소프트웨어 신뢰도 모형 설계에 대한 연구

김희철*, 이상식*
*송호대학 컴퓨터 정보처리과
e-mail : {khc, leess}@songho.ac.kr

A Study on Design for Software Reliability Model

*Hee-Cheul Kim, *Sang-Sik Lee
*Dept. of Computer Information Processing, Songho College

요약

우리의 주변에는 복잡한 소프트웨어 시스템(system)들로 둘러 쌓여 있으며 이러한 시스템의 혜택을 받는 일이 커짐에 따라 소프트웨어 신뢰성의 역할은 점차 커지게 되었다. 고장 시스템이 복잡해지면 고장의 원인이 하나의 수명분포에 의해서만 일어나지 않고 여러 원인이 혼합되어 발생 할 수 있다. 이러한 복잡한 시스템에 의한 우도함수가 적분하기 난해하므로 반복표본을 이용하는 깁스 알고리즘이 제안되었다.

1. 서론

수리가 가능한 시스템이나 소프웨어 신뢰도에서 관측시간 $(0,t]$사이에 발견된 에러수 $N(t)$를 모형화하는데 비동질적 포아송 프로세스(NHPP)가 널리 사용하여 왔다[1]. 강도함수 혹은 고장 발생률(ROCOF) $\lambda(t)=dE[N(t)]/dt$은 t에 대한 단조함수로 흔히 가정한다. 예를 들어 동질적 포아송 과정(HPP), Musa-Okumoto과정, Weibull 과정 그리고 Cox-Lewis 과정은 각각 ROCOF를 상수, 시간에 관한 부분 함수, 멱 함수, 대수선형 함수를 가정하였다[2].

본 논문에서는 기존의 모형인 Goel 이 제시한 일반화 모형[3]과 Yamada, Ohba-Osaki 모형[4]을 재조명하고 이러한 모형과 연관되고 신뢰도 분포로 많이 사용되는 와이블 분포의 특수형태인 레일리(Rayleigh)분포와 겜벨(Gumbel)분포[1] 를 이용한 모형을 제시하고 또, 효율적 모형을 위한 모형선택으로서 편차자승합(SSE)을 이용하여 비교하고자 한다.

본 논문의 구성의 2장에서는 관련연구로서 NHPP 모형과 소프트웨어 신뢰도에 대한 일반적 내용을 요약하였고 3장에서는 이들 모형에 대한 최우추정법을 이용한 모수 추정에 대하여 서술 하였고 또, 혼합모형을 제안하였다.

2. 최우추정법을 이용한 모수추정

시간구간 $(0,t]$에 있어서 신뢰도 성장 시험을 행하여 n개의 에러발생 시각의 데이터 $s_k(k=1,2,\cdots,n;\ 0\le s_1\le s_2\le\cdots\le s_n)$가 관측되고 $s=(s_1,s_2,\cdots,s_n)$로 표현하면 확률변수 $s_1,s_2,\cdots,s_n$ $(s_0=0)$에 대한 우도함수(likelihood function)는 다음과 같다.

$$L_{NHPP}=\left[\prod_{k=1}^{n}\lambda(s_k)\right]\exp[-m(s_n)] \qquad (2.1)$$

단, 강도함수 $\lambda(t)=m'(t)$.

따라서 (2.1)식에서 최우추정법을 사용하여 NHPP모형들의 모수 추정을 할 수 있고, 각 모형에서 모수

인 β_0는 테스트 개시 전에 소프트웨어 내에 잠재하는 총 기대 에러 수를 나타내고, β_1은 잔존에러 1개당 에러 발견율을 의미한다.

2.1 Goel 일반화(generalized) 모형

이 모형에 대한 평균값 함수는 $m(t)=\beta_0(1-e^{-\beta_1 t^d})$, $(\beta_0>0,\ \beta_1>0)$ 이라고 알려져 있다. t을 최종 고장시점 s_n으로 대치하고(d는 테스트 반영 상수) 오류 발견률 $\lambda(t)=\beta_0\beta_1 dt^{d-1}e^{-\beta_1 t^d}$ 를 이용하면 우도함수는 식 (2.1)에 의해

$$L_{GG}(\beta_0,\beta_1 \mid s_1,s_2,\cdots,s_n)$$
$$=\left(\prod_{k=1}^{n}\beta_0\beta_1 d s_k^{d-1}e^{-\beta_1 s_k^d}\right)\exp\left[-\beta_0(1-e^{-\beta_1 s_n^d})\right] \qquad (2.2)$$

가 되고, 로그우도함수를 구하면,

$$\ln L = n\ln\beta_0 + n\ln\beta_1 + n\ln d + (d-1)\sum_{k=1}^{n}\ln s_k - \beta_1\sum_{k=1}^{n}s_k^d - \beta_0(1-e^{-\beta_1 s_n^d}) \qquad (3.3)$$

으로 표현된다.

따라서 각 모수에 대한 최우추정량은 $\hat\beta_0$와 $\hat\beta_1$는 다음과 같은 형태가 된다.

$$\frac{n}{\hat\beta_0}=1-\exp(-\hat\beta_1 s_n^d) \qquad (2.4)$$

$$\frac{n}{\hat\beta_1}=\sum_{k=1}^{n}s_k^d+\hat\beta_0 s_n^d\exp(-\hat\beta_1 s_n^d) \qquad (2.5)$$

식 (2.4) 과 식 (2.5)을 수치적으로 풀어 두 모수를 구할 수 있고 신뢰도는 식 (2.9)에 의해 다음과 같이 나타낼 수 있다.

$$\hat R(x \mid t)$$
$$=\exp[-\hat\beta_0\{\exp(-\hat\beta_1 t^d)-\exp(-\hat\beta_1(t+x)^d)\}] \qquad (2.6)$$

이 모형에서 특히 $d=1$일 때 $m(t)=\beta_0(1-e^{-\beta_1 t})$ 이 되고 Goel-Okumoto 모형[3]과 같은 형태가 된다.

3.2 Yamada, Ohba-Osaki 모형

이 모형에 대한 평균값 함수는 $m(t)=\beta_0[1-(1+\beta_1 t)e^{-\beta_1 t}]$ $(\beta_0>0,\ \beta_1>0)$이라고 알려져 있고 식 (2.2)에 의해 $\lambda(t)=\beta_0\beta_1^2 te^{-\beta_1 t}$ 가 되고, 우도함수는 식 (3.1)과 관련하면 다음과 같다.

$$L_{YOO}(\beta_0,\beta_1 \mid s_1,s_2,\cdots,s_n)$$
$$=\left(\prod_{k=1}^{n}\beta_0\beta_1^2 s_k e^{-\beta_1 s_k}\right)\cdot\exp[-\beta_0[1-(1+\beta_1 s_n)e^{-\beta_1 s_n}] \qquad (2.7)$$

따라서 로그우도함수를 구하면 다음과 같이 표현된다.

$$\ln L = n\ln\beta_0+2n\ln\beta_1+\sum_{k=1}^{n}\ln s_k-\beta_1\sum_{k=1}^{n}s_k - \beta_0[1-(1+\beta_1 s_n)e^{-\beta_1 s_n}] \qquad (2.8)$$

데이터 $s=(s_1,s_2,\cdots,s_n)$가 주어질 때의 모수 β_0와 β_1에 관한 편미분식은 다음과 같은 형태로 유도된다.

$$\frac{\partial\ln L}{\partial\beta_0}=\frac{n}{\beta_0}-1+e^{-\beta_1 s_n}+\beta_1 s_n e^{-\beta_1 s_n}=0$$

$$\frac{\partial\ln L}{\partial\beta_1}=\frac{2n}{\beta_1}-\sum_{k=1}^{n}s_k-\beta_0\beta_1 s_n^2 e^{-\beta_1 s_n}=0$$

따라서 최우추정량 $\hat\beta_0$와 $\hat\beta_1$는 다음과 같은 조건식을 만족한다.

$$\frac{n}{\hat\beta_0}=1-\exp(-\hat\beta_1 s_n)-\hat\beta_1 s_n\exp(-\hat\beta_1 s_n) \qquad (2.9)$$

$$\frac{2n}{\hat\beta_1}=\sum_{k=1}^{n}s_k+\hat\beta_0\hat\beta_1 s_n^2\exp(-\hat\beta_1 s_n) \qquad (2.10)$$

식 (2.9)와 식 (2.10)을 수치적으로 풀어 두 모수를 구할수 있고 신뢰도는 다음과 같다.

$$\hat R(x \mid t)$$
$$=\exp[-\hat\beta_0\{(1+\hat\beta_1 t)\exp(-\hat\beta_1 t) \qquad (2.11)$$
$$-(1+\hat\beta_1(t+x))\exp(-\hat\beta_1(t+x))\}]$$

2.3 제안모형

2.3.1 단순모형

Goel-Okumoto[2]가 제시한 모형은 평균값 함수 $m(t)=\beta_0(1-e^{-\beta_1 t})$이고 Yamada, Ohba-Osaki가 제시한 모형은 평균값 함수가 $m(t)=\beta_0[1-(1+\beta_1 t)e^{-\beta_1 t}]$ $(\beta_0>0,\ \beta_1>0)$이라고 알려져 있다. 이 모형들의 평균값 함수를 일반화 하면 다음과 같이 표현할 수 있다.

$m(t)=\beta_0\cdot F(t)$ (즉, $\lambda(t)=\beta_0\cdot f(t)$)

여기서, $F(t)$는 분포함수이고, $f(t)$는 확률밀도함수를 의미한다. 즉, $F(t)$가 지수분포의 분포함수 $F(t)=(1-e^{-\beta_1 t})$이면 Goel-Okumoto 모형이 되고 형상모수(shape parameter)가 2인 감마분포의 분포함수 $F(t)=(1-(1+\beta_1 t)e^{-\beta_1 t})$이면 Yamada, Ohba-Osaki 모형이 된다.

따라서 본 장에서는 Goel-Okumoto 모형과 연관하여 다루지 않고 신뢰도 분포로 많이 사용되는 와이블 분포의 특수형태인 레일리(Rayleigh)분포와 겜벨(Gumbel)분포[1]를 이용하여 다음과 같은 모형을 제시하고자 한다.

2.3.2 레일리분포 모형

레일리 분포는 형상모수가 2인 와이블 분포의 특수한 경우이므로 평균값 함수는 $m(t)=\beta_0[\,1-e^{-\beta_1 t^2}\,]$ $(\beta_0>0,\ \beta_1>0)$으로 표현할 수 있고 식 (2.1)에 의해 $\lambda(t)=2\beta_0\,\beta_1\,t\,\exp(-\beta_1 t^2)$ 가 되고, 우도함수는 식 (2.1)과 관련하면 다음과 같다.

$$L_{Ray}(\beta_0,\beta_1\mid s_1,s_2,\cdots,s_n)=$$

$$\left\{\prod_{k=1}^{n}2\beta_0\,\beta_1\,x_i\,\exp(-\beta_1 x_k^2)\right\}\cdot\exp[-\beta_0\,(1-e^{-\beta_1 s_n^2})]\quad(2.12)$$

따라서 로그우도함수를 구하면 다음과 같이 표현된다.

$$\ln L=n\ln 2+n\ln\beta_0+n\ln\beta_1+\sum_{k=1}^{n}\ln s_k-\beta_1\sum_{k=1}^{n}s_k^2-\beta_0(1-e^{-\beta_1 s_n^2})\quad(2.13)$$

데이터 $s=(s_1,s_2,\cdots,s_n)$가 주어질 때의 모수 β_0와 β_1에 관한 편미분식은 다음과 같은 형태로 유도된다.

$$\frac{\partial\ln L}{\partial\beta_0}=\frac{n}{\beta_0}-1+e^{-\beta_1 s_n^2}=0$$

$$\frac{\partial\ln L}{\partial\beta_1}=\frac{n}{\beta_1}-\sum_{k=1}^{n}s_k^2-\beta_0\beta_1\,s_n^2\,e^{-s_n^2}=0$$

따라서 최우추정량 $\widehat{\beta}_0$와 $\widehat{\beta}_1$는 다음과 같은 조건식을 만족한다..

$$\frac{n}{\widehat{\beta}_0}=1-\exp(-\widehat{\beta}_1 s_n^2)\quad(2.14)$$

$$\frac{n}{\widehat{\beta}_1}=\sum_{k=1}^{n}s_k^2+\widehat{\beta}_0\widehat{\beta}_1\,s_n^2\,\exp(-s_n^2)\quad(2.15)$$

식 (2.14)와 식 (2.15)을 수치적으로 풀어 두 모수를 구할수 있고 신뢰도는 다음과 같다.

$$\widehat{R}(x\mid t)=\exp[-\beta_0\{\exp(-\beta_1 t^2)-\exp(-\beta_1(t+x)^2)\}]\quad(2.16)$$

3.3.1 겜벨분포 모형

겜벨분포의 경우의 평균값 함수는 $m(t)=\beta_0[\,1-\exp(e^{-t\beta_1})]$ $(\beta_0>0,\ \beta_1>0)$으로 표현할 수 있고 식 (2.1)에 의해 $\lambda(t)=\beta_0\,\beta_1 e^{\beta_1 t}\exp(-e^{\beta_1 t})$ 가 되고, 우도함수는 식 (2.1)과 관련하면 다음과 같다.

$$L_{Gum}(\beta_0,\beta_1\mid s_1,s_2,\cdots,s_n)=$$

$$\left\{\prod_{k=1}^{n}\beta_0\,\beta_1 e^{\beta_1 s_k}\cdot\exp(-e^{\beta_1 s_k})\right\}\exp[-\beta_0\,(1-\exp(e^{-s_n\beta_1}))]$$

따라서 로그우도함수를 구하면 다음과 같이 표현된다.

$$\ln L=n\ln\beta_0+n\ln\beta_1+\beta_1\sum_{k=1}^{n}s_k-\sum_{k=1}^{n}e^{\beta_1 s_k}-\beta_0(1-\exp(e^{-\beta_1 s_n}))\quad(2.17)$$

데이터 $s=(s_1,s_2,\cdots,s_n)$가 주어질 때의 모수 β_0와 β_1에 관한 편미분식은 다음과 같은 형태로 유도된다.

$$\frac{\partial\ln L}{\partial\beta_0}=\frac{n}{\beta_0}-1+\exp(e^{-\beta_1 s_n})=0\quad(2.18)$$

$$\frac{\partial\ln L}{\partial\beta_1}=\frac{n}{\beta_1}+\sum_{k=1}^{n}s_k-\beta_1\left(\sum_{k=1}^{n}s_k e^{s_k}\right)-\exp[\beta_1 s_n(1+e^{s_n})]=0\quad(2.19)$$

식 (2.18)와 식 (2.19)을 수치적으로 풀어 두 모수를 구할수 있고 신뢰도는 다음과 같다.

$$\widehat{R}(x\mid t)=\exp[-\beta_0\{\exp(e^{-t\beta_1})-\exp(e^{-(t+x)\beta_1})\}]\quad(2.20)$$

2.3.2 깁스 샘플링을 이용한 혼합모형

고장 시스템이 복잡해지면 고장의 원인이 하나의 수명분포에 의해서만 일어나지 않고 여러원인이 혼합되어 발생 할 수 있다. 이러한 복잡한 시스템에서 본 연구는 Rayleigh 분포(형상모수가 2인 와이블분포의 특수한 경우)와 Erlang(2) 분포(감마분포에서 형상모수가 2인 정수를 가진)로 이루어진 다음과 같은 혼합모형을 고려 할 수 있다.

$$f_m(x\mid p,\lambda_1,\lambda_2)=pf_1(\lambda_1)+(1-p)f_2(\lambda_2)$$

$$=2p\lambda_1\,x\exp(-\lambda_1 x^2)+(1-p)\lambda_2^2\,x\exp(-\lambda_2 x)\quad(2.21)$$

여기서, 데이터 $D_n=(x_1,x_2,\cdots,x_n)$ 이 주어지면 우도함수는 다음과 같다.

$$L_{MIX}(p,\lambda_1,\lambda_2\mid D_n,Z)$$

$$\propto\prod_{i=1}^{n}\{2p\lambda_1 x_i\exp(-\lambda_1 x_i^2)\cdot Z_i+(1-p)\lambda_2^2 x_i\exp(-\lambda_2 x_i)\cdot(1-Z_i)\}\quad(2.22)$$

단, $Z_i\sim Bernoulli(1,w)$이고 모수 w는 다음과 같다.

$$w=\frac{2p\lambda_1\,x_i\exp(-\lambda_1 x_i^2)}{2p\lambda_1\,x_i\exp(-\lambda_1^2 x_i^2)+(1-p)\lambda_2^2\,x_i\exp(-\lambda_2 x_i)}$$

모수 λ_1,λ_2,p 에 대해 사전밀도는 서로 독립일 때 Akman과 Huwang(1997)이 제시한 모형과 유사한 형태로 다음과 같이 가정한다.

$$\pi(p,\lambda_1,\lambda_2)=\pi_1(p\mid\lambda_1,\lambda_2)\cdot\pi_2(\lambda_1\mid\lambda_2)\cdot\pi_3(\lambda_2)\quad(2.23)$$

위의 식(2.4)에서 다음과 같은 사전밀도함수를 따른다고 가정한다.

(1) $\pi_1(p\mid\lambda_1,\lambda_2)\sim B(\lambda_1,\lambda_2)$

(2) $\pi_2(\lambda_1\mid\lambda_2)\sim\Gamma(a_1,\lambda_2)$

(3) $\pi_3(\lambda_2)\sim\Gamma(a_2,b_2)$

여기서, a_1,a_2 그리고 b_2 는 각 각 기지의 양수이고 $\Gamma(a,b)$ 는 평균이 a/b인 감마분포로 그리고

$B(c,d)$ 는 평균이 $c/(c+d)$ 인 베타분포로 표시하였다. 베이즈 정리에 의해 식(2.22)과 식(2.23)를 이용한 사후 결합밀도함수는 다음과 같다.

$$f(p,\lambda_1,\lambda_2 \mid D_n, Z)$$

$$\propto \prod_{i=1}^{n}[2p\lambda_1 x_i \exp(-\lambda_1 x_i^2)\cdot Z_i + \qquad (2.24)$$

$$(1-p)\lambda_2^2 x_i \exp(-\lambda_2 x_i)\cdot(1-Z_i)]\cdot \pi(p,\lambda_1,\lambda_2)$$

이런 형태에서 깁스 알고리즘에 이용되는 모수에 대한 완전한 사후 조건부분포들은 다음과 같이 주어진 다

(1) $Z_i \mathrm{vert} p,\lambda_1,\lambda_2, D_n \sim Bernoulli(1,w),\ i=1,,n$

$$w=\frac{2p\lambda_1 x_i \exp(-\lambda_1 x_i^2)}{2p\lambda_1 x_i \exp(-\lambda_1^2 x_i^2)+(1-p)\lambda_2^2 x_i \exp(-\lambda_2 x_i)}$$

(2) $f(p \mid \lambda,\beta,Z,D_n) \propto B\left(\sum_{i=1}^{n} Z_i+\lambda_1,\ \sum_{i=1}^{n}(1-Z_i)+\lambda_2\right)$

(3) $f(\lambda_1 \mid p,\lambda_2,Z,D_n) \propto \Gamma\left(\sum_{i=1}^{n} Z_i+a_1,\ \sum_{i=1}^{n} z\, x_i^2-\ln p+\lambda_2\right)$

(4) $f(\lambda_2 \mid \lambda,p,\beta,Z,D_n)$

$$\propto \Gamma\left(2\sum_{i=1}^{n}+a_1+a_2,\ \sum_{i=1}^{n}(1-Z_i)x_i-\ln(1-p)+b_2+\lambda_1\right)$$

위 식을 임의의 양의상수 a_1, a_2 그리고 b_2 를 초기치로 하여 t번 반복에 m번 적용하면 사후평균을 이용한 모수의 추정치 $\hat{p}, \hat{\lambda}_1, \hat{\lambda}_2$를 얻을 수 있다. *Rayleigh*$(R_1(t))$ 과 *Erlang*(2)$(R_2(t))$ 추세를 가진 혼합모형에 대한 혼합 신뢰도 함수는 다음과 같이 정의된다(Musa. Innino 와 Okumoto, 1987. p512).

$$R_m(t)=p(X>T)=\int_t^{\infty} f(p,\lambda_1,\lambda_2 \mid t)dt \qquad (2.25)$$

$$=1-F_m(t)=p\exp(-\lambda_1 t^2)+$$

$$(1-p)\{\exp(-\lambda_2 t)+\lambda_2 t\exp(-\lambda_2 t)\}$$

단, 단순모형에 대한 신뢰도 함수는 각 각 다음과 같다.

$$R_1(t)=\exp(-\beta_1 t^2)$$

$$R_2(t)=\exp(-\beta_2 t)+\beta_2 t\exp(-\beta_2 t)$$

각각의 신뢰도 함수에 앞절의 깁스 추정치를 대입한 신뢰도 $\hat{R}_1(t)$ 과 $\hat{R}_2(t)$, $\hat{R}_m(t)$ 은 임무시간(mission time) t 가 주어지면 t 에 따른 신뢰도 추이를 알아볼 수 있다.

3. 결론

고장 시스템이 복잡해지면 고장의 원인이 하나의 수명분포에 의해서만 일어나지 않고 여러 원인이 혼합되어 발생 할 수 있다. 이러한 복잡한 시스템에 의한 우도함수가 적분하기 난해하므로 반복표본을 이용하는 깁스 알고리즘이 제안되었다. 이 깁스 알고리즘은 MCMC기법으로 바람직한 사후분포로서 정상성 분포를 가지는 마코프 체인에 따라 표본들이 변화하는 알고리즘이다.

이 마코프체인의 추이 측정은 일반적으로 조건부 밀도함수의 곱으로 이루어진다. 깁스 표본의 수렴성 확인을 위하여 Gelman and Rubin(1992)이 제시한 대안적 방법을 도구로 하여 수렴성을 확인할 수 있고 모수의 사전분포가 공액(conjucate)형태를 가진 감마형 혼합모형에 대한 베이지안적 신뢰도 추이가 단순모형보다 완만한 비증가 추세가 되면 바람직한 모형이 된다. 이러한 혼합모형은 감마족 수명분포에만 국한되지 말고 비감마족 수명분포(예, 파레토, 로그노말 분포 등)인 경우의 신뢰도 추이에 대한 비교의 문제와 사후베이즈 요인을 이용한 단순모형과 혼합모형에 대한 모형선택의 문제가 기대된다.

끝으로 본 논문에서의 내용을 참고로 하여 소프트웨어 혼합 신뢰도모형이나 중첩모형에 대한 극한분포이론과 응용에 관한 연구가 기대된다.

참고문헌

[1] Johnson, N. L. and Kotz, S. " *Continuous Univariate Distributions 1*", A wiley-interscience publication, JOHN WILEY & SONS, pp. 272-289, 1970.

[2] Goel, A. L., and Okumoto, K. "Time-Dependent Error-Detection Rate Model for Software Reliability and Other Performance Measures", *IEEE Trans. Rel.,* Vol. R-28, pp. 206-211, 1979.

[3] Goel, A. L. "Software Reliability Model : Assumption, Limitations, and Applicability", *IEEE Trans. Software Eng.,* Vol. SE-11, No. 12, pp. 1411-1423, 1985.

[4] Yamada, S., Ohba, M. and Osaki, S. S-shaped Reliability Growth Modeling for Software Error Detection", *IEEE Trans. Rel.,* Vol. R-32, No. 5, pp. 475-478, 1983.