

메이크햄 수명분포에 의존한 소프트웨어 평균고장간격시간에 관한 모수 추정법 비교 연구

김희철* · 문송철**

A Comparative Study of the Parameter Estimation Method about the Software Mean Time Between Failure Depending on Makeham Life Distribution

Hee Cheul Kim* · Song Chul Moon**

Abstract

For repairable software systems, the Mean Time Between Failure (MTBF) is used as a measure of software system stability. Therefore, the evaluation of software reliability requirements or reliability characteristics can be applied MTBF. In this paper, we want to compare MTBF in terms of parameter estimation using Makeham life distribution. The parameter estimates used the least square method which is regression analyzer method and the maximum likelihood method. As a result, the MTBF using the least square method shows a non-decreased pattern and case of the maximum likelihood method shows a non-increased form as the failure time increases. In comparison with the observed MTBF, MTBF using the maximum likelihood estimation is smallerd about difference of interval than the least square estimation which is regression analyzer method. Thus, In terms of MTBF, the maximum likelihood estimation has efficient than the regression analyzer method. In terms of coefficient of determination, the mean square error and mean error of prediction, the maximum likelihood method can be judged as an efficient method.

Keywords : Software Reliability, NHPP, Makeham Life Distribution, Mean Time Between Failure

Received : 2017. 03. 02. Revised : 2017. 03. 31. Final Acceptance : 2017. 03. 31.

* Namseoul Univ., Dept. of Industrial & Management Engineering, Professor, e-mail : kim1458@nsu.ac.kr

** Corresponding Author, Professor, Department of Computer Science, Namseoul University, 91 Daehak-ro, Seonghwan-eup, Seobuk-gu, Cheonan-si, Chungcheongnam-do, 31020, Korea, e-mail : moon@nsu.ac.kr

1. 서 론

소프트웨어 신제품개발에서 신뢰성 특징을 응용하여 고객이 요구하는 항목의 신뢰성을 갖는 소프트웨어 제품을 보다 경제적인 비용과 개발기간을 투자하여 개발하는 과정이 필요하다. 이런 신뢰특성을 실현하기 위해서는 소프트웨어 개발자의 개발환경과 테스트 비용의 최소비용이 필수불가결한 요인이 된다. 현재까지 많은 소프트웨어 신뢰특성 요인분석에 대한 연구가 제안되어 왔다. 여러 연구 중에서 비동질적 포아송 과정(Non-Homogeneous Poisson Process; NHPP)에 기초를 둔 신뢰성 모형[Gokhale and Trivedi, 1999; Goel and Okumoto, 1978]은 결함 관찰요인 측면에서는 효율적인 모형으로 알려져 있고 결함이 관찰되면 즉각적으로 디버깅되고 이러한 디버깅 시행에서 새로운 결함이 관찰되지 않는다는 사전 가정에 바탕을 두고 있다. 본 연구에서는 수리가 가능한 소프트웨어 시스템의 경우 평균고장간격시간(Mean time between failure; MTBF)은 시스템 안정성에 대한 기본척도로 사용된다. 따라서 소프트웨어 신뢰도의 요구조건이나 신뢰도 평가는 평균고장간격시간을 이용한다[Lutfiah Ismail, 2014]. 최근에 수명분포가 와이블 분포를 따르는 경우 MTBF 측면에서 모수 추정법 비교에 있어서 최우추정법보다는 회귀분석기법인 최소자승법을 이용한 모수추정법이 효율적임을 확인하기도 하였다[Lutfiah Ismail, 2014]. 본 논문에서는 이를 위하여 NHPP에 기초하고 모수추정은 회귀분석기법인 최소자승법과 최우추정법을 이용하고 수명분포는 메이크햄 분포를 적용하여 비교하고자 한다.

2. 메이크햄 분포 모형

메이크햄 분포(Makeham)는 수치해석적인 데

이터 분석모형에 사용되는 분포로서 강도함수는 다음과 같이 알려져 있다[Manton et al., 1986].

$$\lambda(t) = ae^{bt}, a, b > 0 \quad (1)$$

여기서도, a 는 척도수이고 b 는 형상모수를 의미하고 식 (1)을 이용하면 평균값함수는 다음과 같이 표현 할 수 있다[Kim, 2015; Tae, 2015].

$$m_{Mkm}(t) = \int_0^t \lambda(s) ds = \frac{a}{b} (e^{bt} - 1) \quad (2)$$

또한, MTBF은 다음관계로 표현 된다[Lutfiah Ismail, 2014].

$$MTBF(t) = \frac{1}{\lambda_{Mkm}(t)} = a^{-1} e^{-bt} \quad (3)$$

식 (3)은 자연로그를 취하여 다음과 같은 관계식으로 변환 할 수 있다.

$$\ln[MTBF(t)] = -\ln a - bt \quad (4)$$

2.1 최우추정법을 이용한 MTBF 추정

메이크햄 과정에 대한 강도함수인 식 (1)과 평균값함수인 식 (2)를 이용하면 NHPP 모형의 로그 우도함수는 다음과 같은 형태로 유도 할 수 있다[Tae, 2015].

$$\begin{aligned} \ln L_{NHPP}(\Theta | x) & \\ &= \ln \left[\left(\prod_{i=1}^n \lambda(x_i) \right) \exp[-m(x_n)] \right] \\ &= n \ln a + b \sum_{i=1}^n x_i - \frac{a}{b} (e^{b x_n} - 1) \end{aligned} \quad (5)$$

즉, a 와 b 에 관하여 편미분을 시행하여 식 (6)과 식 (7)과 같은 조건식을 충족하는 \hat{a}_{MLE} 와 \hat{b}_{MLE} 을 수치 해석적 기법으로 추정할 수 있다[Hee, 2016].

$$\frac{\partial \ln L_{NHPP}(\Theta | \underline{x})}{\partial a} = \frac{n}{a} - \frac{1}{b} (e^{b x_n} - 1) = 0 \quad (6)$$

$$\text{즉, } \hat{a} = \frac{n b}{e^{b x_n} - 1}$$

$$\frac{\partial \ln L_{NHPP}(\Theta | \underline{x})}{\partial b} \quad (7)$$

$$= b^2 \sum_{i=1}^n x_i + a e^{b x_n} - a b x_n e^{b x_n} - a = 0$$

단, $\underline{x} = (x_1 \leq x_2 \leq x_3 \cdots \leq x_n)$, $\Theta = (a, b)$ 는 모수공간을 의미한다.

따라서 \hat{a}_{MLE} 와 \hat{b}_{MLE} 을 이용하면 MTBF를 추정할 수 있다[Lutfiah Ismail, 2014].

2.2 최소자승법을 이용한 MTBF 추정

식 (4)을 이용하면 최소자승 추정치인 기율기 \hat{b}_{reg} 와 절편 \hat{a}_{reg} 은 다음과 같이 유도된다[Lutfiah Ismail, 2014 ; Sheldom, 2000].

$$\hat{b}_{REG} = \frac{-\sum_{i=1}^n x_i \ln(MTBF(x_i)) + \sum_{i=1}^n x_i \sum_{i=1}^n \ln(MTBF(x_i))}{n} \quad (8)$$

$$\sum_{i=1}^n (x_i)^2 - \frac{[\sum_{i=1}^n x_i]^2}{n}$$

$$\hat{a}_{REG} = e^{-\frac{\sum_{i=1}^n \ln(MTBF(x_i))}{n}} - \hat{b}_{REG} \frac{\sum_{i=1}^n x_i}{n} \quad (9)$$

따라서 \hat{a}_{REG} 와 \hat{b}_{REG} 을 이용하면 MTBF를 추정할 수 있다.

2.3 효율적 모형을 위한 판단 기준

모형에 대한 효율성 평가를 위한 척도로서 결정계수, 평균제곱오차, 평균 예측오차를 고려할

수 있다. 이러한 척도들을 비교하여 효율적인 모형을 판단할 수 있다[Lutfiah Ismail, 2014; Kuei-Chen, 2008].

2.3.1 결정계수(Coefficient of Determination; R^2)

R^2 (결정계수)는 관찰 값의 차이에 대한 설명력을 나타내는 도구로서 다음과 같은 형태로 알려져 있다[Kim, 2015; Tae, 2015].

$$R^2 = 1 - \frac{\sum_{i=1}^n (m(x_i) - \hat{m}(x_i))^2}{\sum_{i=1}^n \left(m(x_i) - \frac{\sum_{j=1}^n m(x_j)}{n} \right)^2} \quad (10)$$

즉 비교에 있어서 결정계수 값이 크면 상대적으로 효율적인 모형으로 판단된다.

2.3.2 평균제곱오차(Mean Squared Error; MSE)

평균제곱오차는 실제 관찰 값과 예측 값에 대한 차이를 측정하는 기준으로서 다음과 같이 정의된다[Kim, 2015; Tae, 2015].

$$MSE = \frac{\sum_{i=1}^n (m(x_i) - \hat{m}(x_i))^2}{n-p} \quad (11)$$

단, $m(x_i)$ 은 시간(0, x_i)까지 나타난 고장들의 누적개수를 의미하고 $\hat{m}(x_i)$ 는 x_i 시점까지 평균 값 함수로부터 추정된 고장의 누적개수를 의미하고 n 은 파라미터를 추정하는데 사용된 자료의 수를 나타내며, p 는 모수의 수를 의미한다. 즉, 작은 평균제곱오차 값을 가진 모형은 효율적인 모형으로 판단할 수 있다.

2.3.3 평균예측오차(Mean Error of Prediction;

$MEOP$)

$MEOP$ 는 MTBF 측면에서 고려된 실제 데이터와 예측된 값 사이의 변화의 절대 값의 합계

이다. *MEOP*의 평가 식은 다음과 같이 정의된다 [Lutfiah Ismail, 2014].

$$MEOP = \frac{1}{n-p+1} \sum_{i=1}^n |MTBF(x_i; a, b) - \widehat{MTBF}(x_i; a, b)| \quad (12)$$

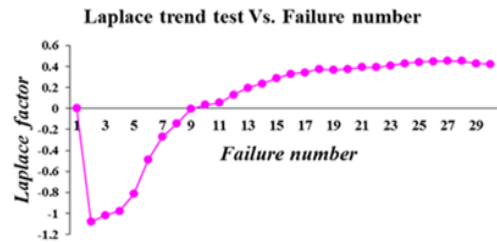
3. 소프트웨어 MTBF와 고장시간 특성 분석

이 절에서는 소프트웨어 고장시간자료 정보 [Sylwia, 2007](Failure time data)를 이용하여 논문에서 제안하는 소프트웨어 신뢰모형들에 대한 신뢰도 특성을 비교분석하고자 한다. 이 자료의 고장시간은 13.853 시간단위에 30번의 고장이 일어난 자료정보이며 <Table 1>에 요약 되어 있다. 또한 자료에 대한 신뢰성을 확보하기 위하여 자료에 대한 추세검정이 시행해야 하는데 논문에서는 라플라스 추세검정(Laplace trend test)[kim, 2015]을 이용하였다. 이 검정을 시행한 결과 <Figure 1>에서 라플라스요인(Laplace Factor)의 간격이 -2와 2사이에서 나타남으로서 즉, 극단값(Extreme value)이 포함될 확률이 거의 없으므로 이 데이터를 바탕으로 신뢰도 특성을 비교 분석하는 시행이 효율적임을 시사하고 있다 [Kim, 2015; Kanoun and Laprie, 1996].

모수추정은 최우추정법을 기초한 비선형 방정식의 계산방법은 수치 해석적 기본 방법인 이분법(Bisection method)을 사용하였다. 이러한 계산은 초기 값을 0.001과 5.000을, 허용한계(Tolerance for width of interval)는 10^{-5} 을 부여하고 수렴성을 확보하기 위하여 충분한 반복 횟수인 100번을 C-언어를 사용하여 모수추정이 이루어졌다. 최우추정법의 추정결과와 최소자승법을 이용한 모수추정을 바탕으로 모형에 대한 적합성의 기준으로서 R^2 (결정계수), *MSE*(평균제곱오차)와 *MEOP*(평균 예측오차)가 <Table 2>에 요약 되었다. 이 표에서 관찰 값의 차이에 대한 설명력을 나타내는 기준인 결정계수는 최소자승법으로 추정된 값보다 최우추정법으로 추정된 값이 큰 값을 가지므로 관찰값에 대한 설명력은 최우추정법이 효율적 방법으로 나타나고 있다. 평균값함수 측면에서는 예측된 값과 실제 값 사이의 편차를 설명하는 평균제곱오차도 최소자승법으로 추정된 값보다 최우추정법으로 추정된 값이 작은 값을 가지므로 최우추정법이 효율적 방법으로 나타나고 있다. 구체적으로 고장시점에

<Table 1> Failure Time Data of Software[Sylwia, 2007]

Failure Number	Failure Time (hours)	Failure Number	Failure Time (hours)
1	0.0094	16	10.0192
2	0.0500	17	10.4077
3	0.4064	18	10.4791
4	4.6307	19	11.0706
5	5.1741	20	11.325
6	5.8808	21	11.5284
7	6.3348	22	11.9226
8	7.1654	23	12.0294
9	7.2316	24	12.0740
10	8.2604	25	12.1835
11	9.2962	26	12.3549
12	9.3812	27	12.5381
13	9.5223	28	12.8049
14	9.8783	29	13.4615
15	9.9346	30	13.8530



<Figure 1> Laplace Trend Test

차)와 *MEOP*(평균 예측오차)가 <Table 2>에 요약 되었다. 이 표에서 관찰 값의 차이에 대한 설명력을 나타내는 기준인 결정계수는 최소자승법으로 추정된 값보다 최우추정법으로 추정된 값이 큰 값을 가지므로 관찰값에 대한 설명력은 최우추정법이 효율적 방법으로 나타나고 있다. 평균값함수 측면에서는 예측된 값과 실제 값 사이의 편차를 설명하는 평균제곱오차도 최소자승법으로 추정된 값보다 최우추정법으로 추정된 값이 작은 값을 가지므로 최우추정법이 효율적 방법으로 나타나고 있다. 구체적으로 고장시점에

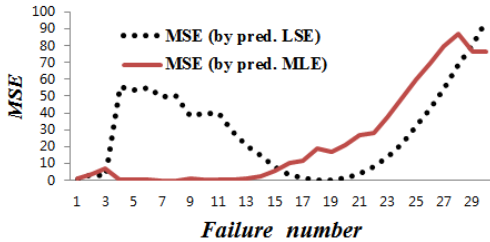
따른 평균제곱오차는 <Figure 2>에 요약되었다. 이 요약그림에서 전반부(15번째 고장까지)는 최우추정법으로 추정된 값이 작은 값을 가지며 후반부(16번째 고장부터)는 최소자승법으로 추정된 값이 작은 값을 가지는 속성을 보여주고 있다.

<Table 2> MSE and R^2 , LSE and MEOP of the Each Model

Model	Makeham	Model Comparison		
		R^2	MSE	MEOP
MLE	$\hat{b} = 0.0819$ $\hat{a} = 0.8249$	0.8864	62.6252	0.2392
LSE	$\hat{b} = -0.1447$ $\hat{a} = 3.3031$	0.6284	69.6271	0.8249

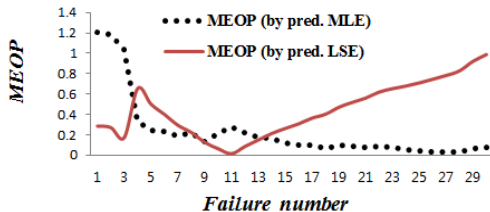
Note) MLE : Maximum likelihood estimation;
LSE : Least square estimation;
 R^2 : Coefficient of determination;
MSE : Mean square error;
MEOP : Mean error of prediction

Mean square error Vs. Failure number



<Figure 2> Mean Square Error

Mean error of prediction Vs. Failure number

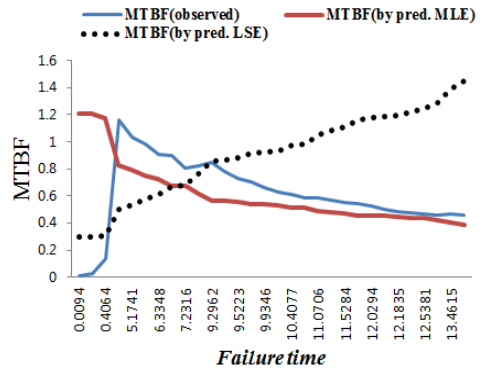


<Figure 3> Comparison of MEOP Estimates for each Model

<Figure 3>에서는 MTBF 측면에서 고려된 예측된 값 사이의 변화도의 절대 값인 평균예측

오차도 전체적으로 비교적 최우추정법으로 추정된 값이 작은 값을 가지므로 최우추정법이 효율적 방법으로 나타나고 있다.

MTBF Vs. Failure time



<Figure. 4> Comparison of MTBF Estimates for each Model

<Figure 4>의 MTBF의 추이는 고장시간이 증가함에 따라 최우추정법으로 추정된 MTBF는 비-증가형태를 나타내고 있는 반면 최소자승법으로 추정된 MTBF는 비-감소형태를 나타내고 있다. 결국 참값인 관찰된 MTBF과의 비교에 있어서 최우추정법으로 추정된 MTBF가 최소자승법으로 추정된 MTBF보다 참값과의 차이의 폭이 작은 추세를 보이고 있다. 따라서 MTBF 측면에서도 최우추정법으로 추정된 MTBF가 적합도가 더 높음을 알 수 있다.

4. 결론

빅 데이터 처리를 위하여 소프트웨어의 성능을 비교 분석하는 과정에서 결함의 발생은 거의 피할 수 없는 상황이 현실이다. 본 연구에서는 수리가 가능한 소프트웨어 시스템의 경우 소프트웨어 평균고장간격시간은 소프트웨어 시스템 안정성에 대한 기본척도로 사용된다. 따라서 소프트웨어 신뢰특성의 요구사항이나 신뢰속성 평

가는 평균고장간격을 이용 하였다. 모수추정방법에 있어서 회귀분석기법인 최소자승법과 우도함수 방법을 적용한 모수추정방법의 효율성에 대하여 비교 하였고, 수명분포는 메이크헴 분포를 이용하였다.

본 논문의 결과 최우추정법으로 추정된 평균고장간격시간은 비-증가형태를 나타내고 있는 반면 최소자승법으로 추정된 평균고장간격시간은 비-감소형태를 나타내고 있다. 결국 참값인 관찰된 평균고장간격시간과의 비교에 있어서 최우추정법으로 추정된 평균고장간격시간이 최소자승법으로 추정된 평균고장간격시간보다 참값과의 차이의 폭이 작은 추세를 보이고 있다. 따라서 평균고장간격시간 측면에서도 최우추정법으로 추정된 평균고장간격시간이 적합도가 더 높음을 알 수 있다. 따라서 최소자승추정법 보다는 우도함수를 이용하는 최우추정법을 적용한 모수 추정법이 효율적 추정방법으로 간 주 될 수 있다.

따라서 모수의 관점에서 본 논문에서 적합 시킨 소프트웨어 고장자료에 의한 모수 추정 값이기 때문에 다른 자료를 적합 시키면 모수 추정 결과는 달라질 수 있지만 최소자승추정과 최우추정법의 모수 추정법을 적용한 모형에서 최우추정법을 적용한 모형이 효율적인 모형으로 판단 할 수 있다. 최근에 Lutfiah Ismail[2014]은 신뢰성 분야에서 많이 사용되는 와이블 수명분포를 적용한 연구결과는 회귀분석 방법인 최소자승추정법이 최우추정법보다 효율적 접근방법이 됨을 확인하였지만 본 논문에서 사용된 메이크헴 수명분포를 적용했을 때는 결정계수, 평균제곱오차 및 평균예측오차도 비교적 최우추정법이 효율적 방법으로 나타나고 있다.

이 연구를 통하여 소프트웨어 운용자들은 수명분포에서 파생된 평균고장간격시간을 파악하는데 도움을 줄 수 있으리라 판단된다.

References

- [1] Goel, A. L. and Okumoto, K, "Time-dependent fault detection rate model for software and other performance measures", *IEEE Trans. Reliab.*, Vol. 28, No. 3, 1978, pp. 206-211.
- [2] Gokhale, S. S. and Trivedi, K. S. A., "A time/structure based software reliability model", *Annals of Software Engineering*, Vol. 8, No. 1, 1999, pp. 85-121.
- [3] Kanoun, K. and Laprie, J. C., "Handbook of Software Reliability Engineering", R. Lyu, Editor, chapter Trend Analysis. McGraw-Hill New York, NY, 1996, pp. 401-437.
- [4] Kim, H. C., Kim, J. B., and Moon, S. C., "A Comparative Study on Software Reliability Model for NHPP Intensity Function Following a Decreasing Pattern", *Journal of Information Technology Applications and Management*, Vol. 23, No. 4, 2016, pp. 117-125.
- [5] Kim, H.-C., "A Performance Analysis of Software Reliability Model using Lomax and Gompertz Distribution Property", *Indian Journal of Science and Technology*, Vol. 9, No. 20, 2016, pp. 1-6.
- [6] 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*, Vol. 8, No. 34, 2015, pp. 1-7.
- [7] 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*, Vol. 93, No. 10, 2008, pp. 1410-1421.
- [8] Lutfiah Ismail, A. T., "Testing the Performance of the Power Law Process Model Considering the Use of Regression Approach", *International Journal of Software Engineering and Applications*, Vol. 5, No. 5, 2014, pp. 35-46.
- [9] Manton, K. G., Stallard, E., and Vaupel, J. W., "Alternative Models for the Heterogeneity of Mortality Risks Among the Aged", *Journal of the American Statistical Association*, Vol. 81, No. 395, 1986, pp. 635-644.
- [10] Ross, S. M., "Introduction to Probability and Statistics for Engineers and Scientists", Academic Press, San Diego, USA, 2000, pp. 541-567.
- [11] Sylwia, K. B., "Makeham's Generalized Distribution", *Computational Methods in Science and Technology*, Vol. 13, No. 2, 2007, pp. 113-120.
- [12] Yang, T.-J., "The Comparative Study of NHPP Software Reliability Model Based on Log and Exponential Power Intensity Function", *The Journal of Korea Institute of Information, Electronics, and Communication Technology*, Vol. 8, No. 6, 2015, pp. 445-452.
- [13] Yoo, T.-H., "The Infinite NHPP Software Reliability Model based on Monotonic Intensity Function", *Indian Journal of Science and Technology*, Vol. 8, No. 14, 2015, pp. 1-7.

■ 저자소개



Hee Cheul Kim

He received a MS and a PhD from the Department of Statistics at Dongguk University in 1992 and 1998, He received a MS and a PhD from the

Department of Statistics at Dongguk University in 1992 and 1998, respectively. He has been a professor in the Department of Industrial & Management Engineering at Namseoul University since 2005. His current research interests include software reliability engineering, computer statistics, and information systems.



Song Chul Moon

He received a MS and a PhD from the KAIST. Dept. Info. Engineering and Kookmin Univ. Dept. Info. Mgt. He is currently a Chairman for JITAM (Journal

if Information Technology Application and Management) in Korea Information Technology Application and Management Society. Main interesting field : S/W Engineering, Ubiquitous computing, Datamining, Mgt. Info. Sys.