

지수화 지수 분포에 의존한 NHPP 소프트웨어 신뢰성장 모형에 관한 연구

김희철*

The Study for NHPP Software Reliability Growth Model based on Exponentiated Exponential Distribution

Kim Hee Cheul *

요약

유한고장수를 가진 비동질적인 포아송 과정에 기초한 모형들에서 잔존 결함 1개당 고장 발생률은 일반적으로 상수, 혹은 단조증가 및 단조 감소 추세를 가지고 있다. 본 논문에서는 기존의 소프트웨어 신뢰성 모형인 Goel-Okumoto 모형과 Yamada-Ohba-Osaki 모형을 재조명하고 최근에 Gupta 와 Kundu(2001)에 의해서 2 모수 감마분포나 와이불 분포의 대체모형으로서 여러 가지 수명자료를 분석함에 있어서 효율적 분포가 됨이 밝혀진 지수화된 지수 분포를 이용한 모형을 제안하였다. 고장 간격시간으로 구성된 자료를 이용한 모수추정 방법은 최우추정법과 일반적인 수치해석 방법인 이분법을 사용하여 모수 추정을 실시하고 효율적인 모형 선택은 편차자승합(SSE), AIC 통계량 및 콜모고로프 거리를 적용하여 모형들에 대한 효율성 입증방법을 설명하였다. 소프트웨어 고장 자료 분석에서는 NTDS 자료를 통하여 분석하였다. 이 자료들에서 기존의 모형과 지수화된 지수 분포 모형의 비교를 위하여 산술적 및 라플라스 검정, 편의 검정 등을 이용하였다.

Abstract

Finite failure NHPP models presented in the literature exhibit either constant, monotonic increasing or monotonic decreasing failure occurrence rates per fault. In this paper, Goel-Okumoto and Yamada-Ohba-Osaki model was reviewed, proposes the exponentiated exponential distribution reliability model, which made out efficiency substituted for gamma and Weibull model(2 parameter shape illustrated by Gupta and Kundu(2001) Algorithm to estimate the parameters used to maximum likelihood estimator and bisection method, model selection based on SSE, AIC statistics and Kolmogorov distance, for the sake of efficient method, was employed. Analysis of failure using NTDS data set for the sake of proposing shape parameter of the exponentiated exponential distribution was employed. This analysis of failure data compared with the exponentiated exponential distribution model and the existing model(using arithmetic and Laplace trend tests, bias tests) is presented.

▶ Keyword : 소프트웨어 신뢰도 모형(Software Reliability Model), 비동질적인 포아송 과정 (Nonhomogeneous Poisson Process), 소프트웨어 신뢰도(Software Reliability), 지수화된 지수분포(Exponentiated Exponential Distribution), 편의 검정(Bias Tests), 편차자승합(Sum of the Squared Errors), 콜모고로프 거리(Kolmogorov Distance)

• 제1저자 : 김희철
• 접수일 : 2006.09.20, 심사일 : 2006.10.16, 심사완료일 : 2006.11.18
* 남서울대학교 산업경영공학과 전임강사

I. 서론

소프트웨어 테스트 단계에서 소프트웨어 고장수(Number of failure)와 고장간격시간에 의해 소프트웨어 고장현상을 수리적으로 모형화 하면 소프트웨어에 대한 평가를 보다 쉽게 할 수 있으며 신뢰도 모형에 의해 소프트웨어 고장수, 소프트웨어 고장발생간격시간, 소프트웨어 신뢰도 및 고장률 등의 신뢰성 평가측도들이 추정되어 미래의 고장시간을 예측할 수도 있다.

소프트웨어 고장시간은 수명자료가 된다. 따라서 비음(Nonnegative)의 값을 가지기 때문에 이 분야에서는 주로 지수분포, 와이블분포, 감마분포 등 일반화 감마 분포가 많이 사용되어 왔다[1]. 그러나 이러한 감마분포 외에도 통계 자료 해석시 자료를 대수변환 이 후에 정규분포로 처리하면 되는 대수정규분포(Lognormal distribution), 이 분포와 유사한 대수로지스틱 분포(Loglogistic distribution)도 이 분야에 사용이 가능하대[2]. 또, 자유도(Degree of freedom; df)에 의존하는 카이제곱(χ^2 distribution) 분포도 적용이 가능하다. 이러한 카이제곱 분포는 변화가 일어날 때까지의 대기시간을 나타내는 데 많이 사용되는 감마분포(Gamma Distribution)의 특수한 경우로도 접근 할 수 있기 때문에 소프트웨어 신뢰성 수명분포로 설명 할 수 있다[3].

본 논문에서는 최근에 Gupta 와 Kundu[4,5]에 의해서 2 모수 감마분포나 와이블분포의 대체모형으로서 여러 가지 수명자료를 분석함에 있어서 효율적 분포가 됨이 밝혀져 있는 지수화된 지수 분포를 적용하고자 한다. 이러한 지수화된 지수 분포를 수명분포로 사용한 NHPP 모형에 대한 신뢰성 척도를 추정하고 이를 바탕으로 형상모수에 따른 모형의 효율성과 그 특성을 알아보하고자한다.

본 논문의 2장에서는 관련연구로서 유한 고장 NHPP 모형에 대하여 서술하였고 3장에서는 지수화된 지수 분포 신뢰성 모형 및 최우추정법을 이용한 모수추정에 대하여 설명하고 4장에서는 실제 고장자료를 이용하여 각 모형에 대한 모수추정 및 모형비교를 실시하였으며 마지막으로 5장에서 결론을 내릴 하였다.

II. 이론적 배경에 대한 고찰

신뢰도에서 관측시간 $(0, t]$ 사이에 발견된 고장수 $N(t)$ 를 모형화하는데 비동질적 포아송 과정(Nonhomogeneous poisson

process; NHPP)이 널리 사용하여 왔다. 이 과정(Process)에서 강도함수(Intensity function) 혹은 고장 발생률(Rate of occurrence of failure; ROCOF)

$\lambda(t) = dE[N(t)]/dt$ 은 t 에 대한 단조(Monotonic)함수로 흔히 가정 한다[3]. 이 범주에서 지금까지 알려진 모형들은 Goel-Okumoto 모형, Weibull 모형 그리고 Cox-Lewis 모형등이 있는데 이 모형들에 대한 강도함수는 각 각 시간에 의존한 함수, 멱(Power) 함수, 대수선형(Loglinear) 함수를 가정하였다[6,7].

NHPP 모형에서 평균값 함수 $m(t)$ (Mean value function)와 강도 함수 $\lambda(t)$ 는 다음과 같은 관계로 표현 할 수 있다.

$$m(t) = \int_0^t \lambda(s)ds, \quad \frac{dm(t)}{dt} = \lambda(t) \dots\dots\dots (2.1)$$

$N(t)$ 는 모수 $m(t)$ 를 가진 포아송 확률밀도함수(Probability density function)로 알려져 있다. 즉,

$$P\{N(t) = n\} = \frac{[m(t)]^n \cdot e^{-m(t)}}{n!}, \quad n = 0, 1, 2, \dots, \infty \dots\dots\dots (2.2)$$

이처럼 시간관련 모형(Time domain models)들은 NHPP에 의해서 확률 고장 과정으로 설명이 가능하다. 이러한 모형들은 고장 강도 함수 $\lambda(t)$ 가 다르게 표현됨으로서 평균값 함수 $m(t)$ 도 역시 다르게 나타난다. 이러한 NHPP 모형들은 유한 고장 모형과 무한 고장 범주로 분류한다[8]. 유한 고장(Finite failure) NHPP 모형들은 충분한 테스트 시간이 주어지면 결함들(Faults)의 기대값이 유한 값 ($\lim_{t \rightarrow \infty} m(t) = \theta < \infty$)을 가지고 반면에 무한 고장(Infinite failure) NHPP 모형들은 무한 값을 가진다고 가정 된다. 유한 고장 NHPP 모형에서 충분한 테스트 시간이 주어졌을 때 탐색되어 질 수 있는 결함의 기대값을 θ 라고 표현하고 $F(t)$ 를 분포함수라고 표현하면 유한 고장 NHPP모형의 평균값 함수는 다음과 같이 표현 할 수 있다[6].

$$m(t) = \theta F(t) \dots\dots\dots (2.3)$$

(2.3)식으로 부터 순간고장 강도함수(Instantaneous failure intensity) $\lambda(t)$ 는 다음과 같이 유도된다.

$$\lambda(t) = \theta F'(t) \dots\dots\dots (2.4)$$

(2.4)식을 다음과 같이 변형하여 표기 할 수도 있다.

$$\lambda(t) = [\theta - m(t)] \frac{F'(t)}{1-F(t)} = [\theta - m(t)] h(t) \dots\dots\dots (2.5)$$

단, $h(t) = \frac{F'(t)}{1-F(t)}$ 는 위험함수(Hazard function, 고장률 함수)으로서 소프트웨어 결함당 고장 발생률을 의미하고 $[\theta - m(t)]$ 은 t 시점에서 소프트웨어에 남아있는 결함들의 기대값을 나타낸다. $[\theta - m(t)]$ 의 값은 시점 t 에 대한 단조 비증가 함수(Monotonically nonincreasing function)가 된다. 즉, 시간이 지남에 따라 그 고장을 찾아 제거하는 디버깅 과정을 거치면서 제거되기 때문에 감소성을 가진다.

$\lambda(t)$ 는 $h(t)$ 의 속성에 따라 달라지며 그 추세는 상수나 증가, 혹은 감소(증가)하다가 증가(감소)하는 패턴을 가질 수도 있다.

시간 $(0, t]$ 까지 조사하기 위한 시간 절단(Time truncated) 모형은 n 번째 까지 고장시점 자료를

$$x_k = \sum_{i=1}^k t_i \quad (k=1,2,\dots,n; 0 \leq t_1 \leq t_2 \leq \dots \leq t_n) \dots\dots\dots (2.6)$$

이라고 하면 데이터 집합 D_t 는 $\{n, x_1, x_2, \dots, x_n; t\}$ 와 같이 구성된다. n 번째까지 고장시점이 관찰된 고장 절단 모형 일 경우에 데이터 집합 D_{x_n} 은 $\{x_1, x_2, \dots, x_n\}$ 으로 구성되며 이 시간 절단 모형에서의 θ 를 모수공간이라고 표시하면 우도함수는 다음과 같이 알려져 있다[6,8].

$$L_{NHPP}(\theta | D_{x_n}) = \left(\prod_{i=1}^n \lambda(x_i) \right) \exp(-m(x_n)) \dots\dots\dots (2.7)$$

이 분야의 기본적 모형인 Goel-Okumoto 모형은 $h(t)$ 가 정수 패턴을 가짐으로서 시점 t 에 독립이고 잘 알려진 Yamada, Ohba-Osaki 모형은 단조 비감소 패턴을 가진다 [2].

NHPP 모형에서 테스트 시점 x_n 에서 소프트웨어 고장이 일어난다고 하는 가정하에서 신뢰구간 $(x_n, x_n + t]$ (단, t 는 임무시간(Mission time))사이에서 소프트웨어의 고장이 일어나지 않을 확률인 신뢰도(Reliability)

$\hat{R}(t | x_n)$ 는 다음과 같이 됨이 알려져 있다[9].

$$\begin{aligned} \hat{R}(t | x_n) &= e^{-\int_{x_n}^{x_n+t} \lambda(\tau) d\tau} \dots\dots\dots (2.8) \\ &= \exp[-\{m(t+x_n) - m(x_n)\}] \end{aligned}$$

2.1 지수화된 지수분포

이 장에서 지수화된 지수 분포(Exponentiated exponential distribution; EED)[4]에 대하여 설명하고자 한다.

원래 지수화된 지수분포는 Gompertz-Verhulst 분포 함수

$$G(t) = (1 - \rho e^{-t\tau})^\alpha, \quad (t > \frac{1}{\tau} \ln \rho, \rho, \alpha, \tau > 0) \dots\dots\dots (2.9)$$

에서 $\rho=1$ 인 경우를 의미하고 일반화된 지수 분포로 알려져 있다. 이러한 Gompertz-Verhulst 분포함수는 초기에 인간 사망률 도표(Human mortality table)와 모집단 인구성장(Growth)를 설명하기 위하여 Gompertz와 Verhulst가 정의한 분포함수이다. 따라서 지수화된 2모수 지수 분포는 (2.9)과 관련하여 다음과 같이 정의 된다.

$$F(t | \alpha, \tau) = (1 - e^{-\tau t})^\alpha \quad (t > 0, \alpha, \tau > 0) \dots\dots\dots (2.10)$$

(2.10) 식에서 $\alpha(>0)$ 와 $\tau(>0)$ 는 각각 형상모수(Shape parameter) 와 척도모수(Scale parameter)를 의미하고 확률밀도함수는 다음과 같이 유도 된다.

$$f(t | \alpha, \tau) = \alpha \tau (1 - e^{-\tau t})^{\alpha-1} e^{-\tau t} \quad (t > 0, \alpha, \tau > 0) \dots\dots\dots (2.11)$$

이러한 지수화된 지수 분포는 최근에 Gupta 와 Kundu[2, 5]에 의해서 2 모수 감마분포나 와이블 분포의 대체모형으로서 여러 가지 수명자료를 분석함에 있어서 효율적 분포가 됨이 밝혀 졌다.

또, 위험함수는(2.10)과 (2.11)식과 관련하여 다음과 같이 유도 된다.

$$h(t | \alpha, \tau) = \frac{f(t | \alpha, \tau)}{1 - F(t | \alpha, \tau)} = \frac{\alpha \tau (1 - e^{-\tau t})^{\alpha-1} e^{-\tau t}}{1 - (1 - e^{-\tau t})^\alpha} \dots\dots\dots (2.12)$$

위험함수의 형태는 <그림 1>에서 보여 주듯이 $\alpha < 1$ 인 경우에는 감소함수이고 $\alpha = 1$ 인 경우에는 상수, $\alpha > 1$ 인 경우에는 증가 위험 함수 형태가 된다.

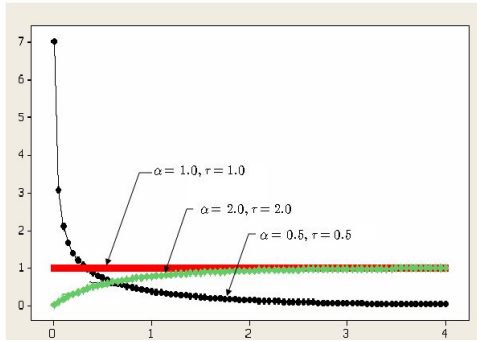


그림 1. 지수화된 지수 분포의 위험함수
Fig. 1 Hazard function of EED

2.2 기존의 신뢰성 모형

기존의 모형은 주로 수명 분포가 다음과 같은 감마분포의 분포함수를 사용하고 있다[1].

$$F_{Ga}(t | k, \beta) = 1 - e^{-\beta t} \left(\sum_{i=0}^{k-1} \frac{(\beta t)^i}{i!} \right) \quad (t, k > 0, \beta > 0)$$

..... (2.13)

(2.13)식에서 감마분포의 확률밀도함수는 다음과 같이 유도 할 수 있다[10].

$$f_{Ga}(t | k, \beta) = \frac{\beta^k}{\Gamma(k)} t^{k-1} e^{-\beta t} \quad (k, \beta > 0, t > 0)$$

..... (2.14)

따라서 (2.3)식과 (2.4)에 관련하여 소프트웨어 결함(Fault)당 고장(Failure) 발생률이 일정하거나 증가 혹은 감소하는 특징을 가지는 모형이 감마 모형이 되고 유한 고장 NHPP모형의 평균값 함수 $m(t)$ 와 강도함수 $\lambda(t)$ 는 다음과 같이 표현 할 수 있다.

$$m(t) = \theta F_{Ga}(t | k, \beta) = \theta \left[1 - e^{-\beta t} \left(\sum_{i=0}^{k-1} \frac{(\beta t)^i}{i!} \right) \right]$$

..... (2.15)

$$\lambda(t) = \theta f_{Ga}(t | k, \beta) = \theta \frac{\beta^k}{\Gamma(k)} t^{k-1} e^{-\beta t}$$

..... (2.16)

(2.15)식의 분포함수가 폐쇄형(close form)의 형태이므로 기존의 모형은 주로 함수는 k 값이 양의 정수일 때만을 고려하고 있다.

이 분야에서 가장 기본적인 모형은 Goel-Okumoto 모

형[11] 이다. 이 모형은 결함당 고장발생 시간의 분포(수명 분포)를 지수분포를 가정하였다. 따라서 소프트웨어 결함(Fault)당 고장(Failure) 발생률이 일정한 형태를 가지며 (2.13)식과 (2.14)식에서 형상모수 $k=1$ 인 경우와 같이 된다. 이 모형을 개선한 모형이 잘 알려진 S 모형(Yamada, Ohba-Osaki 모형[12])이다. 따라서 이 모형은 수명 분포를 감마분포 중에서 형상모수 $k=2$ 인 분포를 말한다.

2.2.1 Goel-Okumoto모형

Goel-Okumoto 모형[7,11]에 대한 평균값 함수는 $m(t|\theta, \beta_1) = \theta(1 - e^{-\beta_1 t})$ ($\theta > 0, \beta_1 > 0$)이라고 알려져 있다.

시간 $(0, t]$ 까지 조사하기 위한 시간 절단(Time truncated)모형은 n 번째 까지 고장시점 자료를

$$x_k = \sum_{i=1}^k t_i \quad (k = 1, 2, \dots, n; 0 \leq t_1 \leq t_2 \leq \dots \leq t_n)$$

이라고 하면 데이터 집합 D_t 는 $\{n, x_1, x_2, \dots, x_n; t\}$ 와 같이 구성된다.

t 을 최종 고장시점 x_n 으로 대치하고 고장 발생률 $\lambda(t|\theta, \beta_1) = \theta \beta_1 e^{-\beta_1 t}$ 을 이용하면 유도함수는 다음과 같이 표현 할 수 있다.

$$L_{GO}(\theta, \beta_1 | D_n) = \left(\prod_{k=1}^n \theta \beta_1 e^{-\beta_1 x_k} \right) \exp[-\theta(1 - e^{-\beta_1 x_n})]$$

..... (2.17)

최우추정법(MLE)을 이용하기 위한 로그우도함수를 구하면,

$$\ln L_{GO} = n \ln \theta + n \ln \beta_1 - \beta_1 \sum_{k=1}^n x_k - \theta(1 - e^{-\beta_1 x_n})$$

..... (2.18)

으로 표현된다.

따라서 고장절단모형에서의 모수 θ 와 β_1 에 관한 편미분 식은 다음과 같이 유도할 수 있다.

$$\frac{\partial \ln L_{GO}}{\partial \theta} = \frac{n}{\theta} - 1 + e^{-\beta_1 x_n} = 0,$$

$$\frac{\partial \ln L_{GO}}{\partial \beta_1} = \frac{n}{\beta_1} - \sum_{k=1}^n x_k - \theta x_n e^{-\beta_1 x_n} = 0$$

각 모수에 대한 최우추정량 $\hat{\theta}_{MLE}$ 와 $\hat{\beta}_{1,MLE}$ 은 다음식

을 만족한다.

$$\frac{n}{\theta} = 1 - \exp(-\hat{\beta}_1 x_n), \dots\dots\dots (2.19)$$

$$\frac{n}{\hat{\beta}_1} = \sum_{k=1}^n x_k + \hat{\theta} x_n \exp(-\hat{\beta}_1 x_n) \dots\dots\dots (2.20)$$

(2.19)과 (2.20)식을 수치적으로 풀어 두 모수를 구할 수 있다.

2.2.2 Yamada-Ohba-Osaki 모형

Yamada-Ohba-Osaki 모형[7,12]에 대한 평균값 함수는 $m(t|\theta, \beta_2) = \theta [1 - (1 + \beta_2 t)e^{-\beta_2 t}]$ ($\theta > 0, \beta_2 > 0$) 이라고 알려져 있고 $\lambda(t|\theta, \beta_2) = \theta \beta_2^2 t e^{-\beta_2 t}$ 가 되고 우도함수는 다음과 같다.

$$L_{YOO}(\theta, \beta_2 | D_x) = \left(\prod_{k=1}^n \theta \beta_2^2 x_k e^{-\beta_2 x_k} \right) \dots\dots\dots (2.21)$$

$$\cdot \exp[-\theta [1 - (1 + \beta_2 x_n) e^{-\beta_2 x_n}]]$$

따라서 로그우도함수를 구하면 다음과 같이 표현된다.

$$\ln L_{YOO} = n \ln \theta + 2n \ln \beta_2 + \sum_{k=1}^n \ln x_k - \beta_2 \sum_{k=1}^n x_k \dots\dots\dots (2.22)$$

$$- \theta [1 - (1 + \beta_2 x_n) e^{-\beta_2 x_n}]$$

따라서 고장절단 모형에서의 모수 θ 와 β_2 에 관한 편미분식은 다음과 같은 형태로 유도된다.

$$\frac{\partial \ln L_{YOO}}{\partial \theta} = \frac{n}{\theta} - 1 + e^{-\beta_2 x_n} + \beta_2 x_n e^{-\beta_2 x_n} = 0,$$

$$\frac{\partial \ln L_{YOO}}{\partial \beta_2} = \frac{2n}{\beta_2} - \sum_{k=1}^n x_k - \theta \beta_2 x_n^2 e^{-\beta_2 x_n} = 0$$

따라서 각 모수에 대한 최우추정량은 $\hat{\theta}_{MLE}$ 와 $\hat{\beta}_{2MLE}$ 은 다음 식을 만족한다.

$$\frac{n}{\hat{\theta}} = 1 - \exp(-\hat{\beta}_2 x_n) - \hat{\beta}_2 x_n \exp(-\hat{\beta}_2 x_n) \dots\dots\dots (2.23)$$

$$\frac{2n}{\hat{\beta}_2} = \sum_{k=1}^n x_k + \hat{\theta} \hat{\beta}_2 x_n^2 \exp(-\hat{\beta}_2 x_n) \dots\dots\dots (2.24)$$

(2.23) 과 (2.24)식을 수치적으로 풀어 두 모수를 구할 수 있다.

III. 제안된 지수화된 지수분포를 이용한 신뢰성 모형

지수화된 지수분포에 대한 평균값 함수는 (2.3)식과 (2.4)식을 연관하여 평균값 함수 $m(t|\alpha, \tau)$ 와 고장발생률 $\lambda_{EED}(t|\alpha, \tau)$ 을 다음과 같이 유도 할 수 있다.

$$m(t|\alpha, \tau) = \theta F_{EED}(t|\alpha, \tau) = \theta (1 - e^{-\tau t})^\alpha \dots\dots\dots (3.1)$$

$$\lambda(t|\alpha, \tau) = \theta f_{EED}(t|\alpha, \tau) = \theta \alpha \tau (1 - e^{-\tau t})^{\alpha-1} e^{-\tau t}$$

$$\dots\dots\dots (3.2)$$

단, $t > 0, \alpha, \tau > 0, \theta > 0.$

시간(0,t] 까지 조사하기 위한 시간 절단(Time truncated) 모형은 n 번째 까지 고장시점 자료를

$$x_k = \sum_{i=1}^k t_i \quad (k=1,2,\dots,n; 0 \leq t_1 \leq t_2 \leq \dots \leq t_n) \dots\dots\dots (3.3)$$

이라고 하면 데이터 집합 D_t 는 $\{n, x_1, x_2, \dots, x_n; t\}$ 와 같이 구성된다.

t 을 최종 고장시점 x_n 으로 대치하고 평균값 함수 $m(t|\alpha, \tau)$ 와 고장발생률 $\lambda_{EED}(t|\alpha, \tau)$ 을 이용하면 우도함수 $L_{EED}(\theta, \alpha, \tau | D_x)$ 는 다음과 같이 표현 할 수 있다.

$$L_{EED}(\theta, \alpha, \tau) = \prod_{k=1}^n \theta \alpha \tau (1 - e^{-\tau x_k})^{\alpha-1} \dots\dots\dots (3.4)$$

$$\cdot e^{-\tau x_k} \exp(-\theta (1 - e^{-\tau x_n})^\alpha)$$

따라서 최우추정법을 이용하기 위한 지수화된 지수분포 모형 로그우도함수는 다음과 같이 유도된다.

$$\ln L_{EED}(\theta, \alpha, \tau) = n \ln \theta + n \ln \alpha + n \ln \tau \dots\dots\dots (3.5)$$

$$+ (\alpha - 1) \ln \sum_{k=1}^n (1 - e^{-\tau x_k})$$

$$- \tau \sum_{k=1}^n x_k - \theta (1 - e^{-\tau x_n})^\alpha$$

형상 모수 α 값은 상수(사전에 알고 있는 경우)라고 가정했을 때 최우추정법을 이용하기 위하여 (3.5)식을 θ 와 τ 에 대하여 편미분을 하면 다음과 같은 식을 유도 할 수 있다.

$$\frac{\partial \ln L_{EED}}{\partial \theta} = \frac{n}{\theta} - (1 - e^{-\tau x_n})^\alpha = 0 \dots\dots\dots (3.6)$$

$$\frac{\partial \ln L_{EED}(\theta, \alpha, \tau)}{\partial \tau} = \frac{n}{\tau} + \frac{\partial (\alpha - 1) \ln \sum_{k=1}^n (1 - e^{-\tau x_k})}{\partial \tau} - \sum_{k=1}^n x_k - \frac{\partial \theta (1 - e^{-\tau x_n})^\alpha}{\partial \tau} \dots\dots\dots (3.7)$$

따라서 고장절단 모형에서의 각 모수에 대한 최우추정량은 $\hat{\theta}$ 와 $\hat{\tau}$ 은 다음 식을 만족한다.

$$\frac{n}{\hat{\theta}} = (1 - e^{-\hat{\tau} x_n})^\alpha \dots\dots\dots (3.8)$$

$$\frac{n}{\hat{\tau}} = - \frac{\partial (\alpha - 1) \ln \sum_{k=1}^n (1 - e^{-\hat{\tau} x_k})}{\partial \hat{\tau}} + \sum_{k=1}^n x_k + \frac{\partial \hat{\theta} (1 - e^{-\hat{\tau} x_n})^\alpha}{\partial \hat{\tau}} \dots\dots\dots (3.9)$$

(3.8) 식과 (3.9) 식을 비선형 연립방정식(수치해석적 방법)을 이용하여 풀면 최우추정치 $\hat{\tau}_{MLE}$ 와 $\hat{\theta}_{MLE}$ 의 값을 구할 수 있다.

본 논문에서는 지수화 지수분포의 형상 모수 α 값은 1, 1.5, 2, 2.5, 3 과 같은 상수인 경우를 본 논문에서는 제시하고자 한다. 예를 들어 $\alpha = 1$ 인 경우에는 Goel-Okumoto 모형과 같고 $\alpha = 2$ 인 경우에는 다음과 같은 연립방정식을 유도할 수 있으며 다른 경우도 유사한 방법을 이용 할 수 있다.

$$\frac{n}{\hat{\theta}} = (1 - e^{-\hat{\tau} x_n})^2 \dots\dots\dots (3.10)$$

$$\frac{n}{\hat{\tau}} = - \ln \sum_{k=1}^n x_k e^{-\hat{\tau} x_k} + \sum_{k=1}^n x_k - 2 \hat{\theta} (1 - e^{-\hat{\tau} x_n}) x_n e^{-\hat{\tau} x_n} \dots\dots\dots (3.11)$$

IV. 소프트웨어 고장 자료 분석

이 장에서 실제 고장자료를 이용하여 형상모수에 따른 지수화된 지수화 신뢰도 모형을 분석하고자 한다. 이 고장 자료는 NTDS(Naval Tactical Data System)에 의해 발생된 소프트웨어 고장자료로서 Goel과Okumoto[11], Mazzuchi 와 Soyer[12], Kuo와 Yang[8] 등이 이 고장 자료를 이용하여 소프트웨어 모형을 제안 한 바 있다. 본 연구도 이 자료를 이용하고자 한다. 자료는 <표 1>에 요약 되었다.

표 1. 고장 간격 자료
Table 1. Failure interval time

Failure number	Failure Interval Time	Failure Time	Failure Time $\times 10^{-2}$
1	9	9	0.09
2	12	21	0.21
3	11	32	0.32
4	4	36	0.36
5	7	43	0.43
6	2	45	0.45
7	5	50	0.5
8	8	58	0.58
9	5	63	0.63
10	7	70	0.7
11	1	71	0.71
12	6	77	0.77
13	1	78	0.78
14	9	87	0.87
15	4	91	0.91
16	1	92	0.92
17	3	95	0.95
18	3	98	0.98
19	6	104	1.04
20	1	105	1.05
21	11	116	1.16
22	33	149	1.49
23	7	156	1.56
24	91	247	2.47
25	2	249	2.49
26	1	250	2.5
27	87	337	3.37
28	47	384	3.84
29	12	396	3.96
30	9	405	4.05
31	135	540	5.4

<표 1>의 자료를 이용하여 제시하는 신뢰모형들을 분석하기 위하여 우선 자료에 대한 추세 검정이 선행 되어야 한다[2,13].

추세 분석에는 산술평균 검정(Arithmetic mean test) 과 라플라스 추세 검정(Laplace trend test)등이 있다. 이 검정을 실시한 결과 <그림 2>에서 산술평균 검정결과 고장 시간이 증가함에 따라 산술 평균이 거의 증가 추세를 보이고 있으므로 신뢰성장(Reliability growth)이되고 있음을 나타내고 있고 <그림 3>에 나타난 라플라스 추세 검정의 결

과도 리플라스 요인(Factor)이 고장시간이 증가함에 따라 음수로서 감소하기 때문에 역시 신뢰성장(Reliability growth)이 되고 있음을 나타내고 있다. 따라서 이 자료를 가지고 신뢰성장 모형을 제시하는 것이 효율적임을 시사하고 있다[2.13]. 본 논문에서는 모수추정을 용이하게 하기 위하여 실제자료에서 수치변환된 자료($failure\ times \times 10^{-2}$)를 이용하였다.

소프트웨어 신뢰성 모형의 모수 추정은 최우추정법을 이용하였고 비선형 방정식의 계산방법은 수치해석적 기본 방법인 이분법(Bisection method)을 사용하였다. 이러한 계산은 초기값을 10^{-4} 와 1 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. 그러나 $\alpha=3$ 이상인 경우는 근이 수렴되지 않아 본 논문에서는 고려대상에서 제외 시켰다.

기존의 모형과 지수화된 지수화 신뢰도 모형에 대한 모수의 추정값들의 결과는 <표 2>에 요약되었다.

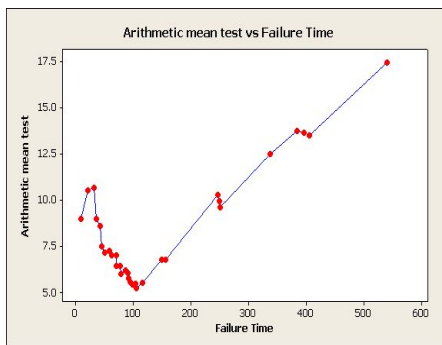


그림 2. 산술평균 검정
Fig. 2 Arithmetic mean test

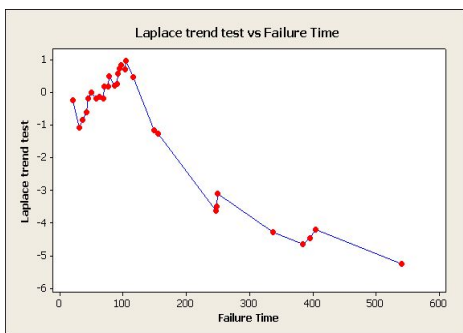


그림 3. 리플라스 추세 검정

Fig. 3 trend test

이 분야에서는 일반적으로 모형 선택의 하나의 방법으로 편차자승합(SSE [9])을 이용할 수 있다. 본 논문은 효율적 모형의 선택을 위하여 SSE 및 SSE 기준을 보완하는 AIC(Akaike Information Criterion) 통계량을 이용하였다. AIC 통계량은 적용모형에 대하여 자유도와 관련된 우도함수를 최대화 시켜주는 척도이다. 이 척도는 다음과 같이 정의된다[9].

$$AIC = -2 \log(\max \text{ of likelihood function}) + 2N$$

단, N 은 적용모형에 대한 모수의 수를 의미하고 이러한 AIC 척도는 SSE와 마찬가지로 작은 값을 가지는 모형이 상대적으로 효율적인 모형으로 간주된다. 이러한 결과는 <표 3>에 요약되었다. 이 표에서 지수화된 지수화 분포의 형상 모수(α)를 이용한 신뢰도 모형들이 이 분야에서 기존에 알려진 모형인 Yamada-Ohba-Osaki 모형이나 Goel-Okumoto 모형에 비해 상대적으로 효율적 모형으로 나타나고 있다.

예측 오류(Prediction error)의 비정상성(Nonstationarity)에 대한 측도는 Kolmogorov 거리(Distance)[2.14]로 측정되는데 이 거리가 클수록 상대적으로 비정상성을 내포하고 있다. <그림 4>는 S-Plus 소프트웨어[15]를 이용하여 Kolmogorov 검정에 대한 그림을 보여주고 있고 이 그림에서도 지수화된 지수화 신뢰도 모형이 상대적으로 비정상성 속성이 덜 내포하고 있음을 알 수 있다. <표 4>는 Kolmogorov 거리를 요약한 표로서 모형에 대한 치우침(Bias)을 나타내고 있다. 이 표에서도 지수화된 지수화 신뢰도 모형($\alpha=2, 3$)들이 상대적으로 비정상성 속성이 덜 내포하고 있음을 알 수 있다.

표 2. 각 모형의 모수 추정값
Table 2. Estimator of each model

Model	MLE
Goel-Okumoto Model (EED Model($\alpha=1$))	$\hat{\beta}_{1,MLE} = 0.39812$
	$\hat{\theta}_{MLE} = 35.07$
Yamada-Ohba-Osaki Model	$\hat{\beta}_{2,MLE} = 0.68351$
	$\hat{\theta}_{MLE} = 34.86$
EED Model ($\alpha=1.5$)	$\hat{\tau}_{MLE} = 0.87532$
	$\hat{\theta}_{MLE} = 31.42$
EED Model ($\alpha=2$)	$\hat{\tau}_{MLE} = 0.99339$
	$\hat{\theta}_{MLE} = 31.32$
EED Model ($\alpha=2.5$)	$\hat{\tau}_{MLE} = 1.0382$
	$\hat{\theta}_{MLE} = 31.29$
EED Model ($\alpha=3$)	$\hat{\tau}_{MLE} = 2.0041$
	$\hat{\theta}_{MLE} = 31.14$

표 3. 모형들에 대한 SSE 및 AIC 의 값
Table 3. SSE and AIC of each model

Model	SSE	AIC
Goel-Okumoto Model (EED Model($\alpha=1$))	9013.172	109.1432
Yamada-Ohba-Osaki Model	5049.463	104.0325
EED Model($\alpha=1.5$)	2726.814	98.0673
EED Model($\alpha=2$)	2613.983	94.3456
EED Model($\alpha=2.5$)	2467.353	90.1275
EED Model($\alpha=3$)	1154.564	88.8945

Kolmogorov-test -- plot

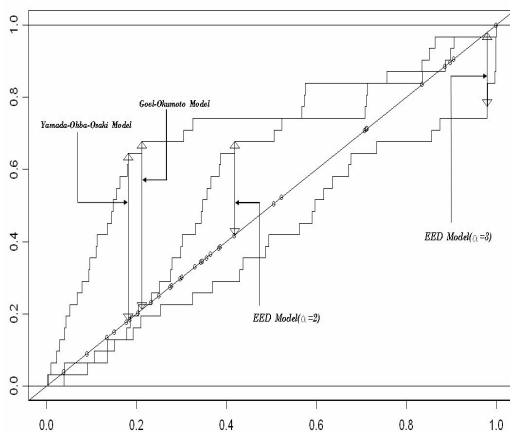


그림 4. 모형에 대한 Kolmogorov 검정 그림
Fig. 4 Kolmogorov test of each model

표 4. 모형에 대한 Kolmogorov 거리
Table 4. Kolmogorov distance of each model

Model	Kolmogorov distance
Goel-Okumoto Model($\alpha=1$)	0.464987
Yamada-Ohba-Osaki Model	0.258784
EED Model($\alpha=2$)	0.223357
EED Model($\alpha=3$)	0.204768

t 는 임무시간(Mission time)에 대한 즉, 소프트웨어의 고장이 일어나지 않을 확률인 신뢰도(Reliability)

$\hat{R}(t | x_n)$ 는 <그림 5>와 같다. 이 그림에서도 지수화 된 지수화 신뢰도 모형($\alpha=2, 3$)들이 상대적으로 효율적인 신뢰도를 나타내고 있다.

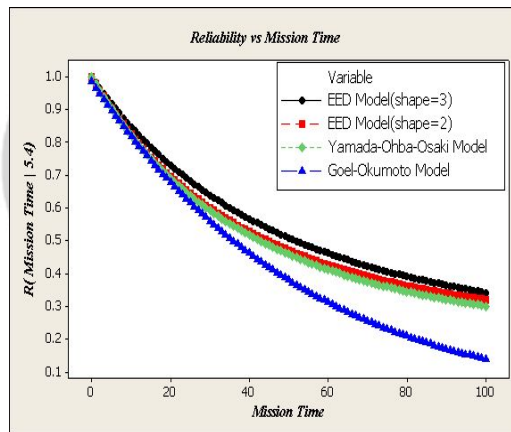


그림 5. 모형에 대한 신뢰도
Fig. 5 Reliability of each model

v. 결론

소프트웨어 신뢰성은 개발의 최종단계에 있는 테스트 공정이나 실제 사용단계에 있어서 소프트웨어 내에 존재하는 고장 수나 고장 발생 시간에 의해서 효과적으로 평가할 수 있는 상황으로 그 평가 기술이 중요하게 된다. 따라서 소프트웨어 개발의 테스트 공정이나 실제사용단계에 있어서 고장 발생 환경이나 고장 발생 현상을 수리적으로 모형화가

가능하면 평가를 할 수 있다. 테스트 시간이나 혹은 실행 시간, 발생된 고장 수와 고장 발생시간과의 관계를 효율적으로 관리함으로써 소프트웨어 신뢰도를 성장 시킬 수 있다. 이러한 과정을 소프트웨어 성장 과정이라고 볼 수 있다.

본 논문에서는 지수화된 지수화 분포를 이용한 지수화된 지수화 신뢰도 모형을 제안하였다. 고장 간격시간으로 구성된 NTDS 자료를 이용하여 기존의 모형과 지수화된 지수화 신뢰도 모형에 대하여 최우 추정법을 이용하여 모수 추정을 실시하였다.

효율적인 모형 비교를 위한 편차자승합의 결과는 지수화된 지수화 모형이 기존에 잘 알려진 Yamada-Ohba-Osaki 모형이나 Goel-Okumoto보다 우수함을 보이고 있고 콜모고로프 거리의 결과도 지수화된 지수화 모형이 상대적으로 정상성에 가까운 모형으로 간주할 수 있다. 분석된 자료에 대한 평가에서도 산술평균 검정과 라플라스 추세 검정을 실시한 결과도 신뢰성장이 되고 있음을 나타내고 있다. 신뢰도에 대한 평가도 역시 지수화된 지수화 모형이 신뢰도가 높음을 알 수 있다.

참고문헌

- [1] 김 희철. "일반화감마분포를 이용한 NHPP 소프트웨어 신뢰도 모형에 관한 연구", 한국컴퓨터정보학회 논문지, 10권 6호, pp. 27-35, 2005.
- [2] S. S. Gokhale and K. S. Trivedi. "A time/structure based software reliability model". *Annals of Software Engineering*, 8, pp. 85-121, 1999
- [3] 김 희철. "카이제곱 NHPP에 의한 소프트웨어 신뢰성장 모형에 관한 연구", 한국컴퓨터정보학회논문지, 제11권 1호, pp. 45-53, 2006.
- [4] R. D. Gupt and D. Kundu, "Exponentiated exponential family; an alternative to gamma and Weibull", *Biometrical Journal*, Vol. 43, pp. 117-130, 2001.
- [5] R. D. Gupt and D. Kundu, "Discriminating between gamma and generalized exponential distribution", *Journal of Statistical Computation and Simulation*, Vol. 74, No. 2, pp. 107-121, 2004.
- [6] J. F. Lawless. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, New York, 1981.
- [7] H. Pham and L. Nordmann and X. Zhang "A General Imperfect-Software -Debugging Model with S-Shaped Fault-Detection Rate". *IEEE Trans. on reliability*, Vol, 48, No 2, pp, 169-175, 1999.
- [8] L. Kuo and T. Y. Yang. "Bayesian Computation of Software Reliability". *Journal of the American Statistical Association*, Vol.91, pp. 763-773, 1996.
- [9] S. Yamada, M. Ohba and S. Osaki. "S-Shaped Reliability Growth Modeling for Software Error Detection". *IEEE Trans. on Reliability*. R-32(5) : pp. 475-485, Dec. 1983.
- [10] V. K. Rohatgi. *Statistical inference*, pp. 398-416. JOHN WILEY & SONS, INCI, New York, 1984.
- [11] A. L. Goel and K. Okumoto. "Time-Dependent Error-Detection Rate Models for Software Reliability and Other Performance Measures". *IEEE Trans. on Reliability*, R-28(3):pp. 206-211, Aug. 1979.
- [12] T. A. Mazzuchi and R. Soyer. "A Bayes Empirical Bayes Model for Software Reliability". *IEEE Transactions on Reliability*, 37, pp 248-254, 1988
- [13] K. Kanoun and J. C. Laprie. *Handbook of Software Reliability Engineering*, MRLyu, Editor, chapter Trend Analysis, pp.401-437.
- [14] S. Brocklehurst and B. Littlewood, *Handbook of Software Reliability Engineering*, MRLyu, Editor, chapter Techniques for Prediction Analysis and Recalibration, pp.119-166.
- [15] S. Selvin. *Modern Applied Biostatistical Methods Using S-Plus*, pages 141-184. Oxford University Press, New York, 1998.

저 자 소 개



김 희 철

1992년 동국대학교 통계학과 석사

1998년 동국대학교 통계학과 박사

2000년 3월 ~ 2004년 2월

송호대학 정보산업계열 조교수

2005년 3월~현재 남서울대학교

산업경영공학과 전임강사

관심분야: 소프트웨어 신뢰성공학,

웹프로그래밍, 전산통계, 인터넷비즈니스

K C I