

NHPP에 기초한 소프트웨어 신뢰도 모형에 대한 비교연구

간 광 현^{*} · 김 희 철^{**} · 이 병 수^{***}

요 약

본 논문에서는 비동질 포아송 프로세스(NHPP)에 기초한 소프트웨어 에러 현상에 대한 확률 모형을 고려했다. 고장 패턴은 NHPP에 대한 강도함수와 평균값 함수로서 나타낼 수 있다. 따라서 본 논문에서는 기존의 모형인 Goel 이 제시한 일반화모형[2]과 Yamada, Ohba-Osaki 모형 [1]을 재조명하고 이러한 모형과 연관되고 신뢰도 분포로 많이 사용되는 와이블 분포의 특수형태인 레일리(Rayleigh)분포와 겐벨(Gumbel)분포 [5] 를 이용한 모형을 제시하고, 또 효율적 모형을 위한 모형선택으로서 편차자승합(SSE)을 이용하여 비교하였다. 모수의 추정을 위해서 최우추정법(MLE)과 일반적인 수치해석적 방법인 이분법을 이용하였다. 수치적인 예에서는 실측자료인 NTDS 자료[4]를 이용하여 모수 및 신뢰도를 추정하였고 편차자승합을 이용한 모형비교의 결과를 나열하였다.

The Comparative Study for Software Reliability Models Based on NHPP

Kang-Hyun Kan^{*} · Hee-Cheul Kim^{**} · Byoung-soo Lee^{***}

ABSTRACT

This paper presents a stochastic model for the software failure phenomenon based on a nonhomogeneous Poisson process (NHPP). The failure process is analyzed to develop a suitable mean value function for the NHPP; expressions are given for several performance measure. Actual software failure data are compared with generalized model by Goel dependent on the constant reflecting the quality of testing. The performance measures and parametric inferences of the new models, Rayleigh and Gumbel distributions, are discussed. The results of the new models are applied to real software failure data and compared with Goel-Okumoto and Yamada, Ohba and Osaki models. Tools of parameter inference was used method of the maximum likelihood estimate and the bisection algorithm for the computing nonlinear root. In this paper, using the sum of the squared errors, model selection was employed. The numerical example by NTDS data was illustrated.

키워드 : 소프트웨어 신뢰도 모형(Software Reliability Model), 비동질적인 포아송 프로세스(Nonhomogeneous Poisson Process), 레일리분포(Rayleigh Distribution), 겐벨 분포(Gumbel Distribution), 편차자승합(Sum of the Squared Error), 소프트웨어 신뢰도(Software Reliability)

1. 서 론

우리의 주변에는 복잡한 소프트웨어 시스템(system)들로 둘러 쌓여 있으며 이러한 시스템의 혜택을 받는 일이 커짐에 따라 소프트웨어 신뢰성의 역할은 점차 커지게 되었다.

그러나 시스템이 에러가 나면 그 원인을 찾아 필요할 경우 새로운 디자인을 개발하거나 새로운 기술을 도입하게 된다. 따라서 시간이 지남에 따라 신뢰도의 증가가 기대되어 진다.

소프트웨어 테스트 단계에서 소프트웨어 에러수(number of error)와 에러간격시간에 의해 소프트웨어 에러현상을 수리적으로 모형화 하면 소프트웨어에 대한 평가를 쉽게

할 수 있으며 신뢰도 성장 모형에 의해 소프트웨어 에러수, 소프트웨어 에러발생간격시간, 소프트웨어 신뢰도 및 에러율 등의 신뢰성 평가측도들이 추정되어 예측할 수 있다.

수리가 가능한 시스템이나 소프트웨어 신뢰도에서 관측시간(0, t) 사이에 발견된 에러수 $N(t)$ 를 모형화하는데 비동질적 포아송 프로세스(NHPP; non-homogeneous poisson process)가 널리 사용하여 왔다[5]. 강도함수(intensity function) 혹은 고장 발생률(ROCOF; rate of occurrence of failure) $\lambda(t) = dE[N(t)]/dt$ 에 대한 단조(monotonic)함수로 흔히 가정한다. 예를 들어 동질적 포아송 과정(HPP; homogeneous poisson process), Musa-Okumoto과정, Weibull 과정 그리고 Cox-Lewis 과정은 각각 ROCOF를 상수, 시간에 관한 부분(fraction) 함수, 멱(power) 함수, 대수선형(log-linear) 함수를 가정하였다[3].

본 논문에서는 기존의 모형인 Goel 이 제시한 일반화 모형[2]과 Yamada, Ohba-Osaki 모형[1]을 재조명하고 이러한 모형과 연관되고 신뢰도 분포로 많이 사용되는 와이블 분

* 본 연구는 한국과학재단 지정 인천대학교 멀티미디어 연구센터의 지원에 의한 것입니다.

† 정 회 원 : 한림정보대학 전산정보처리과 교수

†† 정 회 원 : 송호대학 컴퓨터정보처리과 교수

††† 총신회원 : 인천대학교 컴퓨터공학과 교수

논문접수 : 2000년 10월 31일, 심사완료 : 2001년 6월 18일

포의 특수형태인 레일리(Rayleigh)분포와 겐벨(Gumbel)분포 [5]를 이용한 모형을 제시하고, 또 효율적 모형을 위한 모형 선택으로 편차자승합(sum of the squared errors ; SSE)을 이용하여 비교하고자 한다.

본 논문의 구성의 2장에서는 관련 연구로서 NHPP 모형과 소프트웨어 신뢰도에 대한 일반적 내용을 요약하였고 3장에서는 이들 모형에 대한 최우추정법을 이용한 모수 추정에 대하여 서술하였고 4장에서는 실측자료인 NTDS 자료를 이용하여 각 모형에 대한 모수 추정 및 모형 비교에 대해서 설명하였다. 마지막으로 5장에서는 결론 및 향후 연구방향을 제시하였다.

2. 관련 연구

2.1 NHPP 모형

어떤 시간 t 까지에 발견된 총 에러수를 나타내는 계수과정 $\{N(t), t \geq 0\}$ 을 고려하자. 이 확률 과정에 대하여 단위 시간당 발견되는 에러수를 나타내는 강도함수 $\lambda(t)$ 를 가지는 비동질 포아송 프로세스(NHPP)를 따른다면 테스트 공정 혹은 운용단계에 있어서 에러 발견사상에 대하여 다음의 가정을 만족함이 알려져 있다.

- (가정1) $N(0) = 0$. 즉, 테스트 시각 $t=0$ 에서는 에러가 발견되지 않는다.
- (가정2) $\{N(t), t \geq 0\}$ 은 독립증분을 갖는다. 즉, 서로 다른 테스트 구간(시간)에서 발견되는 총 에러수는 통계적으로 독립이다.
- (가정3) $\Pr\{N(t + \Delta t) - N(t) = 1\} = \lambda(t)\Delta t + o(\Delta t)$. 즉, 임의의 극소 구간(시간) Δt 에서 한 개의 에러가 발견될 확률은 강도함수 $\lambda(t)$ 에 비례한다.
- (가정4) $\Pr\{N(t + \Delta t) - N(t) \geq 2\} = o(\Delta t)$. 즉, 임의의 극소 테스트 구간(시간) Δt 에서 두개 이상의 에러가 발견될 확률은 무시해도 좋을 만큼 적다.

위의 가정하에서 $N(t)$ 를 시간 $(0, t]$ 사이에 발견된 에러수라 정의하면 $N(t)$ 는 평균값 함수(mean value function) $m(t) = E[N(t)]$ 인 NHPP에 의해 다음과 같이 모형화 될 수 있다[6].

$$P(N(t) = n) = \frac{m(t)^n}{n!} e^{-m(t)} \quad n = 0, 1, 2, \dots \quad (2.1)$$

단,
$$m(t) = \int_0^t \lambda(x) dx. \quad (2.2)$$

$m(t)$ 가 t 에 대한 비감소함수(nondecreasing function) 추세를 가진 미분가능함수이면 강도함수(intensity function)는 $\lambda(t) = m'(t)$ 가 됨이 알려져 있다. $\lambda(t)$ 가 상수($m(t)$ 가 선형(linear) 추세이면 동질적 포아송 과정이고, t 에 대

한 함수형태이면 NHPP가 된다.

초기의 에러 총수를 N , 시점 t 에 있어서 에러 잔존수(number of errors remaining)를 $n(t)$ 라 하면, 시점 t 에서의 잔존 확률(신뢰도)은

$$R(t) = \frac{n(t)}{N} \quad (2.3)$$

가 되고 시점 t 까지의 에러 누적 확률 $F(t) = 1 - R(t)$ 가 된다. 양변을 미분하면

$$f(t) = -\frac{d}{dt} R(t) \quad (2.4)$$

이므로 강도함수 $\lambda(t) = -\frac{f(t)}{R(t)}$ 이되고 이식을 식 (2.4)과 관련하면 다음과 같이 나타낼 수 있다.

$$\lambda(t) = \frac{-\frac{d}{dt} R(t)}{R(t)} = \frac{-R'(t)}{R(t)} \quad (2.5)$$

식 (2.5)의 양변을 적분하면, $-\int_0^t \lambda(t) dt = \int_0^t \frac{R'(t)}{R(t)} dt$ 이고, 식 (2.2)과 관련하면 $-m(t) = \ln R(t)$ 가 되고, 양변에 exp를 취하면 신뢰도는 다음과 같다.

$$R(t) = \exp[-m(t)] \quad (2.6)$$

2.2 소프트웨어 신뢰도

n 번째까지 고장시점이 관찰된 고장 절단 모형일 경우 x_i 를 에러 발생시간 간격으로 하고, $s_n = \sum_{i=1}^n x_i$ 라 한다면 데이터 집합 D_s 은 $\{s_1, s_2, \dots, s_n\}$ 으로 구성된다.

우도함수를 구하기 위하여 $1 - F(t) = R(t) = \exp[-m(t)]$ 을 이용하여 양변을 미분하면 $-f(t) = -m'(t)\exp[-m(t)]$ 가 되고, 밀도함수는 $f(t) = \lambda(t)\exp[-m(t)]$ 으로 표현이 가능하다.

따라서 s_1 에서부터 s_n 까지의 결합확률밀도함수는 다음과 같다[5].

$$L_{s_n} = \left[\prod_{k=1}^n \lambda(s_k) \right] \exp[-m(s_n)] \quad (2.7)$$

테스트에 의해 최종적으로 발견되는 총기대 에러수 즉, 테스트 개시전에 소프트웨어 내에 잠재하는 총기대 에러수를 β_0 라고 할 때 일반적으로는 $m(\infty) = \beta_0$ 인 것으로부터

$$\lim_{t \rightarrow \infty} \Pr(N(t) = n) = \frac{\beta_0^n}{n!} e^{-\beta_0} \quad (n = 0, 1, 2, \dots)$$

로 되며 $N(t)$ 의 극한분포는 평균치가 β_0 인 포아송 분포에 따르는 것을 알 수 있다. 이러한 사실로부터 NHPP에

기초한 소프트웨어 신뢰도 성장 모델에 관한 정량적인 신뢰성 평가척도를 도출할 수 있다.

테스트 시각 t 에 있어서 소프트웨어 내의 기대잔존 에러수는 확률변수 $\{N(\infty) - N(t)\}$ 의 기대치를 고려하는 것에 의해

$$N_r(t) = E[N(\infty) - N(t)] = \beta_0 - m(t) \quad (2.8)$$

로 된다. 또 테스트 시각 t 에서 소프트웨어 고장이 일어난다고 하는 조건하에서 신뢰구간 $(t, t+x)$ 에 있어서 소프트웨어 에러가 일어나지 않을 조건부 확률은 $R(x | t) = F(t+x) - F(t)$ 로 놓고 정리하면,

$$R(x | t) = \exp[-\{m(t+x) - m(t)\}] \quad (t \geq 0, x \geq 0) \quad (2.9)$$

로 되며 소프트웨어 신뢰도(software reliability)라고 부른다.

3. 최우추정법을 이용한 모수추정

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

$$L_{NHPP} = \left[\prod_{k=1}^n \lambda(s_k) \right] \exp[-m(s_n)] \quad (3.1)$$

단, 강도함수 $\lambda(t) = m'(t)$.

따라서 식 (3.1)에서 최우추정법을 사용하여 NHPP모형들의 모수 추정을 할 수 있고, 각 모형에서 모수인 β_0 는 테스트 개시 전에 소프트웨어 내에 잠재하는 총 기대 에러 수를 나타내고, β_1 은 잔존에러 1개당 에러 발견율을 의미한다.

3.1 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. t 을 최종 고장시점 s_n 으로 대치하고(d 는 테스트 반영 상수) 오류 발견률 $\lambda(t) = \beta_0 \beta_1 d t^{d-1} e^{-\beta_1 t^d}$ 를 이용하면 우도함수는 식 (3.1)에 의해

$$L_{GC}(\beta_0, \beta_1 | s_1, s_2, \dots, s_n) = \left(\prod_{k=1}^n \beta_0 \beta_1 d s_k^{d-1} e^{-\beta_1 s_k^d} \right) \exp[-\beta_0(1 - e^{-\beta_1 s_n^d})] \quad (3.2)$$

가 되고, 로그우도함수를 구하면,

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

으로 표현된다.

따라서 $s = (s_1, s_2, \dots, s_n)$ 가 주어질 때의 모수 β_0 와 β_1 에 관한 편미분식은 다음과 같이 유도할 수 있다.

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

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

따라서 각 모수에 대한 최우추정량은 $\hat{\beta}_0$ 와 $\hat{\beta}_1$ 는 다음과 같은 형태가 된다.

$$\frac{n}{\hat{\beta}_0} = 1 - \exp(-\hat{\beta}_1 s_n^d) \quad (3.4)$$

$$\frac{n}{\hat{\beta}_1} = \sum_{k=1}^n s_k^d + \hat{\beta}_0 s_n^d \exp(-\hat{\beta}_1 s_n^d) \quad (3.5)$$

식 (3.4) 과 식 (3.5)을 수치적으로 풀어 두 모수를 구할 수 있고 신뢰도는 식 (2.9)에 의해 다음과 같이 나타낼 수 있다.

$$\hat{R}(x | t) = \exp[-\hat{\beta}_0 \{ \exp(-\hat{\beta}_1 t^d) - \exp(-\hat{\beta}_1(t+x)^d) \}] \quad (3.6)$$

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 $\lambda(t) = \beta_0 \beta_1^2 t e^{-\beta_1 t}$ 가 되고, 우도함수는 식 (3.1)과 관련하면 다음과 같다.

$$L_{YOO}(\beta_0, \beta_1 | s_1, s_2, \dots, s_n) = \left(\prod_{k=1}^n \beta_0 \beta_1^2 s_k e^{-\beta_1 s_k} \right) \cdot \exp[-\beta_0[1 - (1 + \beta_1 s_n)e^{-\beta_1 s_n}]] \quad (3.7)$$

따라서 로그우도함수를 구하면 다음과 같이 표현된다.

$$\ln L = n \ln \beta_0 + 2n \ln \beta_1 + \sum_{k=1}^n \ln s_k - \beta_1 \sum_{k=1}^n s_k - \beta_0[1 - (1 + \beta_1 s_n)e^{-\beta_1 s_n}] \quad (3.8)$$

데이터 $s = (s_1, s_2, \dots, s_n)$ 가 주어질 때의 모수 β_0 와 β_1 에 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.

$$\frac{n}{\hat{\beta}_0} = 1 - \exp(-\hat{\beta}_1 s_n) - \hat{\beta}_1 s_n \exp(-\hat{\beta}_1 s_n) \quad (3.9)$$

$$\frac{2n}{\hat{\beta}_1} = \sum_{k=1}^n s_k + \hat{\beta}_0 \hat{\beta}_1 s_n^2 \exp(-\hat{\beta}_1 s_n) \quad (3.10)$$

식 (3.9)와 식 (3.10)을 수치적으로 풀어 두 모수를 구할 수 있고 신뢰도는 식 (2.9)에 의해 다음과 같다.

$$\widehat{R}(x|t) = \exp[-\widehat{\beta}_0 \{(1 + \widehat{\beta}_1 t) \exp(-\widehat{\beta}_1 t) - (1 + \widehat{\beta}_1(t+x)) \exp(-\widehat{\beta}_1(t+x))\}] \quad (3.11)$$

3.3 제안모형

Goel-Okumoto[3]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 일반화하면 다음과 같이 표현할 수 있다.

$$m(t) = \beta_0 \cdot F(t) \quad (\text{즉, } \lambda(t) = \beta_0 \cdot f(t))$$

여기서, $F(t)$ 는 분포함수이고, $f(t)$ 는 확률밀도 함수를 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 관련하여 다루지 않고 신뢰도 분포로 많이 사용되는 와이블 분포의 특수형태인 레일리(Rayleigh)분포와 겐벨(Gumbel)분포[5]를 이용하여 다음과 같은 모형을 제시하고자 한다.

3.3.1 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 (2.2)에 의해 $\lambda(t) = 2\beta_0 \beta_1 t \cdot \exp(-\beta_1 t^2)$ 가 되고, 우도함수는 식 (3.1)과 관련하면 다음과 같다.

$$L_{Ray}(\beta_0, \beta_1 | s_1, s_2, \dots, s_n) = \left\{ \prod_{k=1}^n 2\beta_0 \beta_1 s_k \exp(-\beta_1 s_k^2) \right\} \cdot \exp[-\beta_0(1 - e^{-\beta_1 s_n^2})] \quad (3.12)$$

따라서 로그우도함수를 구하면 다음과 같이 표현된다.

$$\ln L = n \ln 2 + n \ln \beta_0 + n \ln \beta_1 + \sum_{k=1}^n \ln s_k - \beta_1 \sum_{k=1}^n s_k^2 - \beta_0(1 - e^{-\beta_1 s_n^2}) \quad (3.13)$$

데이터 $s = (s_1, s_2, \dots, s_n)$ 가 주어질 때의 모수 β_0 와 β_1 에 관한 편미분식은 다음과 같은 형태로 유도된다.

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

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

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

$$\frac{n}{\widehat{\beta}_0} = 1 - \exp(-\widehat{\beta}_1 s_n^2) \quad (3.14)$$

$$\frac{n}{\widehat{\beta}_1} = \sum_{k=1}^n s_k^2 + \widehat{\beta}_0 \widehat{\beta}_1 s_n^2 \exp(-\widehat{\beta}_1 s_n^2) \quad (3.15)$$

식 (3.14)와 식 (3.15)을 수치적으로 풀어 두 모수를 구할 수 있고 신뢰도는 식 (2.9)에 의해 다음과 같다.

$$\widehat{R}(x|t) = \exp[-\widehat{\beta}_0 \{\exp(-\widehat{\beta}_1 t^2) - \exp(-\widehat{\beta}_1(t+x)^2)\}] \quad (3.16)$$

3.3.2 겐벨분포 모형

겐벨분포의 경우의 평균값 함수는 $m(t) = \beta_0 [1 - \exp(-e^{-\beta_1 t})]$ ($\beta_0 > 0, \beta_1 > 0$)으로 표현할 수 있고 식 (2.2)에 의해 $\lambda(t) = \beta_0 \beta_1 e^{-\beta_1 t} \exp(-e^{-\beta_1 t})$ 가 되고, 우도함수는 식 (3.1)과 관련하면 다음과 같다.

$$L_{Gum}(\beta_0, \beta_1 | s_1, s_2, \dots, s_n) = \left\{ \prod_{k=1}^n \beta_0 \beta_1 e^{-\beta_1 s_k} \exp(-e^{-\beta_1 s_k}) \right\} \exp[-\beta_0(1 - \exp(e^{-\beta_1 s_n}))]$$

따라서 로그우도함수를 구하면 다음과 같이 표현된다.

$$\ln L = n \ln \beta_0 + n \ln \beta_1 + \beta_1 \sum_{k=1}^n s_k - \sum_{k=1}^n e^{-\beta_1 s_k} - \beta_0(1 - \exp(e^{-\beta_1 s_n})) \quad (3.17)$$

데이터 $s = (s_1, s_2, \dots, s_n)$ 가 주어질 때의 모수 β_0 와 β_1 에 관한 편미분식은 다음과 같은 형태로 유도된다.

$$\frac{\partial \ln L}{\partial \beta_0} = \frac{n}{\beta_0} - 1 + \exp(e^{-\beta_1 s_n}) = 0 \quad (3.18)$$

$$\frac{\partial \ln L}{\partial \beta_1} = \frac{n}{\beta_1} + \sum_{k=1}^n s_k - \beta_1 \left(\sum_{k=1}^n s_k e^{-s_k} \right) - \exp[\beta_1 s_n(1 + e^{-s_n})] = 0 \quad (3.19)$$

식 (3.18)와 식 (3.19)을 수치적으로 풀어 두 모수를 구할 수 있고 신뢰도는 식 (2.9)에 의해 다음과 같다.

$$\widehat{R}(x|t) = \exp[-\beta_0\{\exp(e^{-t\beta_1}) - \exp(e^{-(t+x)\beta_1})\}] \quad (3.20)$$

4. 수치적인 예

고장간격 시간 x_i 에 대한 자료는 실제 관측 데이터(NTDS 자료, [4])를 이용하여 모수를 추정하고 신뢰도를 구하고자 한다. 관측 데이터는 최종고장시간 $s_{26}=250$ ($s_0=0$)이고 각 고장 간격시간에 대한 자료는 <표 1>에 요약하였다.

<표 1> NTDS 데이터($x_i = s_i - s_{i-1}$)

고장번호 (i)	고장간격 시간 (x_i)	누적고장 시간 (s_i)	고장번호 (i)	고장간격 시간 (x_i)	누적고장 시간 (s_i)
1	9	9	14	9	87
2	12	21	15	4	91
3	11	32	16	1	92
4	4	36	17	3	95
5	7	43	18	3	98
6	2	45	19	6	104
7	5	50	20	1	105
8	8	58	21	11	116
9	5	63	22	33	149
10	7	70	23	7	156
11	1	71	24	91	247
12	6	77	25	2	249
13	1	78	26	1	250

각 모형에 대한 모수 추정방법은 최우추정법을 사용하고 근을 구하는 반복방법은 일반적인 수치해석적 방법인 이분법(bisection)으로 추정치를 구하였다. β_0 와 β_1 의 근을 구하는 과정에서 S-PLUS(Version 3.2) 소프트웨어를 사용하였으며 β_0 의 초기치는 20과 40을 β_1 는 0.00001과 0.05를 주고 100번 반복을 수행한 결과를 제시하였다. Yamada-Ohba-Osaki 모형(YOO)과 Goel의 일반화(generalized) 모형(GG)에서 $d=1$ 인 경우인 Goel-Okumoto 모형(GO), 본 논문에서 제안된 레일리(Rayleigh)분포모형(RM), 겐벨(Gumbel)분포모형(GM)에 대한 β_0 와 β_1 의 추정치는 <표 2>와 같다.

<표 2> 각 모형에 대한 모수 추정치

모형	추정치	$\widehat{\beta}_0$	$\widehat{\beta}_1$
GO		33.99	0.00057902
YOO		27.54	0.01854490
RM		27.31	0.00043456
GM		32.72	0.00017851

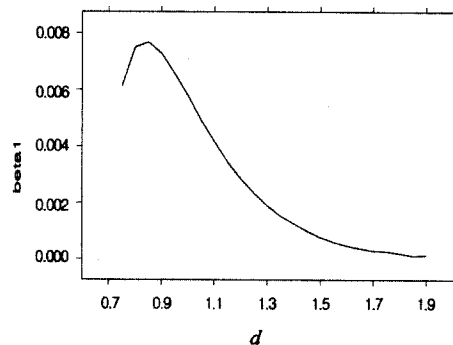
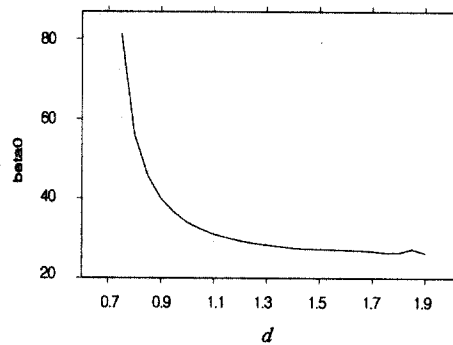
<표 2>에서 Hossain 과 Dahiya[4]는 같은 NTDS 자료

를 가지고 최우추정법을 사용해 GO 모형에 대한 추정치 $\widehat{\beta}_0$ 는 33.99, $\widehat{\beta}_1$ 는 0.000579를 추정한 바 있는데 본 연구의 추정치와 거의 일치한다.

실측 자료인 NTDS 자료를 테스트 반영상수(d) 따른 GG 모형의 특성을 알아보기 위해 테스트 반영상수 d 가 0.75에서 1.90까지 변화시키면서 모수를 추정하면 다음 <표 3>과 같다. 이 표에서 d 값이 증가함에 따라 β_0 의 추정치는 감소 패턴을 보이고 β_1 의 추정치는 증가하다가 급격히 감소하는 추세를 보인다. 이에 대한 요약 그림은 (그림 1)에 나타내었다. 즉, 테스트 개시 전에 소프트웨어 내에 잠재하는 총 기대 에러 수(β_0)는 d 값이 커짐에 따라 감소추세를 보이고 잔존에러 1개당 에러 발견율(β_1)은 약 0.8시점에서는 증가하다가 0.95근방에서 감소하는 추세가 있음을 의미한다.

<표 3> GG 모형에서 d 값에 따른 모수의 추정치

d	$\widehat{\beta}_1$	$\widehat{\beta}_0$	d	$\widehat{\beta}_1$	$\widehat{\beta}_0$
0.75	0.00612421	81.3577690	1.35	0.001534370	27.9778353
0.80	0.00750771	56.1332582	1.40	0.001265300	27.5472069
0.85	0.00768966	45.7634087	1.45	0.000996794	27.3772309
0.90	0.00729057	40.0108445	1.50	0.000765710	27.3245197
0.95	0.00656777	36.5014861	1.55	0.000598272	27.2051822
1.00	0.00579027	33.9932191	1.60	0.000468178	27.0882697
1.05	0.00492248	32.3997769	1.65	0.000368042	26.9647514
1.10	0.00418561	31.0417683	1.70	0.000290574	26.8389909
1.15	0.00345686	30.1726040	1.75	0.000266748	26.3987557
1.20	0.00286717	29.3792831	1.80	0.000197160	26.4452319
1.25	0.00234191	28.8083580	1.85	0.000109593	27.3737062
1.30	0.00190012	28.3519581	1.90	0.000124762	26.2952881



(그림 1) GG 모형에서 d 값에 따른 $\widehat{\beta}_0$ 와 $\widehat{\beta}_1$ 값의 변화

효율적 모형의 비교를 위해 고장번호(i)는 시점 s_i 에 관찰된 실제고장의 수 ($n_i(s_i)$)이고 $\widehat{m}_i(s_i)$ 는 시점 s_i 에서 추정된 고장의 누적수를 나타내므로 실제 고장수와 추정된 고장수의 편차자승합을 계산해서 작은 모형이 효율적인 모형이라고 간주 할 수 있다. 즉, 다음과 같이 나타낼 수 있다.

$$C_{mv} = \sum_{i=1}^n (n_i(s_i) - \widehat{m}_i(s_i))^2 \quad (4.1)$$

단, n_i 는사이에 관찰된 오류의 수이고 $\widehat{m}_i(s_i)$ 는 (추우추정법으로 구한) 평균값 함수의 추정값을 의미한다. 예를 들어 GG모형에 대한 실제 고장수와 추정된 고장수의 편차자승합을 계산하는 요약 통계량은 <표 4>와 같다.

<표 4> GG모형에 대한 평균값함수의 추정값과 편차자승합

고장번호 (i)	고장간격 시간 (x_i)	누적고장 시간 (s_i)	실제고장수 ($n_i(s_i)$)	평균값함수의 추정값 ($\widehat{m}_i(s_i)$)	편차자승합
1	9	9	1	1.72610292	0.52723
2	12	21	2	3.89200831	3.57970
3	11	32	3	5.74946254	7.55954
4	4	36	4	6.39610111	5.74130
5	7	43	5	7.49229891	6.21155
6	2	45	6	7.79742371	3.23073
7	5	50	7	8.54495413	2.38688
8	8	58	8	9.69688676	2.87942
9	5	63	9	10.3902136	1.93269
10	7	70	10	11.3277594	1.76295
11	1	71	11	11.4586193	0.21033
12	6	77	12	12.2280646	0.05201
13	1	78	13	12.3537265	0.41767
14	9	87	14	13.4525336	0.29972
15	4	91	15	13.9228110	1.16034
16	1	92	16	14.0386883	3.84674
17	3	95	17	14.3823214	6.85224
18	3	98	18	14.7200969	10.7582
19	6	104	19	15.3781208	13.1180
20	1	105	20	15.4856959	20.3799
21	11	116	21	16.6276450	19.1175
22	33	149	22	19.6480844	5.53151
23	7	156	23	20.2178941	7.74011
24	91	247	24	25.8599384	3.45937
25	2	249	25	25.9535829	0.90932
26	1	250	26	25.9598990	0.00161
편차자승합	(C_{mv})				129.66647

유사한 방법으로 편차자승합을 계산한 결과는 <표 5>에 요약 되었다.

<표 5>에서는 식 (4.1)을 이용한 편차자승합을 나타내었다. 이 표에서 GG 모형은 테스트 반영상수 d 가 약 1.1와 1.5사이가 기본적 모형인 GO 모형($d=1$)에 비해서 NTDS 데이터에서는 효율적인 모형으로 간주할 수 있다.

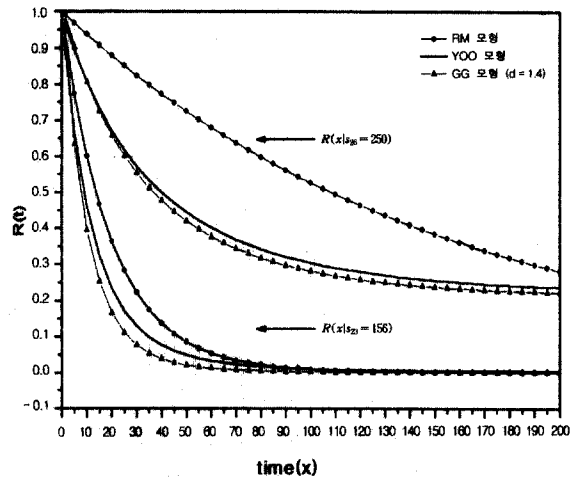
따라서 본 논문에서 제시한 RM 모형이 편차자승합(38.437532)이 다른 모형에 비해 작으므로 보다 효율적인 모형

<표 5> 편차자승합에 의한 모형비교

모형	(C_{mv})	모형	(C_{mv})
RM 모형	38.437532	GM 모형	84.817954
YOO 모형	46.94658	GG 모형($d=1.4$)	79.12883
GG 모형($d=0.8$)	176.7161	GG 모형($d=1.5$)	104.7357
GG 모형($d=0.9$)	151.8326	GG 모형($d=1.6$)	136.7901
GO 모형($d=1.0$)	129.66647	GG 모형($d=1.7$)	165.5301
GG 모형($d=1.1$)	111.3138	GG 모형($d=1.8$)	171.5643
GG 모형($d=1.2$)	97.79414	GG 모형($d=1.9$)	174.6575
GG 모형($d=1.3$)	89.82807		

으로 간주 할 수 있다. 그러나 GM 모형은 YOO 모형이나 GO 모형($d=1.2, 1.3, 1.4, 1.5$)보다 비효율적인 모형으로 나타나고 있다.

그러므로 신뢰도 추정을 위하여 효율적인 모형인 RM 모형, GG 모형($d=1.4$), YOO 모형을 이용하여 시점 s_i 가 두시점(s_{23}, s_{26})에서 신뢰도를 추정하고자 한다. 즉, n 이 23이후의 신뢰도 $\widehat{R}(x|s_{23}=156)$ 와 n 이 26이후의 신뢰도 $\widehat{R}(x|s_{26}=250)$ 를 계산하여 비교하고자 한다.



(그림 2) 각 모형별 신뢰도

(그림 2)에서는 임무시간(x ; mission time)에 따른 신뢰도함수를 그린 것이다. 이 그림에서 전체적으로 신뢰도 $(R(x|t))$ 는 비증가 추세를 보이고 있고 우리가 기대하였듯이 n 이 26이후의 신뢰도 $\widehat{R}(x|s_{26}=250)$ 가 20이후의 신뢰도 $\widehat{R}(x|s_{23}=156)$ 보다는 신뢰성장(신뢰도 높음)이 이루어졌음을 알수 있다.

5. 결 론

소프트웨어의 신뢰성은 개발의 최종단계에 있는 테스트 공정이나 실제 사용 단계에 있어서 소프트웨어 내에 존재하는 에러 수나 소프트웨어의 고장발생시간에 의해

효과적 평가를 할 수 있는 것으로 그 평가 기술이 중요하게 된다. 소프트웨어 개발의 테스트 공정이나 실제 사용단계에 있어서 에러 발생상황이나 소프트웨어 고장 발생현상을 수리적 모형화가 가능하다면 평가를 할 수 있다. 테스트의 개발 상황의 파악, 테스트에 의해 미 발생되었던 고장에 대한 보수 코스트의 예측 등 구체적인 소프트웨어 개발의 보수관리문제에도 적용 가능하다. 따라서 테스트 시간 혹은 실행시간과 발생된 고장 수나 소프트웨어 고장의 발생시간과의 관계를 소프트웨어 신뢰도 성장과정이라고 볼 수 있다.

본 논문에서는 비동질 포아송 프로세스(NHPP)에 기초한 소프트웨어 에러 현상에 대한 확률 모형을 고려했다. 고장 패턴은 NHPP에 대한 강도함수와 평균값 함수로서 나타낼 수 있다. 따라서 본 논문에서는 기존의 모형인 Goel이 제시한 일반화모형과 Yamada, Ohba-Osaki 모형을 재조명하고 이러한 모형과 연관되고 신뢰도 분포로 많이 사용되는 와이블 분포의 특수형태인 레일리(Rayleigh)분포와 겐벨(Gumbel)분포를 이용한 모형을 제시하고 또, 효율적 모형을 위한 모형선택으로서 편차자승합을 이용하여 비교하였다. 모수의 추정을 위해서 최우추정법(MLE)과 일반적인 수치해석적 방법인 이분법을 이용하였다. 수치적인 예에서는 실측자료인 NTDS 자료를 이용하여 모수 및 신뢰도를 추정하였고 상대오차 합을 이용한 모형비교의 결과를 나열하였다.

그 결과 Goel 일반화(generalized) 모형은 테스트 반영상수 d 가 약 1.1와 1.5사이가 기본적 모형인 Goel-Okumoto 모형($d=1$)에 비해서 NTDS 데이터에서는 효율적인 모형으로 간주할 수 있다. 그리고 본 논문에서 제시한 레일리분포 모형은 편차자승합이 다른 모형에 비해 작으므로 보다 효율적인 모형이 됨을 알 수 있었다.

신뢰도 추정을 위하여 실제고장수가 23이후의 신뢰도와 26이후의 신뢰도를 계산한 결과 임무시간(x ; mission time)에 따른 신뢰도함수는 전체적으로 신뢰도가 비증가 추세를 보이고 있고 n 이 26이후의 신뢰도가 23이후의 신뢰도 보다는 우리가 기대한 것처럼 높음(신뢰성장)을 알 수 있다.

향후 추가 정보를 이용한 베이지안 분석과 신뢰도 성장 모형에 대한 분포 이론 및 응용에 관한 연구가 기대된다.

참 고 문 헌

[1] Charnes, A., Frome, E. L. and Yu, P. L. "The Equivalence of Generalized Least Squares and Maximum Likelihood Estimates in the Exponential Family," *Journal of the American*

Statistical Association, Vol.7, No.353, pp.169-171, 1976.

- [2] Goel, A. L. "Software Reliability Model : Assumption, Limitations, and Applicability," *IEEE Trans. Software Eng.*, Vol.SE-11, No.12, pp.1411-1423, 1985.
- [3] 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.
- [4] Hossain, S. A. and Dahiya, R. C. "Estimating the Parameters of a Non-homogeneous Poisson-Process Model for Software Reliability," *IEEE Trans. Rel.*, Vol.R-42, No.4, pp.604-612, 1993.
- [5] Johnson, N. L. and Kotz, S. "Continuous Univariate Distributions 1," A Wiley-Interscience publication, JOHN WILEY & SONS, pp.272-289, 1970.
- [6] Lawless, J. F. "Statistical Models and Methods for Lifetime Data," pp.494-500, 1981.
- [7] Pham, H. and Nordmann, L. and Zhang, X. "A General Imperfect-Software-Debugging Model with S-Shaped Fault-Detection Rate," *IEEE Trans. Rel.*, Vol.48, No.2, pp.169-175, 1999.
- [8] Okumoto, K. "A Statistical Method for Software Quality Control," *IEEE Trans. Software Eng.*, Vol.SE-11, No.12, pp. 424-431, 1985.
- [9] Ross, Sheldon M. "Introduction to Probability Model," Academic press, Inc. pp.277-281, 1997.
- [10] Venables, W. N., and Ripley, B. D. "Modern Applied Statistics with S-Plus," Springer-Verla, 1994.
- [11] 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.
- [12] Yamada, S., Ohtera, H. and Narihisa, H. "Software Reliability Growth Models with Testing-Effort," *IEEE Trans. Rel.*, Vol.R-35, No.1, pp.19-23, 1986.
- [13] Yamada, S. and Osaki, S. "Software Reliability Growth Modeling : Models and application," *IEEE Trans. Software Eng.*, Vol.SE-11, No.12, pp.1431-1437, 1985.

간 광 현

E-mail : khkan@sun.hallym-c.ac.kr

1995년~현재 한림대학교 전자계산학과
박사과정 수료.

1990년~현재 한림정보산업대학 전산정보
처리과 부교수.

관심분야 : 데이터베이스, 소프트웨어 공학,
멀티미디어



김 희 철

E-mail : khc@songho.ac.kr
1998년 동국대학교 대학원 통계학과 졸업
(이학박사)
2000~현재 송호대학 컴퓨터정보처리과
전임강사
관심분야 : 소프트웨어신뢰성공학, 웹프로
그래밍, 전산통계

이 병 수

E-mail : bsi@incheon.ac.kr
1999년 경기대학교 대학원 전자계산학과
졸업(이학박사)
1981년~현재 시립인천대학교
컴퓨터정보통신학부 교수
관심분야 : 웹프로그래밍, 소프트웨어공학,
의사결정지원시스템