Case study II: Table III shows failure times of 21 failures. MLE of failure rate Φ is given by $\hat{\Phi} = 0.00868$. The MLE and MVUE of reliability are given by

$$\hat{R}(t) = e^{-0.00868t} \quad \text{and} \quad \tilde{R}(t) = 0.9524 \left(1 - \frac{t}{2417.53} \right)^{21}$$

Table IV shows MLE and MVUE of R(t) and error in R(t) values.

TABLE III

Failure Number	Time of Failure
1	15.7
2	29.39
3	41.14
4	56.47
5	75.61
6	98.83
7	112.42
8	125.61
9	129.39
10	133.45
11	138.94
12	141.41
13	143.67
14	144.63
15	144.95
16	145.16
17	146.25
18	146.7
19	147.26
20	148.15
21	152.4

TABLE IV

Failure time	R(t)-MLE ($\hat{R}(t)$)	R(t)-MVUE ($\tilde{R}(t)$)	Error
15.7	0.8726017	0.8305943	0.042007448
29.39	0.7748349	0.7366425	0.038192387
41.14	0.6997058	0.6641610	0.035544796
56.47	0.6125286	0.5797608	0.032767779
75.61	0.5187699	0.4886687	0.030101191
98.83	0.4240752	0.3964032	0.027672
112.42	0.3768886	0.3503693	0.026519236
125.61	0.3361173	0.3105917	0.025525632
129.39	0.3252682	0.3000100	0.025258123
133.45	0.3140051	0.2890273	0.024977781
138.94	0.2993926	0.2747839	0.024608777
141.41	0.2930421	0.2685960	0.024446096
143.67	0.2873496	0.2630507	0.024298873
144.63	0.2849651	0.2607283	0.024236777
144.95	0.2841747	0.2599585	0.024216135
145.16	0.2836572	0.2594546	0.024202603
146.25	0.2809861	0.2568535	0.024132557
146.7	0.2798907	0.2557870	0.024103729
147.26	0.2785335	0.2544656	0.024067926
148.15	0.2763901	0.2523789	0.024011185
152.4	0.2663798	0.2426370	0.023742771

From the calculations of the above table, we can see that as the failure time goes on increasing, the error factor in the two estimates goes on decreasing exponentially. Thus for a very large failure time, the error factor almost reaches zero.

V. CONCLUSION

One of the best measures of software quality is the reliability. The reliability can be estimated using various methods of estimation. Two such estimators of reliability are the maximum likelihood estimator and the Minimum Variance Unbiased estimator. MLEs are consistent and sufficient, while MVUEs are unbiased and efficient. Thus, differences in two estimators exist. The difference or the error in the two estimates was found for Jelinski-Moranda model, which is an exponential class software reliability model. From the two case studies, it is found that the estimated values as obtained by MVUE are less than the values obtained by MLE. Further, on finding the difference between the two estimates, we can see that the error goes on decreasing exponentially as failure time goes on increasing. As the failure time reaches infinity, the error reaches zero, which means that the two estimates coincide for a very large failure time.

ACKNOWLEDGMENT

The authors would like to thank NCET, Bangalore and NITK, Surathkal for their support in preparing this paper.

REFERENCES

[1] John. D. Musa, Anthony Iannino, Kazuhira Okumot, “Software Reliability Measurement, Prediction, Application”, International Edition, MC-Graw Hill, 1991.
 [2] Michael R. Lyu, “Hand book of Software Reliability Engineering”, IEEE Computer Society Press, McGraw Hill, 2004.
 [3] S. C. Gupta and V. K. Kapoor: “Theory of Estimation” in *Fundamentals of Mathematical Statistics*, 9th Edition, Sultan Chand & Sons, 1996.

[4] John D. Musa, “Software Reliability Engineering”, Second Edition, Tata McGraw-Hill, 2004.
 [5] Peter Spreij: “Parameter Estimation for a Specific Software Reliability Model”, *IEEE Transactions on Reliability*, Vol R-34, No. 4, 1985.
 [6] S. K. Sinha, B. K. Kale, “Exponential Failure Model” in *Life Testing and Reliability Estimation*, Wiley Eastern Limited, 1980.
 [7] B. Roopashri Tantri, Murulidhar N. N., “An efficient estimator of reliability for exponential class software reliability models”, *Lecture Notes on Software Engineering*, Vol 2, No. 3, pp. 201-204, 2014.
 [8] H. Joe, N. Reid: “On the Software Reliability Models of Jelinski-Moranda and Littlewood”, *IEEE Transactions on Reliability*, Vol R-34, No. 3, 1985.