

# 변화점을 고려한 소프트웨어 신뢰도 성장모형의 모수추정

김도훈<sup>1</sup> · 박천건<sup>2</sup> · 남경현<sup>3</sup>

<sup>1</sup>경기대학교 응용정보통계학과; <sup>2</sup>경기대학교 수학과; <sup>3</sup>경기대학교 응용정보통계학과

(2008년 5월 접수, 2008년 6월 채택)

## 요약

비동질적 포아송과정(NHPP) 모형은 신뢰성 공학에서 소프트웨어 내에 남아있는 결함발견현상을 설명하는데 자주 사용된다. 이때 결함발견율은 연속적이며 단조함수를 가정하였으나 현실적으로 소프트웨어 시험환경, 전략 및 자원 할당 등으로 인해 결함발견율이 변하는 경우가 있다. 본 논문은 결함발견율이 변화하는 변화점 문제를 고려한 소프트웨어 신뢰도 성장모형(SRGM)을 고려하여 모수를 추정하는데 목적이 있다. 이를 위해 자료를 모의생성한 후 평균값 함수의 각 모수를 최우추정법과 최소제곱법을 이용하여 추정하며, 결함발견구간이 일정하게 증가하는 경우와 일정하게 증가하지 않는 경우를 각각 고려한다. 이때 각 모수의 적합도 비교·평가를 통하여 변화점을 고려한 SRGM에서의 최적 추정법을 수치적 방법으로 판단한다.

주요어: 소프트웨어 신뢰도 성장모형, 비동질적 포아송과정, 결함발견율, 평균값 함수, 변화점 문제.

## 1. 서론

요즘들어 소프트웨어 시스템은 컴퓨터 시스템에서 그 중요성이 점차 증대되고 있으며, 이에 따라 소프트웨어 품질요소에 대한 중요성이 부각되고 있다. 이중 소프트웨어 신뢰성은 품질요소에서 가장 기본적인 관심대상이 되었다. 소프트웨어 신뢰성이란 소프트웨어가 의도하는 기능을 제대로 수행하는가를 평가하는 것이며, 이 속성을 정량적으로 측정하는 측도중의 하나가 소프트웨어 신뢰도이다 (Pham, 1993). 이러한 소프트웨어 시스템의 신뢰도를 측정하고 평가하는 일이 소프트웨어 개발에서 중요한 문제로 대두되고 있으며, 소프트웨어 품질보증 및 보전에 대한 관심도 점차 증대되고 있다. 일반적으로 소프트웨어 개발과정은 네 단계로 구성되어 있다: 명세서(specification), 설계(design), 코딩(coding) 및 시험(testing) (Pham, 1993). 소프트웨어 결함은 소프트웨어 개발의 마지막 단계인 시험단계(testing phase)에서 탐지되고 수정되며, 이러한 결함 발견현상은 소프트웨어 신뢰도 성장모형(software reliability growth model: SRGM)으로 기술할 수 있다.

지난 20여년 동안 다양한 통계적 모형들이 소프트웨어 신뢰도를 평가하기 위해 제안되어 왔다. 이중 비동질적 포아송 과정(nonhomogeneous Poisson process: NHPP)에 기초한 소프트웨어 신뢰도 성장모형은 실제적인 소프트웨어 신뢰성 공학 면에서 매우 성공적인 이론이라는 것이 증명되었다 (Musa 등,

<sup>1</sup>(443-760) 경기도 수원시 영통구 의의동 산 94-6, 경기대학교 응용정보통계학과, 박사과정수료.

E-mail: dhkim@kyonggi.ac.kr

<sup>2</sup>(443-760) 경기도 수원시 영통구 의의동 산 94-6, 경기대학교 수학과, 시간강사.

E-mail: chungunpark@paran.com

<sup>3</sup>교신저자: (443-760) 경기도 수원시 영통구 의의동 산 94-6, 경기대학교 응용정보통계학과, 교수.

E-mail: knam@kyonggi.ac.kr

1987). 대부분의 SRGM들은 다양한 가정을 통하여 모형을 전개하는데, 결함발생 시 수정과정에서 모든 결함이 즉시 수정 및 제거되고 새로운 결함이 시스템에 도입되지 않는다는 것이다. 이러한 결함 수정 활동을 완전 수정(perfect debugging)이라 한다. 그러나 완전 수정의 가정은 실제 소프트웨어 시험 과정이나 운영과정에서 비현실적인 면이 많다. 즉, 시험과정이나 운영과정상에서 수정시 결함을 수정하지도 못하고 새로운 결함이 추가적으로 도입되는 불완전 수정 환경(imperfect debugging)을 간주하여 모형을 전개하는 것이 타당할 것이다 (Ohba, 1984; Yamada 등, 1992; Pham, 1993).

또 다른 소프트웨어 신뢰도 성장모형의 기본 가정 중의 하나는 결함발견율과 관련된 가정이다. 많은 NHPP SRGM들은 각각의 결함이란 서로 독립적이고 확률적으로 발생하는 것으로, 결함탐지과정동안 동일한 분포에 따라 발생한다 (Musa 등, 1987). 그러나 더 현실적인 상황 하에서 결함분포는 여러 요인 즉, 작업환경, 시험전략 및 작업할당과 같은 요인들에 의해 영향을 받을 수 있다. 이러한 요인들은 소프트웨어 시험단계 동안 변화하며, 이들은 소프트웨어 고장강도함수(failure intensity function)가 비단조적으로 증가하거나 감소하는 결과를 나타낼 수 있다. 이를 변화점 문제(change-point problem)라 하며 Zhao (1993), Chang (1997, 2001), Shyur (2003) 등에 의해 연구되었다. Zhao (1993)는 변화점을 고려한 Jelinski-Moranda (1972) 모형에서 최우추정법을 이용하여 모수를 추정하였다. 일반적으로 SRGM의 모수추정을 위해서는 최우추정법을 사용하여 왔으나, 최우추정법 이외에 최소제곱법을 이용하여 추정하였다 (Chang, 2001). Chang (2001)은 변화점을 고려한 Goel-Okumoto (1979) 모형으로 최소제곱법을 이용하여 모수를 추정하였으며, Shyur (2003)은 변화점을 고려한 Pham (1993) 모형으로 최우추정법을 이용하여 모수를 추정하였다.

본 논문에서는 소프트웨어 내에 결함발견현상을 기술하는 NHPP에서 결함발견율이 변화하는 변화점 문제를 고려한 SRGM의 모수를 추정하는데 목적이 있다. 이를 위해 자료를 모의생성한 후 평균값 함수의 각 모수를 최우추정법과 최소제곱법을 이용하여 추정하며, 결함발견구간이 일정하게 증가하는 경우와 일정하게 증가하지 않는 경우를 각각 고려한다. 이때 각 모수의 적합도 비교·평가를 통하여 변화점을 고려한 SRGM에서의 최적 추정법을 수치적인 방법으로 판단한다.

다음은 본 논문에서 사용되는 기호를 요약한것이다.

$N(t)$	시간 $t$ 까지 탐지된 누적 결함수를 나타내는 계수과정(counting process)
$a$	소프트웨어 내의 초기 결함수
$a(t)$	소프트웨어 시스템 내에 잠재되어 있는 누적 결함수
$b$	결함발견율
$\tau$	결함발견율의 변화점
$m(t)$	평균값함수, 시간 $t$ 까지 탐지된 기대누적결함수
$\lambda(t)$	강도함수, 시간 $t$ 에서의 순간결함 탐지율

## 2. 변화점을 고려한 NHPP 모형

대부분의 SRGM들은 결함발견율이 상수이거나 단조적으로 증가하는 함수로 가정한다. Yamada 등 (1983)은 Goel-Okumoto (1979) 모형을 수정하여 학습효과로 인한 수정과정에서의 결함발견율이 증가하는 함수로 가정하였다. 그러나 결함발견율은 시험전략과 자원배분과 같은 많은 요인들로부터 영향을 받을 수 있다. 소프트웨어 시험과정 동안 잠재적인 결함발견함수는 특정한 시점  $\tau$ 에서 변화가 가능하며, 이를 변화점이라 한다. 따라서 소프트웨어 신뢰도 모형에서 변화점 문제를 고려한다면 보다 더 현실적인 상황을 반영한 모형을 세울 수 있다.

Zhao (1993)는 변화점 배분을 추정하기 위해 Jelinski와 Moranda (1972)의 모형을 변형하였다. Chang

(1997)은 Goel-Okumoto (1979)에서 변화점 문제를 고려하였다. 여기서 미지의 모수  $\tau$ 가 변화점이 된다. Chang (1997)의 모형에서 결합발견율은 다음과 같이 정의된다.

$$b(t) = \begin{cases} b_1, & 0 \leq t \leq \tau, \\ b_2, & t > \tau, \end{cases}$$

여기서  $\tau$ 는 변화점이다. 가정으로부터 Goel-Okumoto 모형의 평균값함수  $m(t)$  그리고 강도함수  $\lambda(t)$ 를 도출할 수 있다.

$$m(t) = \begin{cases} a[1 - \exp(-b_1t)], & 0 \leq t \leq \tau, \\ a[1 - \exp(-b_1\tau - b_2(t - \tau))], & t > \tau, \end{cases} \quad (2.1)$$

$$\lambda(t) = \frac{dm(t)}{dt} = \begin{cases} ab_1 \exp(-b_1t), & 0 \leq t \leq \tau, \\ ab_2 \exp(-b_1\tau - b_2(t - \tau)), & t > \tau. \end{cases}$$

본 논문에서도 변화점을 고려한 Goel-Okumoto (1979)모형의 평균값함수 식 (2.1)을 이용하여 모의 실험자료 및 실제자료에 적용한 후 모수를 추정하는데 사용할 것이다.

### 3. 모수추정

#### 3.1. 최우추정법

$N(t)$ 를 시간  $t$ 까지 탐지된 누적 결함수를 나타내는 계수과정(counting process)이라 하자. 결함자료  $(t_k, y(t_k))$ ,  $k = 1, 2, \dots, n$ ;  $0 < t_1 < \dots < t_n$ 는 소프트웨어 개발 과정에서 관찰될 수 있으며,  $y(t_k)$ 는 주어진 시점  $(0, t_k]$  사이에서 발견된 누적 결함수를 나타낸다. 따라서 평균값함수  $m(t)$ 를 갖는 NHPP 모형에 기초한 결합확률밀도함수 즉, 우도함수는

$$L = \Pr[N(t_1) = y_1, \dots, N(t_n) = y_n] \\ = \prod_{k=1}^n \frac{[m(t_k) - m(t_{k-1})]^{(y_k - y_{k-1})}}{(y_k - y_{k-1})!} \times \exp[-(m(t_k) - m(t_{k-1}))] \quad (3.1)$$

이며,  $t_0 = 0, y_0 = 0$ 이다. 식 (3.1)에서 양변에 자연로그를 취하면

$$\ln L = \sum_{k=1}^{N(\tau)} \{(y_k - y_{k-1}) \ln[m(t_k) - m(t_{k-1})] - \ln(y_k - y_{k-1})!\} \\ + \sum_{k=N(\tau)+1}^n \{(y_k - y_{k-1}) \ln[m(t_k) - m(t_{k-1})] - \ln(y_k - y_{k-1})!\} - m(t_n). \quad (3.2)$$

과 같다. 위 식 (3.2)에 (2.1)을 대입하여 전개하면

$$\ln L = \sum_{k=1}^{N(\tau)} \{(y_k - y_{k-1}) \ln[a(\exp(-b_1t_{k-1}) - \exp(-b_1t_k))]\} \\ + \sum_{k=N(\tau)+1}^n \{(y_k - y_{k-1}) \ln[a(\exp(-b_1\tau - b_2(t_{k-1} - \tau)) - \exp(-b_1\tau - b_2(t_k - \tau)))]\} \\ - a[1 - \exp(-b_1\tau - b_2(t_n - \tau))]. \quad (3.3)$$

표 4.1. 모의실험시 모수값

$a$	$b_1$	$b_2$	$\tau$
500	0.05	0.1	30
1,000	0.05	0.1	30
1,500	0.05	0.1	30
500	0.01	0.1	30
500	0.005	0.1	30
500	0.05	0.2	30
500	0.05	0.5	30

표 4.2. 모의자료에 대한 모수추정값(변화점이 존재하지 않는 경우)

모수	추정법	$\hat{a}$	$\hat{b}$
$n = 30, a = 500, b = 0.05$	MLE	525.66989 (0.97725)	0.05345 (0.00313)
	LSE	548.57185 (1.54233)	0.05545 (0.00569)
$n = 60, a = 500, b = 0.05$	MLE	505.12547 (0.89985)	0.05254 (0.00222)
	LSE	511.35699 (0.97567)	0.05348 (0.00540)

와 같다. 따라서 식 (3.3)을 최대화 하는 모수  $a, b_1, b_2, \tau$ 에 대한 추정량은 모수 각각에 대하여 편미분하여 0을 대입하여 풀면 얻을 수 있다.

$$\frac{\partial \ln L}{\partial a} = \frac{\partial \ln L}{\partial b_1} = \frac{\partial \ln L}{\partial b_2} = \frac{\partial \ln L}{\partial \tau} = 0.$$

### 3.2. 최소제곱법

최소제곱법에 의한 모수의 추정량을 구하기 위하여 Chang (2001)에서 제시한 방법을 이용하였다.  $E\{N(t)\} = m(t)$ 이므로 최소제곱법을 이용하여 추정량을 구하면

$$\begin{aligned} S(a, b_1, b_2, \tau) &= \sum_{k=1}^n \{N(t_k) - m(t_k)\}^2 \\ &= \sum_{k=1}^{N(\tau)} [k - a\{1 - \exp(-b_1 t_k)\}]^2 \\ &\quad + \sum_{k=N(\tau)+1}^n [k - a\{1 - \exp(-b_1 \tau - b_2(t_k - \tau))\}]^2. \end{aligned} \quad (3.4)$$

과 같다. 식 (3.4)를 최소화 하는 모수  $a, b_1, b_2, \tau$ 에 대한 추정량은 모수 각각에 대하여 편미분하여 0을 대입하여 풀면 얻을 수 있다.

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b_1} = \frac{\partial S}{\partial b_2} = \frac{\partial S}{\partial \tau} = 0.$$

표 4.3. 모의자료에 대한 모수추정값( $t_k$ 가 일정하게 증가하는 경우)

모수	추정법	$\hat{a}$	$\hat{b}_1$	$\hat{b}_2$	$\hat{\tau}$
$a = 500, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	512.00570 (0.92561)	0.04499 (0.00107)	0.12467 (0.00899)	27.15649 (1.06076)
	LSE	510.74358 (1.08563)	0.05245 (0.00100)	0.12254 (0.00723)	32.88780 (0.97802)
$a = 1,000, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	1006.91900 (0.83531)	0.04672 (0.00462)	0.12167 (0.01100)	28.98670 (0.91077)
	LSE	1003.14848 (0.96988)	0.05201 (0.01006)	0.11085 (0.01038)	31.98777 (0.86759)
$a = 1,500, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	1502.04000 (0.96146)	0.04803 (0.00370)	0.11165 (0.00977)	29.24997 (1.06619)
	LSE	1502.07654 (0.98334)	0.05154 (0.00893)	0.10988 (0.01002)	30.48282 (0.82469)
$a = 500, b_1 = 0.01, b_2 = 0.10, \tau = 30$	MLE	524.92058 (1.01743)	0.00890 (0.00011)	0.13409 (0.00968)	27.70179 (1.04539)
	LSE	522.64899 (0.97274)	0.01254 (0.00510)	0.12758 (0.01003)	32.41428 (0.86307)
$a = 500, b_1 = 0.005, b_2 = 0.10, \tau = 30$	MLE	543.98140 (1.02028)	0.00422 (0.00053)	0.15866 (0.00908)	28.55549 (0.97917)
	LSE	540.65844 (0.90576)	0.00548 (0.00097)	0.14540 (0.01091)	31.28281 (0.82877)
$a = 500, b_1 = 0.05, b_2 = 0.20, \tau = 30$	MLE	510.98292 (0.91898)	0.04521 (0.00089)	0.19528 (0.00172)	29.32182 (0.88980)
	LSE	508.23239 (0.88001)	0.04588 (0.00082)	0.19882 (0.00120)	29.42101 (0.83324)
$a = 500, b_1 = 0.05, b_2 = 0.50, \tau = 30$	MLE	509.23442 (0.93598)	0.04682 (0.00090)	0.49421 (0.00532)	30.52120 (0.75292)
	LSE	505.19298 (0.90578)	0.04892 (0.00084)	0.49485 (0.00509)	29.62348 (0.77297)

#### 4. 모의실험

제안된 모형의 추정량에 대한 성능을 알아보기 위해 모의실험을 통하여 알아보았다. 이때 결합자료  $(t_k, y(t_k)), k = 1, 2, \dots, n; 0 < t_1 < \dots < t_n$ 에서  $t_k$ 가 일정하게 증가되는 경우와 일정하게 증가되지 않고 증가하는 시간이 랜덤한 경우로 나누어 각각 60개의 자료를 생성하는 모의실험을 하였다. 모의실험 시 사용한 모수는 표 4.1과 같으며, 반복수는 100으로 하였다.

##### 4.1. 변화점이 존재하지 않는 경우

변화점이 존재하지 않는 경우 최우추정법과 최소제곱법을 이용하여 평균값 함수  $m(t)$ 의 모수인  $a, b$ 를 수치적인 방법에 의해 구한 결과(평균과 표준편차)는 표 4.2와 같다. 표 4.2의 결과에서 LSE보다 MLE가 모수와 좀 더 일치함을 알 수 있으며, 결합발견구간이  $n = 30$ 보다는  $n = 60$ 에서의 추정값이 근사함을 알 수 있다.

표 4.4. 모의자료에 대한 모수추정값( $t_k$ 가 랜덤한 경우)

모수	추정법	$\hat{a}$	$\hat{b}_1$	$\hat{b}_2$	$\hat{\tau}$
$a = 500, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	519.95950 (1.08563)	0.05795 (0.00101)	0.11543 (0.00923)	45.47400 (0.97802)
	LSE	515.64812 (0.99922)	0.05556 (0.00112)	0.11055 (0.00821)	42.32309 (0.96680)
$a = 1,000, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	1010.81600 (0.96989)	0.05581 (0.01006)	0.11113 (0.01038)	36.26144 (0.86759)
	LSE	1008.99612 (0.90099)	0.05465 (0.00882)	0.01082 (0.01011)	34.22323 (0.88282)
$a = 1,500, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	1507.90100 (0.98334)	0.05304 (0.00893)	0.10495 (0.01002)	35.59128 (0.82878)
	LSE	1506.54899 (0.88472)	0.05222 (0.00877)	0.10295 (0.00982)	34.87720 (0.83221)
$a = 500, b_1 = 0.01, b_2 = 0.10, \tau = 30$	MLE	528.40795 (0.97274)	0.01168 (0.00051)	0.12379 (0.01003)	42.46525 (0.86307)
	LSE	521.65458 (0.98810)	0.01070 (0.00046)	0.11988 (0.00882)	40.23292 (0.84421)
$a = 500, b_1 = 0.005, b_2 = 0.10, \tau = 30$	MLE	578.65382 (0.90576)	0.00656 (0.00097)	0.13611 (0.01913)	36.56307 (0.82469)
	LSE	566.03310 (1.03299)	0.00661 (0.00089)	0.12789 (0.01772)	33.39349 (0.82119)
$a = 500, b_1 = 0.05, b_2 = 0.20, \tau = 30$	MLE	515.34823 (1.09820)	0.05476 (0.00118)	0.22082 (0.00909)	33.78921 (0.90782)
	LSE	513.19298 (0.95211)	0.05221 (0.00121)	0.21544 (0.00812)	33.22219 (0.88832)
$a = 500, b_1 = 0.05, b_2 = 0.50, \tau = 30$	MLE	516.39290 (1.10990)	0.05244 (0.00125)	0.52089 (0.00921)	32.99889 (0.94221)
	LSE	512.02932 (0.93227)	0.05152 (0.00128)	0.51585 (0.00778)	32.12298 (0.88732)

#### 4.2. 결함발견시간 $t_k$ 가 일정하게 증가하는 경우

결함발견시간  $t_k$ 가 1만큼씩 일정하게 증가하는 경우 모수  $a, b_1, b_2, \tau$ 의 MLE와 LSE는 표 4.3과 같다. 표 4.3의 결과를 살펴보면 모수  $a$ 가 500, 1,000, 1,500으로 증가할수록 모수값에 좀 더 일치함을 알 수 있다. 변화점 이전의 결함발견을인  $b_1$ 이 0.005, 0.01, 0.05로서 증가할수록 모수값에 일치함을 알 수 있으며, 변화점 이후의 결함발견을인  $b_2$ 가 0.1, 0.2, 0.5로 증가할수록 모수값에 일치함을 알 수 있다. 또한 모수추정방법인 최우추정법과 최소제곱법의 일치정도를 파악한 결과 최우추정법보다는 최소제곱법으로 추정된 추정값이 모수와 좀 더 근사함을 알 수 있다.

#### 4.3. 결함발견시간 $t_k$ 가 랜덤한 경우

결함발견시간  $t_k$ 가 일정하지 않게 증가하는 경우를 고려하기 위해  $T \sim \text{EXP}(30)$ 을 따르는 60개의 난수를 발생시켜  $t_k$ 를 생성하였다. 표 4.4를 살펴보면  $t_k$ 가 일정한 경우인 표 4.3과 마찬가지로 MLE보다 LSE가 모수에 좀 더 근사함을 알 수 있다. 또한  $b_1$ 과  $b_2$ 의 차이가 점점 커질수록 변화점인  $\tau$ 의 추정값이 모수에 근사함을 알 수 있다. 그림 4.1은 표 4.3에서  $a = 500, b_1 = 0.05, b_2 = 0.10, \tau = 30$ 인 경우

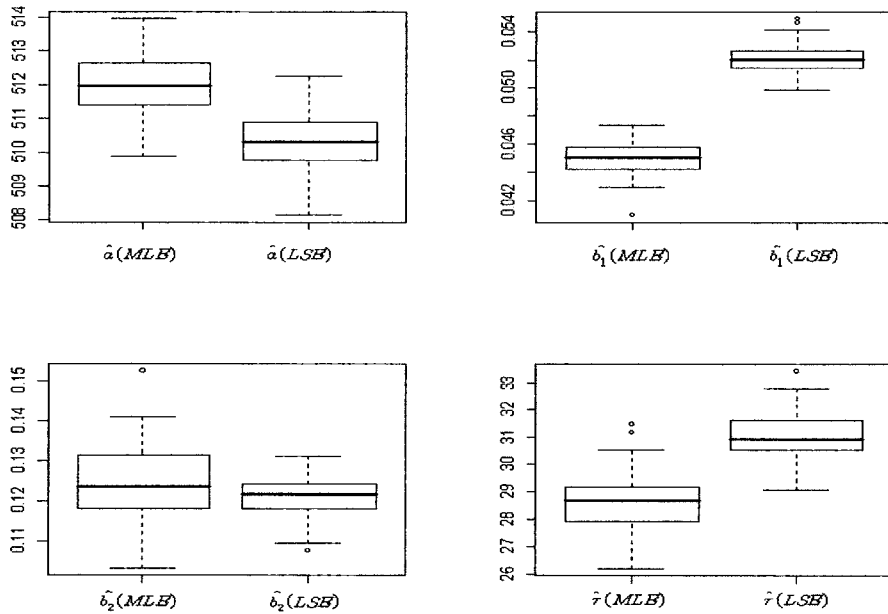


그림 4.1. 모수추정값에 대한 BOXPLOT( $a = 500, b_1 = 0.05, b_2 = 0.10, \tau = 30$ )

에 대한 boxplot을 나타내었다. 그림에서 MLE보다 LSE가 모수에 좀 더 일치함을 알 수 있으며, 편차도 더 작음을 알 수 있다.

#### 4.4. 적합도 비교검정

최우추정법으로 추정한 모형과 최소제곱법으로 추정한 모형과의 적합도를 비교하였다. 이를 위하여 MSE를 모형의 적합도 평가를 위한 기준으로 사용하였다. 우선 MSE는 다음과 같이 정의된다.

$$MSE = \frac{1}{nm} \sum_{j=1}^m \sum_{k=1}^n \{y_j(t_k) - \hat{m}_j(t_k)\}^2, \quad k = 1, \dots, 60, j = 1, \dots, 100.$$

MSE는 제안된 소프트웨어 신뢰도 성장모형이 실제 결함 데이터와 일치정도를 평가할 수 있는 척도로 사용된다. 또한 추정된 모수에 대한 적합도 평가도 함께 수행되며 이에 대한 결과는 표 4.5와 4.6과 같다. 각 모수에 대한 추정값의 MSE값을 보면 대체적으로 MLE보다는 LSE가 더 작게 나오는 것을 알 수 있다. 따라서 변화점을 고려한 경우 MLE보다는 LSE가 더 적합함을 알 수 있다.

### 5. 수치예제

본 논문에서는 변화점을 고려한 Goel-Okutomo (1979)모형에서의 모수를 추정하기 위해 Misra (1983)에 의해 제시된 소프트웨어 결함 데이터를 이용하였으며, 이는 표 5.1과 같다. 자료의 속성은 시험시간에 따라 탐지된 결함의 수를 측정하는 것이다. 식 (4)와 (6)으로부터 Misra(1983) 데이터에 대한 평균값 함수  $m(t)$ 의 모수인  $a, b_1, b_2, \tau$ 를 수치적인 방법에 의해 구한 결과는 표 5.2와 같으며, 이들의 MSE는 표 5.3과 같다. 결과를 살펴보면 모의실험과 마찬가지로 MLE보다는 LSE를 이용하여 추정된 모형에서 MSE가 더 작게 나옴을 알 수 있다.

표 4.5. 모의자료에 대한 적합도 평가( $t_k$ 가 일정할 경우)

모수	추정법	$\hat{m}(t_k)$	$\hat{a}$	$\hat{b}_1$	$\hat{b}_2$	$\hat{\tau}$
$a = 500, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	135.34570	144.98609	0.00003	0.00069	9.19951
	LSE	127.68576	135.74885	0.00003	0.00065	7.84584
$a = 1,000, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	87.54875	48.56525	0.00003	0.00059	1.84798
	LSE	82.55499	36.54844	0.00002	0.00038	1.92887
$a = 1,500, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	61.33544	5.07809	0.00002	0.00023	1.18792
	LSE	56.99855	4.21548	0.00002	0.00014	1.25871
$a = 500, b_1 = 0.01, b_2 = 0.10, \tau = 30$	MLE	182.99851	622.05995	0.00012	0.00125	6.36372
	LSE	163.41585	610.09854	0.00010	0.00111	5.58455
$a = 500, b_1 = 0.005, b_2 = 0.10, \tau = 30$	MLE	301.45689	1935.39050	0.00022	0.00352	3.03580
	LSE	255.65556	1525.48460	0.00018	0.00189	3.02558
$a = 500, b_1 = 0.05, b_2 = 0.20, \tau = 30$	MLE	123.55393	109.10021	0.00005	0.00082	2.65331
	LSE	122.48390	98.55327	0.00004	0.00076	2.44223
$a = 500, b_1 = 0.05, b_2 = 0.50, \tau = 30$	MLE	118.69430	93.55420	0.00005	0.00107	2.44788
	LSE	112.93829	36.20921	0.00004	0.00998	2.09972

표 4.6. 모의자료에 대한 적합도 평가( $t_k$ 가 랜덤한 경우)

모수	추정법	$\hat{m}(t_k)$	$\hat{a}$	$\hat{b}_1$	$\hat{b}_2$	$\hat{\tau}$
$a = 500, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	198.68854	399.54904	0.00006	0.00032	240.39162
	LSE	176.50050	385.45410	0.00005	0.00025	181.24811
$a = 1,000, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	166.54555	117.92507	0.00013	0.00023	39.95085
	LSE	138.45688	101.21540	0.00011	0.0002	30.25451
$a = 1,500, b_1 = 0.05, b_2 = 0.10, \tau = 30$	MLE	100.24544	63.38455	0.00009	0.00012	31.94243
	LSE	85.69960	60.58452	0.00008	0.00009	28.54844
$a = 500, b_1 = 0.01, b_2 = 0.10, \tau = 30$	MLE	312.36633	807.94822	0.00019	0.00067	156.11987
	LSE	285.87455	707.05415	0.00015	0.00065	122.54550
$a = 500, b_1 = 0.005, b_2 = 0.10, \tau = 30$	MLE	885.65459	6187.23812	0.00038	0.00142	57.87338
	LSE	722.32599	3258.75450	0.00033	0.00124	39.54842
$a = 500, b_1 = 0.05, b_2 = 0.20, \tau = 30$	MLE	175.43292	230.48820	0.00007	0.00089	15.02832
	LSE	170.29483	180.92908	0.00007	0.00082	12.93923
$a = 500, b_1 = 0.05, b_2 = 0.50, \tau = 30$	MLE	178.29355	242.82712	0.00006	0.00113	10.99281
	LSE	172.98720	177.28209	0.00005	0.00099	10.22838

## 6. 결론

본 논문은 변화점을 고려한 Goel-Okumoto (1979) 모형을 따르는 자료를 모의생성한 후 최우추정법과 최소제곱법을 이용하여 평균값 함수의 여러 모수들을 추정하였으며, 결합발견구간이 균등하게 증가하는 경우와 균등하게 증가하지 않는 경우를 각각 고려하여 보았다. 또한 각 모수의 적합도 평가를 위하여 MSE를 계산하였다. 이에 대한 결과 변화점이 존재하지 않는 모형에서는 최우추정법이 최소제곱법보다 우수한 추정량의 성질을 보여주었으나 변화점이 존재하는 모형에서는 최소제곱법으로 추정한 추정값이 보다 나은 결과를 나타내었다. 결합발견구간이 균등한 경우와 균등하지 않은 경우를 살펴보면 결합발견구간이 균등한 경우의 MSE값이 더 작게 나오는 것을 알 수 있었으며, 초기결함수인  $a$ 의 값이 커질수록 또한 MSE값이 더 작게 나오는 것을 알 수 있었다. 또한 변화점 전후의 결합발견율인  $b_1$ 과  $b_2$ 의 값이 커질수록 변화점  $\tau$ 에 좀 더 일치함을 알 수 있었다.



표 5.1. 소프트웨어 결함 데이터(MISRA, 1983)

Week	Hour	Number of faults	Week	Hour	Number of faults
1	62.5	15	20	25	5
2	44	6	21	12	2
3	40	8	22	55	5
4	68	8	23	49	6
5	62	8	24	64	9
6	66	4	25	26	1
7	73	4	26	66	4
8	73.5	8	27	49	2
9	92	6	28	52	4
10	71.4	2	29	70	4
11	64.5	7	30	84.5	9
12	64.7	8	31	83	6
13	36	3	32	60	1
14	54	5	33	72.5	3
15	39.5	5	34	90	6
16	68	8	35	58	6
17	61	8	36	60	3
18	62.6	6	37	168	14
19	98.7	12	38	111.5	10

표 5.2. MISRA(1983)데이터에 대한 모수 추정값

구분	$\hat{a}$	$\hat{b}_1$	$\hat{b}_2$	$\hat{\tau}$
MLE	389.96426	0.000423	0.000527	621.98752
LSE	382.65430	0.000376	0.000489	643.77778

표 5.3. MSE에 기초한 적합도 평가

구분	MLE 적용	LSE 적용
MSE	6.2211	5.2329

본 연구는 변화점이 존재하는 경우 모의실험 및 실제 데이터를 통하여 MLE와 LSE의 변화를 살펴보았다. 앞으로의 향후 과제로서는 변화점이 존재하는 경우 MLE보다 LSE가 더 좋은 추정결과를 보이는가에 대한 이론적 고찰을 통하여 이를 입증해 보일 것이다. 더욱이 변화점이 여러개 있는 경우에 대한 추정과 더불어 변화점 존재여부에 대한 검정을 실시할 것이다.

### 참고문헌

Chang, Y. P. (1997). An analysis of software reliability with change-point models, NSC 85-2121-M031-003, National Science Council, Taiwan.

Chang, Y. P. (2001). Estimation of parameters for nonhomogeneous poisson process: Software reliability with change-point model, *Communications in Statistics: Simulation and Computation*, **30**, 623-635.

Goel, A. L. and Okumoto, K. (1979). Time-dependent error-detection rate model for software reliability and other performance measures, *IEEE Transaction on Reliability*, **R-28**, 206-211.

Jelinski, Z. and Moranda, P. B. (1972). *Software reliability research*, Statistical Computer Performance Evaluation, Freiburger, W. Ed., Academic Press, New York.

- Misra, P. N. (1983). Software reliability analysis, *IBM Systems of Journal*, **22**, 465–484.
- Musa, J. D., Iannino, A. and Okumoto, K. (1987). *Software Reliability Measurement Prediction Application*, McGraw-Hill, New York.
- Pham, H. (1993). Software reliability assessment: Imperfect debugging and multiple failure types in software development, EG&G-RAMM-10737, Idaho National Engineering Laboratory.
- Shyur, H. J. (2003). A stochastic software reliability model with imperfect-debugging and change-point, *The Journal of System and Software*, **66**, 135–141.
- Yamada, S., Ohba, M. and Osaki, S. (1983). S-shaped reliability growth modeling for software error detection, *IEEE Transaction on Reliability*, **R-32**, 475–484.
- Yamada, S., Tokuno, K. and Osaki, S. (1992). Imperfect debugging models with fault introduction rate for software reliability assessment, *International Journal of Systems Science*, **23**, 2241–2252.
- Zhao, M. (1993). Change-point problems in software and hardware reliability, *Communications in Statistical-Theory and Mathematics*, **22**, 757–768.

# A Parameter Estimation of Software Reliability Growth Model with Change-Point

Do Hoon Kim<sup>1</sup> · Chun Gun Park<sup>2</sup> · Kyung H. Nam<sup>3</sup>

<sup>1</sup>Dept. of Applied Information Statistics, Kyonggi University;

<sup>2</sup>Dept. of Mathematics, Kyonggi University;

<sup>3</sup>Dept. of Applied Information Statistics, Kyonggi University

(Received May 2008; accepted June 2008)

---

## Abstract

The non-homogeneous Poisson process(NHPP) based software reliability growth models are proved quite successful in practical software reliability engineering. The fault detection rate is usually assumed to be the continuous and monotonic function. However, the fault detection rate can be affected by many factors such as the testing strategy, running environment and resource allocation. This paper describes a parameter estimation of software reliability growth model with change-point problem. We obtain the maximum likelihood estimate(MLE) and least square estimate(LSE), and compare goodness-of-fit.

**Keywords:** Software reliability growth model, NHPP, fault-detection rate, mean value function, change-point problem.

---

---

<sup>1</sup>Ph.D Candidate, Dept. of Applied Information Statistics, Kyonggi University, Suwon 443-760, Korea.

E-mail: dhkim@kyonggi.ac.kr

<sup>2</sup>Part Time Lecture, Dept. of Mathematics, Kyonggi University, Suwon 443-760, Korea.

E-mail: chungunpark@paran.com

<sup>3</sup>Corresponding author: Professor, Dept. of Applied Information Statistics, Kyonggi University, Suwon 443-760, Korea. E-mail: knam@kyonggi.ac.kr