로그형 특성분포에 근거한 소프트웨어 최적 방출시기에 관한 비교 연구

김희철*,박형근**
*남서울대학교 산업경영공학과, **남서울대학교 전자공학과
e-mail: kim1458@nsu.ac.kr, phk315@nsu.ac.kr

The Comparative Study of Software Optimal Release Time Based on Log property Distribution

Kim Hee Cheul*, Park Hyoung Keun**
*Deptment of Industrial Management Engineering, Namseoul University
**Deptment of Electronic Engineering, Namseoul University

요 약
본 연구에서는 소프트웨어 제품을 개발하여 테스팅을 거친 후 사용자에게 인도하는 시기를 결정하는 방출문제에 대하여 연구되었다. 인도시기에 관한 모형은 무한 고장 수에 의존하는 비동질적인 포아송 과정을 적용하였다. 이러한 포아송 과정은 소프트웨어의 결함을 제거하거나 수정 작업 중에도 새로운 결함이 발생될 가능성을 반영하는 모형이다. 적용모형은 여러 수명 분포들을 적합시키는데 효율적인 특성을 가진 콤페르쯔, 파레토, 로그-로지스틱 모형과 같은 로그형 특성분포를 이용하였다. 따라서 소프트웨어 요구 신뢰도를 만족시키고 소프트웨어 개발 및 유지 총비용을 최소화 시키는 방출시간이 최적 소프트웨어 방출 정책이 된다. 본 논문의 수치적인 예에서는 고장 간격 시간 자료를 적용하고 모수추정 방법은 최우추정법을 이용하여 최적 방출시기를 추정하였다.

1. 서론

소프트웨어 신뢰성에 관한 연구 활동은 지난 30년 전부터 행해져 오고 있고 많은 신뢰도 성장 모형들이 소프트웨어에 남아 있는 고장들의 수와 소프트웨어 신뢰도의 추정에 관한 문제들을 제안해 왔다.

일반적으로 소프트웨어 개발과정은 설계단계, 디자인, 코딩 그리고 테스팅 단계를 거친다. 이러한 과정을 거친 후 소프트웨어 제품을 방출하게 되는데 방출이후에 발견되지 않은 고장들이 나타난다면 이것들에 대한 보전 비용(Maintenance cost)은 크게 증가 할 것이다. 결국, 소프트웨어 시스템 시험을 끝내고 그것을 사용자에게 넘기는 시기 결정은 매우 중요한 사항이 된다. 이러한 소프트웨어 방출시간에 대한 연구들은 대부분 유한 고장 NHPP(Non-Homogeneous Poisson Process)모형을 사용하였다[1, 2]. 이러한 유한(Finite)고장 NHPP모형은 소프트웨어가 유한개의 고장이 있고 고장 제거 단계에서는 새로운 고장이 발생하지 않는

다는 가정을 한 모형이다. 그러나 실제 고장 제거 단계에서도 새로운 고장이 발생 할 수 있다. 따라서 본 연구에서는 무한(Infinite) 고장 NHPP 모형을 이용하여 최적 방출시기에 대한 문제를 제안하고자 한다. 이 분야에서는 Musa-Okumoto의 대수 포아송 실행시간 모형[3]과 로그-파우어 모형[4]을 이용한 방출 문제에 대한 문제 들이 이미 연구되었고 최근에도 이와 관련된 문제에 대한 연구는 Yang 과 Xie(2000) 와 Huang(2005)에 의해 보증기간 할인율 및 보증기간 이후의 사후대책 등에 대하여 연구되고 있다[5, 6]. 본 연구에서는 소프트웨어의 결함을 제거하거나 수정 작업 중에도 새로운 결함이 발생될 가능성이 있는 무한고장수를 가진 여러 분포들을 적합시키는데 효율적인 특성을 가진 콤페르쯔, 파레토, 로그-로지스틱 강도함수를 이용한 최적 방출시기에 관한 문제를 다루었다. 본 논문의 적용모형은 여러 수명 분포들을 적합시키는데 효율적인 특성을 가진 콤페르쯔, 파레토, 로그-로지스틱 모형과 같은 로그형 특성분포를 이용하였다

2. 기존모형

1.1. 로그 포아송 실행시간모형

로그 포아송 실행시간(Log Poission execution time)모형[4, 10]은 1984년에 Musa 와 Okumoto에 의해서 소개된 무한 고장 소프트웨어 모형으로 평균값함수와 강도함수는 다음과 같이 알려져 있다.

$$m(t) = \frac{1}{\theta} ln(\lambda_0 \theta t + 1) \quad (1)$$

$$\lambda(t) = \frac{\lambda_0}{\lambda_0 \theta t + 1} \quad (2)$$

한편, (3)식과(.6)식을 이용하면 우도함수는 다음과 같이 유도 할 수 있다.

$$L(\theta,\lambda_0 \mid \underline{x}) = \left(\prod_{i=1}^{n} \frac{\lambda_0}{\lambda_0 \theta x_i + 1}\right) \cdot e^{-\frac{1}{\theta} ln(\lambda_0 \theta x_n + 1)} \quad (3)$$

단, $\underline{x} = (x_1 \le x_2 \le x_3, \cdots \le x_n]$.

모수 최우추정법(MLE)을 이용하기 위한 로그 우도 함수는 다음과 같이 유도된다.

$$\ln L(\theta,\lambda_0 \mid \underline{x}) = n\ln\lambda_0 - \ln\sum_{i=1}^{n}(\lambda_0\theta x_i + 1) - \frac{1}{\theta}ln(\lambda_0\theta x_n + 1) \quad (4)$$

(4)식을 이용하여 최우추정치 $\hat{\theta}_{MLE}$ 와 $\widehat{\lambda_0}_{MLE}$은 다음과 같이 구할 수 있다고 하였다.

$$\hat{\theta}_{MLE} = \frac{1}{n} ln(\hat{\phi}\ \ x_n + 1) \quad (5)$$

$$\widehat{\lambda_0}_{MLE} = \hat{\phi} / \hat{\theta}_{MLE} \quad (6)$$

단, $\phi\left(= \hat{\lambda}_{0_{MLE}} \cdot \hat{\theta}_{MLE}\right)$는 (7) 식의 근이 된다.

즉, 이 ϕ 근을 구하기 위해서는 수치 해석적 방법으로 다음과 같은 식을 이용하여 계산할 수 있다[8].

$$\frac{\partial \ln L(\phi \mid \underline{x})}{\partial \phi} = \frac{n}{\phi} - \sum_{i=1}^{n} \frac{x_i}{\phi x_i + 1} - \frac{n\, x_n}{(\phi\, x_n + 1)\ln(\phi x_n + 1)} = 0 \quad (7)$$

1.2. 로그 파우어 모형

로그 파우어 (Log power)모형[4, 11]은 1999년에 Xie와 Homg에 의해서 발전된 무한 고장 소프트웨어 모형으로 평균값함수와 강도함수는 다음과 같이 알려져 있다.

$$m(t) = a\ln^b(1+t) \quad (8)$$

$$\lambda(t) = \frac{a\, b \ln^{b-1}(1+t)}{1+t} \quad (9)$$

(8)식과 (9)식을 이용하면 우도함수는 다음과 같이 유도 할 수 있다.

$$L(a,b \mid \underline{x}) = \left(\prod_{i=1}^{n} \frac{a\, b\ln^{b-1}(1+x_i)}{1+x_i}\right) \cdot e^{-a\ln^b(1+x_n)} \quad (10)$$

단, $\underline{x} = (x_1 \le x_2 \le x_3, \cdots \le x_n]$.

최우추정법(MLE)을 이용하기 위한 로그 우도 함수는 다음과 같이 유도된다.

$$\ln L(a,b \mid \underline{x}) = n\ln a + n\ln b - (b-1)\ln\left(\sum_{i=1}^{n}\ln(1+x_i)\right) - \sum_{i=1}^{n}\ln(1+x_i) - a\ln^b(1+x_n) \quad (11)$$

(11)식에서 a 와 b 에 대하여 편미분 하여 다음과 같은 식을 만족하는 $\hat{a}_{MLE}$ 와 $\hat{b}_{MLE}$을 수치 해석적 방법으로 계산할 수 있다.

$$\frac{\partial \ln L(a,b \mid \underline{x})}{\partial a} = \frac{n}{a} - \ln^b(1+x_n) = 0 \quad (12)$$

$$\frac{\partial \ln L(a,b \mid \underline{x})}{\partial b} = \frac{n}{b} - \ln\left(\sum_{i=1}^{n}\ln(1+x_i)\right) - a\ln^b(1+x_n)\ln(\ln(1+x_n)) = 0 \quad (13)$$

3. 로그형 특성 모형

3.1 콤페르쯔 모형

콤페르쯔 분포(Gompertz)는 수명분포로 적용 할 수 있으며 다음과 같이 정의 된다.

$$\lambda(t) = \exp(\alpha + \beta t) \quad (14)$$

단, $0 < t < \infty,\ -\infty < \alpha < \infty\ ,\ \beta > 0$.

따라서 평균값 함수는 $m(t) = \int_0^t \lambda(s)ds$ 을 이용하면 다음과 같이 유도 할 수 있다.

$$m(t) = \frac{e^{\alpha}(e^{\beta t} - 1)}{\beta} \quad (15)$$

(14)와(15)식을 이용하면 다음과 같이 표현된다.

$$L(\alpha,\beta \mid \underline{x}) = \exp\left(n\alpha + \beta\sum_{i=1}^{n} x_i - \frac{e^{\alpha}(e^{\beta x_n} - 1)}{\beta}\right) \quad (16)$$

단, $\underline{x} = (x_1 \le x_2 \le x_3, \cdots \le x_n]$.

(16)식에 의해서 α 와 β 에 관한 최우추정치 $\hat{\alpha}_{MLE}$ 와 $\hat{\beta}_{MLE}$는 다음 식을 이용하면 추정 할 수 있다고 하였다[9].

$$\sum_{i=1}^{n} x_i + \frac{n}{\beta} - \frac{n\, x_n}{1 - e^{-\beta x_n}} = 0 \quad (17)$$

$$e^{\alpha} = \frac{n\,\beta}{e^{\beta x_n} - 1} \quad (18)$$

따라서 콤페르쯔 특성을 사용한 최적 방출시간 T_{OP} 는 T_R과 T_C에 대하여 다음을 만족한다[11].

$$T_{OP} = Max(T_C, T_R) \quad (19)$$

(19) 식에서 T_R과 T_C는 다음 두 방정식에 의해서 계산 될 수 있다.

$$\ln R_0 = -\frac{e^{\alpha}}{\beta}\left[e^{\beta(T_R + x)} - e^{\beta T_R}\right] \quad (20)$$

$$T_C = \frac{\ln\left(\frac{-c_3}{(c_1 - c_2)\, e^{\alpha}}\right)}{\beta} \quad (21)$$

단, 신뢰도 $R_0 = \hat{R}(x \mid T)$, x 는 임무시간.

3.2 파레토 모형

확률분포의 꼬리 부분이 정규분포보다 두꺼운(Heavy-tailed) 즉,

다소 극단적인 값들이 발생할 확률을 무시할 수 없는 파레토 (Parto distribution) 분포의 확률 밀도 함수와 분포 함수는 각각 다음과 같다[7].

$$f_{\text{Preto}}(t \mid \alpha, \beta) = \frac{\alpha \beta^{\alpha}}{(\beta + t)^{(1+\alpha)}} \ (\alpha, \ \beta > 0, \ t > 0) \quad (22)$$

$$F_{\text{Preto}}(t \mid \alpha, \beta) = 1 - \left(\frac{\beta^{a}}{(t+\beta)^{\alpha}} \right) \quad (23)$$

위 분포함수에서 강도함수와 평균값 함수는 다음과 같이 표현된다.

$$\lambda(t) = \frac{\alpha}{t+\beta} \quad (24)$$

$$m(t) = \alpha \ln \left(1 + \frac{t}{\beta} \right) \quad (25)$$

(24)와 (25)식을 이용하면 무한 NHPP 고장 우도 함수는 다음과 같다[9].

$$L(\alpha, \beta \mid \underline{x}) = \left(\prod_{i=1}^{n} \frac{\alpha}{\beta + x_i} \right) \exp \left\{ -\alpha \ln \left(1 + \frac{x_n}{\beta} \right) \right\} \quad (26)$$

단, $\underline{x} = (x_1 \le x_2 \le x_3, \cdots \le x_n]$.
최우추정법(MLE)을 이용하기 위한 로그 우도 함수는 (26) 과 관련하여 다음과 같이 유도된다.

$$\ln L(\alpha, \beta \mid \underline{x}) = n \ln \alpha - \sum_{i=1}^{n} \ln(\beta + x_i) - \alpha \ln \left(1 + \frac{x_n}{\beta} \right) \quad (27)$$

(27)식에서 α 와 β 에 대하여 편미분 하여 다음과 같은 식을 만족하는 $\hat{\alpha}_{MLE}$ 와 $\hat{\beta}_{MLE}$을 수치 해석적 방법으로 계산할 수 있다.

$$\frac{\partial \ln L(\alpha, \beta \mid \underline{x})}{\partial \alpha} = \frac{n}{\alpha} - \ln \left(1 + \frac{x_n}{\beta} \right) = 0 \quad (28)$$

$$\frac{\partial \ln L(\alpha, \beta \mid \underline{x})}{\partial \beta} = -\sum_{i=1}^{n} \frac{1}{(\beta + x_i)} - \alpha \left(\frac{1}{\beta + x_n} - \frac{1}{\beta} \right) = 0 \quad (29)$$

따라서 최적 방출시간 T_{OP} 는 T_R과 T_C에 대하여 다음을 만족한다[11].

$$T_{OP} = Max(T_C, T_R) \quad (30)$$

(30)식 에서 T_R과 T_C는 다음 두 방정식에 의해서 계산 될 수 있다.

$$\ln R_0 = -\alpha \left\{ \ln \left(\frac{\beta + T_R + x}{\beta + T_R} \right) \right\} \quad (31)$$

$$T_C = \exp \left\{ \alpha \ln \beta + \frac{-c_3}{(c_1 - c_2)\alpha} \right\} - \beta \quad (32)$$

단, 신뢰도 $R_0 = \hat{R}(x \mid T)$, x 는 임무시간.

4.3 강도함수가 로그-로지스틱 패턴

로그-로지스틱(Log-Logistic)분포[1]는 강도함수(위험함수)가 증가하다가 감소하는 속성을 가지고 있다. 이 분포의 확률밀도 함수와 분포함수는 다음과 같이 표현 됨이 알려져 있다[12].

$$f(t \mid \tau, k) = (\lambda k (\tau t)^{k-1}) / [1 + (\tau t)^k]^2 \quad (33)$$

$$F(t \mid \tau, k) = (\tau t)^k / [1 + (\tau t)^k] \quad (34)$$

단, $t > 0$. 로그-로지스틱 모형을 무한 NHPP로 접근하면 평균값 함수와 강도 함수는 다음과 같이 표현 할 수 있다.

$$m(t \mid \tau, k) = \ln(1 + (\tau t)^k) \quad (35)$$

$$\lambda(t \mid \tau, k) = \frac{\tau k (\tau t)^{k-1}}{1 + (\tau t)^k} = \frac{\tau^k \, k \, t^{k-1}}{1 + (\tau t)^k} \quad (36)$$

(35)와(36)식을 이용하면 우도함수는 다음과 같이 유도 할 수 있다.

$$L(\tau, k \mid \underline{x}) = \left(\prod_{i=1}^{n} \frac{\tau^k \, k \, x_i^{k-1}}{1 + (\tau x_i)^k} \right) \cdot \left[1 + (\tau x_n)^k \right]^{-1} \quad (37)$$

단, $\underline{x} = (x_1 \le x_2 \le x_3, \cdots \le x_n]$.
최우추정법(MLE)을 이용하기 위한 로그 우도 함수는 (55)식과 관련하여 다음과 같이 유도된다.

$$\ln L(\tau, k \mid \underline{x}) = kn \ln \tau + n \ln k + (k-1) \sum_{i=1}^{n} \ln x_i - \sum_{i=1}^{n} \ln(1 + (\tau x_i)^k) - \ln[1 + (\tau x_n)^k] \quad (38)$$

(38)식에서 τ 와 r에 대하여 편미분 하여 다음과 같은 식을 만족하는 $\hat{\tau}_{MLE}$ 와 $\hat{k}_{MLE}$을 수치 해석적 방법으로 계산할 수 있다[12]. 본 연구에서는 $k=2$일 경우를 고려 하였다.

$$\frac{\partial \ln L(\tau, k \mid \underline{x})}{\partial \tau} = \frac{kn}{\tau} - \sum_{i=1}^{n} \frac{x_i^k \, k \, \tau^{k-1}}{1 + (\tau x_i)^k} - \frac{k \tau^{k-1} x_n^k}{1 + (\tau x_n)^k} = 0 \quad (39)$$

$$\frac{\partial \ln L(\tau, k \mid \underline{x})}{\partial k} = n \ln \tau + \frac{n}{k} + \sum_{i=1}^{n} \ln x_i - \sum_{i=1}^{n} \frac{(\tau x_i)^k (\ln \tau x_i)}{1 + (\tau x_i)^k} - \frac{(\tau x_n)^k (\ln \tau x_n)}{1 + (\tau x_n)^k} = 0 \quad (40)$$

따라서 로그-로지스틱분포 모형을 사용한 최적 방출시간 T_{OP} 는 T_R과 T_C에 대하여 다음을 만족한다[11].

$$T_{OP} = Max(T_C, T_R) \quad (41)$$

(41) 식에서 T_R과 T_C는 다음 두 방정식에 의해서 계산 될 수 있다.

$$\ln R_0 = -\ln \left(1 + \{\tau(x + T_R)\}^k \right) + \ln \left(1 + (\tau T_R)^k \right) \quad (42)$$

$$\frac{(c_1 - c_2) \, k \, T_C^{k-1} \tau^k}{\ln(1 + (\tau T_C)^k)} + c_3 = 0 \quad (43)$$

단, 신뢰도 $R_0 = \hat{R}(x \mid T)$, x 는 임무시간.

4. 결론

이 장에서 S27[13]가 인용한 고장 간격 시간 자료(Failure interval time data)를 가지고 강도함수에 따른 최적 방출시기를 분석하고자 한다. 식의 계산방법은 수치해석적 기본 방법인 이분법(Bisection method)을 사용하였다. 이러한 계산은 초기값을 10^{-6} 와 20 을, 허용 한계(Tolerance for width of interval)는 10^{-10} 을 주고 수렴성을 확인 하면서 충분한 반복 횟수인 100번을 C-언어를 이용하여 모수 추정을 수행하였다. 로그-로지스틱 모형은 $k=2$ 일 때를 고정하여 모수 τ을 계산 한 값이 표 1에 요약되었다.

[표 1] 각 모형의 모수 추정값

Model	MLE	
Log Poission execution time	$\hat{\theta}=0.020031$	$\hat{\lambda_0}=0.053067$
Log Power	$\hat{a}=16.058338$	$\hat{b}=0.478584$
Gompertz	$\alpha=-3.72358$	$\hat{\beta}=0.000552$
Pareto	$\alpha=6.345853$	$\hat{\beta}=1.875762$
Log-Logistic	$k=2.0$ **(fixed)**	$\hat{\tau}=0.000069$

[표 2] 최적 방출시간 T_{OP} $(R_0=0.95)$

Model		추정시간	T_{OP}
Log Poission execution time		$\widehat{T_R}=518.4197$ $\widehat{T_C}=556.9223$	556.9223
Log Power		$\widehat{T_R}=136.7676$ $\widehat{T_C}=102.5704$	136.7676
Gompertz		$\widehat{T_R}=629.5485$ $\widehat{T_C}=586.4359$	629.5485
Pareto		$\widehat{T_R}=182.9508$ $\widehat{T_C}=52.55281$	182.9508
Log-Logistic	$k=2.0$	$\widehat{T_R}=99.4723$ $\widehat{T_C}=60.0005$	99.4723

표 2에서는 c_1=5($), c_2=20($) 그리고 c_3=0.5($) 라고 가정하고 시스템 수명시간은 2000시간이고 임무시간을 x을 1.5이고 R_0 을 0.95(95%)를 투입하여 수치해석적 기본 방법인 이분법(Bisection method)을 사용하였다. 이러한 계산은 초기값을 10와 1000을, 허용 한계(Tolerance for width of interval)는 10^{-10} 을 주고 수렴성을 확인 하면서 충분한 반복 횟수인 100번을 C-언어를 이용하여 추정을 수행하였다. 각 모형에 대한 추정시간의 결과와 최적 방출시간은 표 3에 요약되었다.

이 표에서 제안된 콤페르쯔 및 파레토 모형에 대한 최적방출시간은 강도함수가 증가하거나 감소하는 패턴을 따르기 때문에 기존의 로그 파우어 모형보다는 방출시간이 길어서 비효율적이지만 로그 포아송 모형보다는 효율적으로 나타나고 있다. 그러나 제안된 로그-로지스틱모형은 강도함수가 증가하다가 감소하는 패턴이기 때문에 우수한 모형으로 나타나고 있다. 결국 다른 자료를 적용시키면 다르게 나타날 수 있지만 강도함수가 변화하는(증가하다가 감소)하는 로그-로지스틱 모형이 우수한 모형이 될 수 있음을 알 수 있었다.

따라서 제안된 모형들도 이 분야에 새로운 모형으로 선택 할 수 있음을 보여주고 있다.

참고문헌

[1] Gokhale, S. S. and Trivedi, K. S.(1999) "A time/structure based software reliability model", Annals of Software Engineering. 8, pp. 85-121.

[2] 김희철(2005) "일반화감마분포를 이용한 NHPP 소프트웨어 신뢰도 모형에 관한 연구", 한국컴퓨터정보학회논문지, 10권 6호, pp. 27-35.

[3] Musa, J. D and Okumoto, K.(1984.) "A Logarithmic Poisson Execution Time Model for Software Reliability Measurement", Proceeding the 7th International Conference on Software Engineering. pp. 230-238,

[4] Almering, V. and Genuchten, M, V and Cloudt, G. and Sonnemans, P. J. M.(2007) " Using Software Reliability Growth Models in Practice ". IEEE SOFTWARE. pp. 82-88.

[5] Yang, B. and Xie. M.(2000) "A study of operational and testing reliability in software reliability analysis". RELIABILITY ENGINEERING & SYSTEM SAFETY, Vol,70, pp,323-329.

[6] Huang, C. Y.(2005) "Cost-Reliability-optimal release policy for software reliability models incorporating improvements in testing efficiency, The journal of Systems and software, Vol,77, pp,139-155.

[7] Kuo, L. and Yang, T. Y. (1996) "Bayesian Computation of Software Reliability", Journal of the American Statistical Association, Vol.91, pp.763-773.

[8] 김희철, 박형근(2009) "와이블분포특성에 근거한 소프트웨어 최적방출시기에 관한 비교연구", 한국 산학기술학회논문지, 제10권8호, pp. 1903-1910.

[9] Lawless, J. F(1981) "Statistical Models and Methods for Lifetime Data" John Wiley & Sons, New York.

[10] Musa, J. D, Iannino, A. and Okumoto, K.(1987) "Software Reliabi lity: Measurement, Prediction, Application". McGraw Hill, New York.

[11] Xie, M. and Homg, G. Y.(1999) " Software release time determination based on unbound NHPP model". Proceeding of the 24th International Conference on Computers and Industrial Engineering. pp. 165-168.

[12] 김희철(2008) "Log-Logistic 분포모형에 근거한 소프트웨어 최적방출시기에 관한 비교연구", 한국 컴퓨터정보학회 논문지, 제13권7호, pp. 1-9.

[13] K. Kanoun , J. C. Laprie (1996). Handbook of Software Reliability Engineering, M.R.Lyu, Editor, chapter Trend Analysis. McGraw-Hill New York, NY, p.401-437.