

지수화 지수 커버리지 함수를 고려한 ENHPP 소프트웨어 신뢰성장 모형에 관한 연구

A Study on ENHPP Software Reliability Growth Model based on
Exponentiated Exponential Coverage Function

김 희 철 (Kim, Hee Cheul)*

목 차

- I. 서론
 - II. 관련연구
 - III. 제안된 지수화 지수 커버리지모형
 - IV. ENHPP모형에 대한 모수 추정
 - V. 소프트웨어 고장 자료 분석
 - VI. 결론
-

Abstract

Finite failure NHPP models presented in the literature exhibit either constant, monotonic increasing or monotonic decreasing failure occurrence rates per fault. Accurate predictions of software release times, and estimation of the reliability and availability of a software product require quantification of a critical element of the software testing process : test coverage. This model called enhanced non-homogeneous poisson process(ENHPP). In this paper, exponential coverage and S-coverage model was reviewed, proposes the exponentiated exponential coverage reliability model, which made out efficiency substituted for gamma and Weibull model(2 parameter shape illustrated by Gupta and Kundu(2001). In this analysis of software failure data, algorithm to estimate the parameters used to maximum likelihood estimator and bisection method, model selection based on SSE statistics for the sake of efficient

* 남서울대학교 산업경영공학과 교수

model, was employed.

Key words: Test Coverage, Enhanced Non-homogeneous Poission Process, Exponentiated Exponential Distribution, Sum of the Squared Errors

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. 신뢰도에서 관측시간에 발견된 고장수를 모형화 하는데 비동질적 포아송 과정(Non-homogeneous poisson process; NHPP)이 널리 사용하여 왔다. 이러한 NHPP모형은 강도함수(Intensity function)와 평균값함수(Mean value function)에 의존한다[1,2].

소프트웨어 제품의 정확한 인도시기(Release times)를 예측하거나 효율성 및 신뢰성을 예측하기 위해서는 소프트웨어 테스트 과정에서 중요한 요소인 테스트 커버리지(Coverage)를 이용하면 보다 효율적인 테스트 작업을 할 수 있다. 이러한 모형은 기존에 존재하는 NHPP모형에서 테스트 커버리지를 포함하는 모형이 된다. 이런 모형을 ENHPP(Enhanced non-homogeneous poisson process) 이라고 한다[3,4].

본 연구에서는 유한 고장 ENHPP 모형에 대하여 수명 분포가 지수를 가지는 Goel-Okumoto 모형 및 S-형과 이들의 중첩 및 혼합모형 및 을 제시하여 신뢰성 척도를 추정하고 이를 바탕으로 소프트웨어 신뢰성에 대한 특성을 알아보고자 한다. 본 논문의 2장에서는 관련연구로서 테스트 커버리지, 단순 커버리지 및 제안된 중첩 및 혼합 커버리지 함수에 에 대하여 서술하였고 3장에서는 ENHPP에 모수 추정방법에 대하여 나열하였으며 4장에서는 수치적인 예로 실제 고장자료를 이용하여 각 모형에 대한 모수추정 및 모형비교를 실시하였으며 마지막으로 6장에서는 결론을 나열 하였다

II. 관련연구

1. 소프트웨어 신뢰성에 대한 테스트 커버리지

테스트 커버리지를 유한고장 비동질적 포아송 과정 NHPP에 적용 시킬 수 있다. 이러한 적

용 모형을 확장된 비동질적 포아송 과정(ENHPP)이라고 부른다. 이러한 ENHPP 모형의 오류 탐색비율은 숨겨진 잠재 오류 사이트의비와 남아있는 오류함의 기대수와의 곱에 비례한다. 이러한 ENHPP 모형은 다음과 같은 가정을 한다[3, 5].

- (가정 1) 오류들은 모든 잠재 오류 사이트에 대하여 균일하게(uniformly) 분포되어 있다.
- (가정 2) 잠재 오류 사이트가 숨겨졌을 때 사이트에서 오류가 탐색되어 질수 있는 확률은 $K(t)$.
- (가정 3) 수리는 효과적으로 즉시 이루어지고 새로운 오류는 발생되지 않는다. (이 가정은 유한 고장 NHPP 모형과 유사하다)
- (가정 4) 커버리지는 테스트 시간에 대하여 연속적인 단조 비감소 함수(Monotonic non-decreasing function)를 따른다.

이러한 모형을 해석학적으로 표시하면 다음과 같이 표현된다.

$$\frac{dm(t)}{dt} = \tilde{\theta} K(t) \frac{dc(t)}{dt} \quad (2.1)$$

혹은
$$m(t) = \tilde{\theta} \int_0^t K(\tau) c'(\tau) d\tau.$$

단, $\tilde{\theta}$ 는 무한 테스트 시간이 주어졌을 때 탐색될수 있는 오류의 기대수이고 완전한 오류 탐색 커버리지는 $K(t)=1$ 이고 완전 테스트 커버리지는 다음을 만족한다.

$$\lim_{t \rightarrow \infty} c(t) = 1 \quad (2.2)$$

만약 $K(t)$ 가 상수 K 로 가정하면 t 시점에서 탐색 될 오류의 기대수 $m(t)$ 는 다음과 같이 표현 가능하다.

$$m(t) = \tilde{\theta} K c(t) \quad (2.3)$$

(2.3)식은 t 시점에서 탐색 될 오류의 기대수는 오류의 총 오류의 기대수와 t 시점에서의 오류 탐색 커버리지의 곱과 같다. 따라서 $\theta = \tilde{\theta} K$ 로 하면 탐색 될 오류의 기대수 $m(t)$ 는 다음과 같이 또 다른 형태로 표현 가능하다.

$$m(t) = \theta c(t) \quad (2.4)$$

이런 결과로 인하여 고장 강도함수(Intensity function) $\lambda(t)$ 을 적용시키면 다음과 같다.

$$\frac{dm(t)}{dt} = \lambda(t) = \theta c'(t) \quad (2.5)$$

고장 강도 함수 $\lambda(t) = [\theta - m(t)] h(t)$ 임이 됨이 알려져 있으므로[1, 3, 4] (2.5)식을 변형하면 다음과 같이도 표현도 가능하다.

$$\lambda(t) = [\theta - m(t)] \frac{c'(t)}{1 - c(t)} = [\theta - m(t)] h(t) \quad (2.6)$$

그러므로 한 개의 오류당 오류발생비율인 위험함수(Hazard function) $h(t)$ 는 다음과 같이 유도 된다.

$$h(t) = \frac{c'(t)}{1 - c(t)} \quad (2.7)$$

따라서 이러한 ENHPP모형은 테스트가 진행됨에 따라 각각의 오류가 발생할 수 있는 비율은 시간에 따라 변화하기 때문에 시간에 의존하는 고장 발생 비율이라고 할 수 있다. 또, 이러한 위험함수에 해당하는 분포는 t 시점에서 커버리지 함수를 평가 할 수 있다. 이러한 ENHPP 모형은 (불완전한) 커버리지 함수에 대하여 실제적인 상황($c(\infty) < 1$) 을 적용 시킬 수 있다. 결과적으로 조건부 신뢰도(Conditional reliability) $R(t | s)$ 는 다음과 같이 알려져 있다[3].

$$R(t | s) = e^{-\int_s^{s+t} \lambda(\tau) d\tau} = e^{-a[c(s+t) - c(s)]} \quad (2.8)$$

단, t 는 임무시간(Mission time)이고 s 는 마지막 고장 시간이다.

2. 커버리지 함수를 이용한 NHPP 모형(ENHPP)

1) 지수 커버리지 함수

잘 알려진 Goel-Okumoto 모형[3, 7]은 이 분야에서 기본적인 모형이다. 이 모형은 지수 커버리지 함수를 이용한 모형 즉, 오류당 고장 발생 비율이 상수를 가지는 모형이 된다. 이러한 지수 모형을 NHPP로 접근하면 평균값 함수와 강도 함수는 다음과 같이 알려져 있다.

$$m(t|\theta, g) = \theta(1 - e^{-\beta_1 t}) \quad (\theta > 0, \beta_1 > 0) \quad (2.9)$$

$$\lambda(t|\theta, \beta_1) = \theta \beta_1 e^{-\beta_1 t} \quad (2.10)$$

단, θ 는 무한하게 테스트 시간이 주어졌을 때 관찰된 고장의 기대수이고 β_1 는 오류당 고장 발생 비율을 의미한다. 따라서 지수모형에 대한 커버리지 함수 $c(t)$ 와 위험함수 $h(t)$ 는 다음과 같이 유도 된다.

$$c(t|\beta_1) = 1 - e^{-\beta_1 t} \quad (2.11)$$

$$h(t|\beta_1) = \frac{c'(t)}{1 - c(t)} = \beta_1 \quad (2.12)$$

2) S-형 커버리지 함수

S-형 커버리지 함수 모형[3, 8]은 Yamada-Ohba-Osaki 모형으로 알려져 있고 S-형 모형을 NHPP로 접근하면 평균값 함수와 강도 함수는 다음과 같이 알려져 있다.

$$m(t|\theta, \beta_2) = \theta [1 - (1 + \beta_2 t) e^{-\beta_2 t}] \quad (\theta > 0, g > 0) \quad (2.13)$$

$$\lambda(t|\theta, \beta_2) = \theta \beta_2^2 t e^{-\beta_2 t} \quad (2.14)$$

따라서 S-형모형에 대한 커버리지 함수 $c(t)$ 와 위험함수 $h(t)$ 는 다음과 같이 유도 된다.

$$c(t|\beta_2) = 1 - (1 + \beta_2 t) e^{-\beta_2 t} \quad (2.15)$$

$$h(t|\beta_2) = \frac{c'(t)}{1 - c(t)} = \frac{\beta_2^2 t}{1 + \beta_2 t} \quad (2.16)$$

III. 제안된 지수화된 지수 분포 커버리지모형

1. 지수화된 지수 분포

이 절에서 지수화된 지수 분포(Exponentiated exponential distribution; EED)[9]에 대하여 설명하고자 한다. 원래 지수화된 지수분포는 Gompertz-Verhulst 분포함수

$$G(t) = (1 - \rho e^{-tb})^a, t > \frac{1}{b} \ln \rho, \rho, \alpha, b > 0 \quad (3.1)$$

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

$$F(t|a, b) = (1 - e^{-bt})^a \quad (t > 0, \alpha, b > 0) \quad (3.2)$$

(2.18) 식에서 $a(>0)$ 와 $b(>0)$ 는 각각 형상모수(Shape parameter) 와 척도모수(Scale parameter)를 의미하고 확률밀도함수는 다음과 같이 유도 된다.

$$f(t|a, b) = ab(1 - e^{-bt})^{a-1} e^{-bt} \quad (t > 0, a, b > 0) \quad (3.3)$$

이러한 지수화된 지수 분포는 최근에 Gupta 와 Kundu[9, 10]에 의해서 2 모수 감마분포나 와이블 분포의 대체모형으로서 여러 가지 수명자료를 분석함에 있어서 효율적 분포가 됨이 밝혀 졌다.

또, 위험함수는(2.18)과 (2.19)식과 관련하여 다음과 같이 유도 된다.

$$h(t|a, b) = \frac{f(t|a, b)}{1 - F(t|a, b)} = \frac{ab(1 - e^{-bt})^{a-1} e^{-bt}}{1 - (1 - e^{-bt})^a} \quad (3.4)$$

$b=1$ 인 경우에 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.

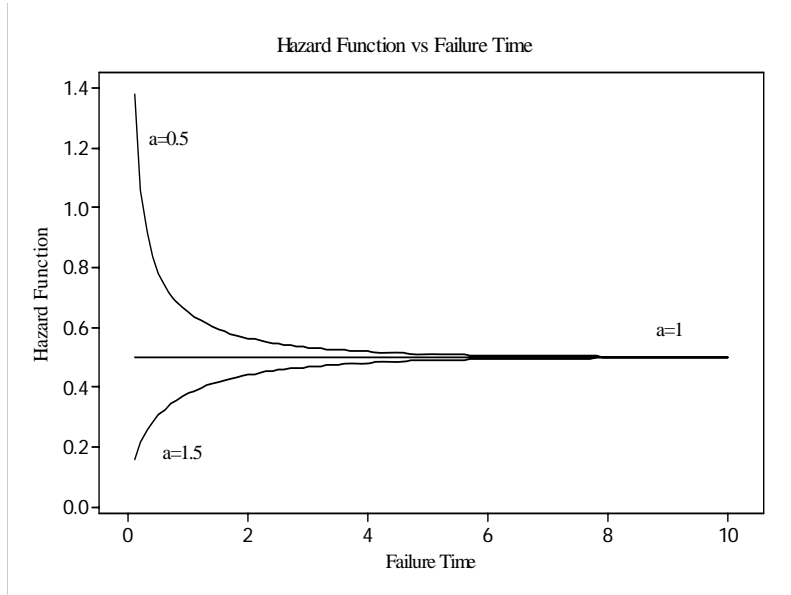


그림 1. 지수화된 지수 분포의 위험함수($b=1$ 인 경우)

2. 지수화된 지수 커버리지 함수

지수화된 지수분포에 대한 평균값 함수는 (2.4)식과 (2.5)식을 연관하여 평균값 함수 $m_{EED}(t|\theta, a, b)$ 와 강도함수는 $\lambda_{EED}(t|\theta, a, b)$ 을 다음과 같이 유도 할 수 있다.

$$m_{EED}(t|\theta, a, b) = \theta F_{EED}(t|a, b) = \theta (1 - e^{-bt})^a \tag{3.6}$$

$$\lambda_{EED}(t|\theta, a, b) = \theta f_{EED}(t|a, b) = \theta ab(1 - e^{-bt})^{a-1} e^{-bt} \tag{3.7}$$

단, $t > 0, a, b > 0, \theta > 0$.

따라서 지수화 지수분포모형에 대한 커버리지 함수 $c(t)$ 와 위험함수 $h(t)$ 는 다음과 같이 유도 된다.

$$c(t) = (1 - e^{-bt})^a \tag{3.8}$$

$$h(t) = \frac{c'(t)}{1 - c(t)} = \frac{ab(1 - e^{-bt})^{a-1} e^{-bt}}{1 - (1 - e^{-bt})^a} \tag{3.9}$$

IV. ENHPP 모형에 대한 모수 추정

Goel-Okumoto 모형[7]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 고장 시점 자료를

$$x_k = \sum_{i=1}^k t_i \quad (k=1, 2, \dots, n; 0 \leq x_1 \leq x_2 \leq \dots \leq x_n)$$

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 우도함수는 다음과 같이 표현 할 수 있다.

$$L_{GO}(\theta, \beta_1 | D_{x_n}) = \left(\prod_{k=1}^n \theta \beta_1 e^{-\beta_1 x_k} \right) \exp[-\theta(1 - e^{-\beta_1 x_n})] \quad (4.1)$$

최우추정법(MLE)을 이용하기 위한 로그우도함수를 구하면 다음과 같이 표현된다.

$$\ln L_{GO} = n \ln \theta + n \ln \beta_1 - \beta_1 \sum_{k=1}^n x_k - \theta(1 - e^{-\beta_1 x_n}) \quad (4.2)$$

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

$$\frac{\partial \ln L_{GO}}{\partial \theta} = \frac{n}{\theta} - 1 + e^{-\beta_1 x_n} = 0, \quad \frac{\partial \ln L_{GO}}{\partial \beta_1} = \frac{n}{\beta_1} - \sum_{k=1}^n x_k - \theta x_n e^{-\beta_1 x_n} = 0 \quad (4.3)$$

따라서 각 모수에 대한 최우추정량 $\hat{\theta}_{MLE}$ 와 $\hat{\beta}_{1MLE}$ 은 다음식을 만족한다.

$$\frac{n}{\hat{\theta}} = 1 - \exp(-\hat{\beta}_1 x_n), \quad \frac{n}{\hat{\beta}_1} = \sum_{k=1}^n x_k + \hat{\theta} x_n \exp(-\hat{\beta}_1 x_n) \quad (4.4)$$

(4.4)식을 수치적으로 풀어 두 모수를 구할 수 있다.

Yamada-Ohba-Osaki 모형[8]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.

$$L_{YOO}(\theta, \beta_2 | D_{x_n}) = \left(\prod_{k=1}^n \theta \beta_2^2 x_k e^{-\beta_2 x_k} \right) \cdot \exp[-\theta [1 - (1 + \beta_2 x_n) e^{-\beta_2 x_n}]] \quad (4.5)$$

따라서 로그우도함수를 구하면 다음과 같이 표현된다.

$$\ln L_{YOO} = n \ln \theta + 2n \ln \beta_2 + \sum_{k=1}^n \ln x_k - \beta_2 \sum_{k=1}^n x_k - \theta [1 - (1 + \beta_2 x_n) e^{-\beta_2 x_n}] \quad (4.6)$$

따라서 Goel-Okumoto 모형과 유사하게 각 모수에 대한 최우추정량은 $\hat{\theta}_{MLE}$ 와 $\hat{\beta}_{2MLE}$ 은 다음식을 만족한다.

$$\frac{n}{\hat{\theta}} = 1 - \exp(-\hat{\beta}_2 x_n) - \hat{\beta}_2 x_n \exp(-\hat{\beta}_2 x_n), \quad \frac{2n}{\hat{\beta}_2} = \sum_{k=1}^n x_k + \hat{\theta} \hat{\beta}_2 x_n^2 \exp(-\hat{\beta}_2 x_n) \quad (4.7)$$

(4.7)식을 수치적으로 풀어 두 모수를 구할 수 있다.

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

$$m_{EED}(t|\theta, a, b) = \theta F_{EED}(t|a, b) = \theta (1 - e^{-bt})^a \quad (4.8)$$

$$\lambda_{EED}(t|\theta, a, b) = \theta f_{EED}(t|a, b) = \theta a b (1 - e^{-bt})^{a-1} e^{-bt} \quad (4.9)$$

단, $t > 0, a, b > 0, \theta > 0$.

(4.8), (4.9)을 이용한 우도함수는 다음과 같다.

$$L_{EED}(\theta, a, b) = \prod_{k=1}^n \theta a b (1 - e^{-b x_k})^{a-1} \cdot e^{-b x_k} \exp(-\theta (1 - e^{-b x_n})^a) \quad (4.10)$$

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

$$\ln L_{EED}(\theta, a, b) = n \ln \theta + n \ln a + n \ln b + (a-1) \ln \sum_{k=1}^n (1 - e^{-b x_k}) - b \sum_{k=1}^n x_k - \theta (1 - e^{-\tau x_n})^a \quad (4.11)$$

형상 모수 a 값은 상수(사전에 알고 있는 경우)라고 가정 했을 때 최우추정법을 이용하기 위하여 (3.4)식을 θ 와 b 에 대하여 편미분을 하면 다음과 같은 식을 유도 할 수 있다.

$$\frac{\partial \ln L_{EED}}{\partial \theta} = \frac{n}{\theta} (1 - e^{-\lambda x_n})^a = 0 \quad (4.12)$$

$$\frac{\partial \ln L_{EED}(\theta, a, b)}{\partial b} = \frac{n}{b} + \frac{\partial (a-1) \ln \sum_{k=1}^n (1 - e^{-b x_k})}{\partial b} - \sum_{k=1}^n x_k - \frac{\partial \theta (1 - e^{-\tau x_n})^a}{\partial b} = 0 \quad (4.13)$$

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

$$\frac{n}{\hat{\theta}} = (1 - e^{-\hat{b} x_n})^a \quad (4.14)$$

$$\frac{n}{\hat{b}} = - \frac{\partial (a-1) \ln \sum_{k=1}^n (1 - e^{-\hat{b} x_k})}{\partial \hat{b}} + \sum_{k=1}^n x_k + \frac{\partial \hat{\theta} (1 - e^{-\hat{b} x_n})^a}{\partial \hat{b}} \quad (4.15)$$

(4.14) 식과 (4.15) 식을 비선형 연립방정식(수치해석적 방법)을 이용하여 풀면 최우추정치 \hat{b}_{MLE} 와 $\hat{\theta}_{MLE}$ 의 값을 구 할 수 있다.

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

$$\frac{n}{\hat{\theta}} = (1 - e^{-\hat{b} x_n})^2, \quad \frac{n}{\hat{b}} = - \ln \sum_{k=1}^n x_k e^{-b t_k} + \sum_{k=1}^n x_k - 2 \hat{\theta} (1 - e^{-\hat{b} x_n}) x_n e^{-\hat{b} x_n} \quad (4.16)$$

V. 소프트웨어 고장자료 분석

이 장에서 실제 고장자료를 이용하여 형상모수에 따른 지수화된 지수화 신뢰도 모형을 분석

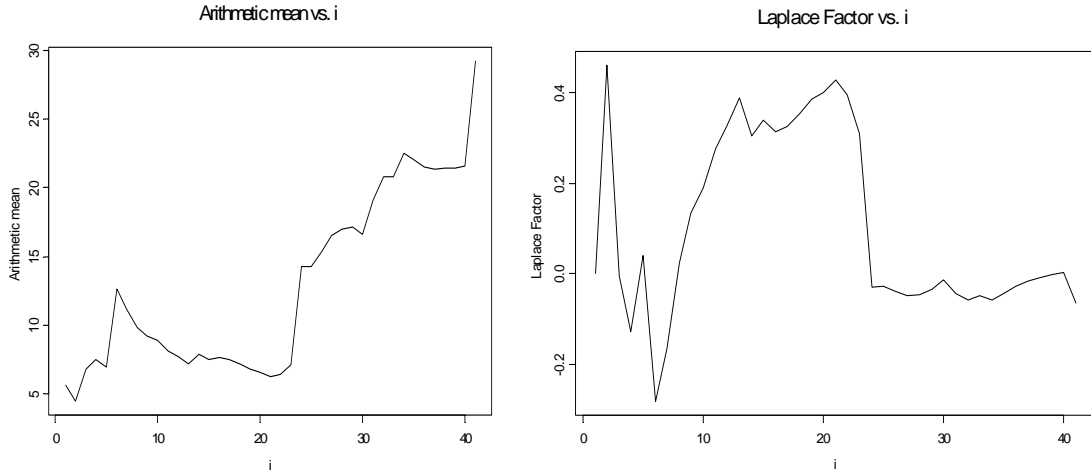
하고자 한다. 본 연구에서는 1197.95 시간단위에 41번의 고장이 발생한 S27[11] 고장자료를 이용하여 커버리지 소프트웨어 모형을 비교 하여 그 특성을 파악하고자 한다. 이 자료는 <표 1>에 나열 되어 있고 제시하는 신뢰 모형들을 분석하기 위하여 우선 자료에 대한 추세 검정이 선행 되어야 한다[3, 11].

추세 분석에는 산술평균 검정(Arithmetic mean test)과 라플라스 추세 검정(Laplace trend test)등이 있다. 이 검정을 실시한 결과 <그림 1>에서 산술평균 검정결과 고장시간이 증가함에 따라 산술 평균이 거의 증가 추세를 보이고 있으므로 신뢰성장(Reliability growth) 속성을 가지고 있고 라플라스 추세 검정의 결과도 라플라스 요인(Factor)이 -2와 2사이에 존재함으로서 신뢰성장(Reliability growth) 속성을 나타내고 있다. 따라서 신뢰성장 모형을 이 자료에 적용시키는 것이 가능하다.[11]

소프트웨어 신뢰성 모형의 모수 추정은 최우추정법을 이용하였고 비선형 방정식의 계산방법은 수치해석적 기본 방법인 이분법(Bisection method)을 사용하였다. 이러한 계산은 초기값을 10^{-2} 와 10 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 지수화 지수 커버리지 모형에서 $a = 1$ 인 경우는 Goel-Okumoto 커버리지 모형이 되기 때문에 본 문에서는 a 가 1.1에서 2까지의 고정값을 고려하였다. 기존의 모형과 지수화 지수 커버리지 모형에 대한 모수의 추정값들의 결과는 <표 2>에 요약되었다.

<표 1> 고장 간격 자료 S27

Failure number	Failure Interval (hours)	Failure Time(hours)	Failure number	Failure Interval (hours)	Failure Time(hours)
1	5.649	5.65	22	9.883	141.71
2	3.271	8.92	23	22.5	164.21
3	11.37	20.29	24	178.638	342.85
4	9.665	29.96	25	13.294	356.14
5	4.76	34.72	26	43	399.14
6	41.235	75.95	27	47.35	446.49
7	2.221	78.17	28	30.15	476.64
8	0.454	78.63	29	20.5	497.14
9	4.397	83.02	30	0.517	497.66
10	6.092	89.11	31	93.5	591.16
11	0.69	89.8	32	74.483	665.64
12	3.056	92.86	33	20.8	686.44
13	0.8	93.66	34	79.5	765.94
14	16.995	110.66	35	7.033	772.98
15	1.333	111.99	36	1.967	774.94
16	10.557	122.55	37	16.617	791.56
17	4.5	127.05	38	24.417	815.98
18	1.667	128.71	39	21.167	837.15
19	0.278	128.99	40	24.8	861.95
20	2.778	131.77	41	336	1197.95
21	0.061	131.83			



<그림 1> 산술평균 및 라플라스 추세 검정

<표 2> 각 모형의 모수 추정값

Model	MLE
G-0	$\hat{\theta} = 43.1927, \hat{\beta}_1 = 0.001488$
S-S	$\hat{\theta} = 41.3255, \hat{\beta}_2 = 0.005750$
EE(a = 1.1)	$\hat{\theta} = 42.7957, \hat{b} = 0.002725$
EE(a = 1.2)	$\hat{\theta} = 41.2583, \hat{b} = 0.002959$
EE(a = 1.3)	$\hat{\theta} = 41.1203, \hat{b} = 0.003174$
EE(a = 1.4)	$\hat{\theta} = 41.0612, \hat{b} = 0.003369$
EE(a = 1.5)	$\hat{\theta} = 41.0333, \hat{b} = 0.003545$
EE(a = 1.6)	$\hat{\theta} = 41.0187, \hat{b} = 0.003740$
EE(a = 1.7)	$\hat{\theta} = 41.0109, \hat{b} = 0.003896$
EE(a = 1.8)	$\hat{\theta} = 41.0065, \hat{b} = 0.004072$
EE(a = 1.9)	$\hat{\theta} = 41.0040, \hat{b} = 0.004229$
EE(a = 2.0)	$\hat{\theta} = 41.0025, \hat{b} = 0.004385$

단, G-O: 지수(Goel-Okumoto) 커버리지 모형, SS: S-형(Yamada-Ohba-Osaki)커버리지모형
 EE: 지수화 지수 커버리지 모형

모형 선택의 하나의 방법으로 편차자승합(SSE [12,13])을 이용할 수 있는데 이 편차자승합이 작으면 상대적으로 효율적인 모형이 된다. 주어진 자료를 이용하여 제시된 모형들에 대한 편차자승합의 값은 <표 3>에 요약되었다. 이 표에서 지수화지수 커버리지 모형이 기존에 알려진 모형인 S-커버리지 모형이나 지수 커버리지 모형에 비해 상대적으로 효율적 모형으로 나타나고 있다.

예측오류(Prediction error)의 비정상성(Nonstationarity)에 대한 측도는 Kolmogorov 거리(distance)[3, 14]로 측정되는데 이 거리가 클수록 상대적으로 비정상성을 내포하고 있다. <그림 2> 은 S-Plus 소프트웨어[14]를 이용하여 Kolmogorov 검정에 대한 그림을 보여주고 있고 이 그림에서도 지수화지수모형($a=1.2$)이 상대적으로 비정상성 속성이 덜 내포하고 있음을 알 수 있다. <표 3>에 Kolmogorov 거리를 계산된 값에서도 지수화지수 커버리지모형의 형태가 상대적으로 비정상성 속성이 덜 내포($a=1.1 \sim 1.7$)하고 있음을 알 수 있다.

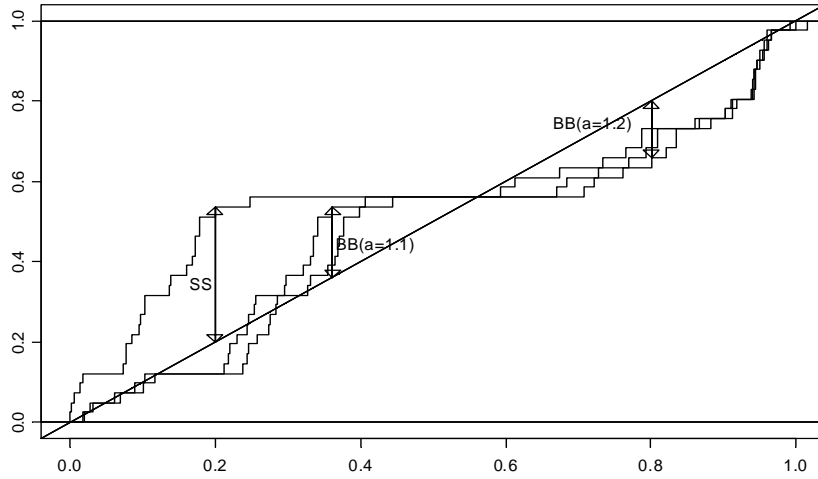
<표 3> 모형들에 대한 SSE 및 Kolmogorov 거리(Distance) 값

Model	SSE	K-D
G-0	2208	0.335302
S-S	1811.842	0.337444
EE($a=1.1$)	493.6929	0.175344
EE($a=1.2$)	480.1956	0.141607
EE($a=1.3$)	654.8205	0.18548
EE($a=1.4$)	757.9131	0.201251
EE($a=1.5$)	1266.339	0.253793
EE($a=1.6$)	1688.788	0.286505
EE($a=1.7$)	2124.052	0.311639
EE($a=1.8$)	2627.857	0.333751
EE($a=1.9$)	3145.819	0.351105
EE($a=2.0$)	3693.72	0.365241

단, G-O: 지수(Goel-Okumoto) 커버리지 모형 , SS: S-형(Yamada-Ohba-Osaki)커버리지모형
 EE: 지수화 지수 커버리지 모형, K-D: Kolmogorov 거리

<그림 3> 은 각 모형에 대하여 위험함수를 그린 그림이다. 이 그림에서는 지수 커버리지 모형은 일정하고 S-커버리지 모형과 지수화 지수 커버리지 모형은 증가형태를 나타내고 있으며 <그림 4> 에서는 형상모수 a 에 따른 지수화 지수 모형에 위험함수의 형태는 전체적으로 a 의 값이 증가 할수록 상대적으로 높은 값을 보이고 있다. <그림 5>에서는 커버리지 함수를 그린 그림이다. ENHPP 모형에서 가정한 대로 연속적인 단조 비감소 함수형태로 나타나고 있다고 형상모수 a 에 따른 지수화 지수 모형에 커버리지 함수의 형태는 전체적으로 a 의 값이 증가 할수록 상대적으로 처음에는 작은 값을 가지다가 높은 값으로 변모하는 패턴을 나타내고 있다(<그림 6>)

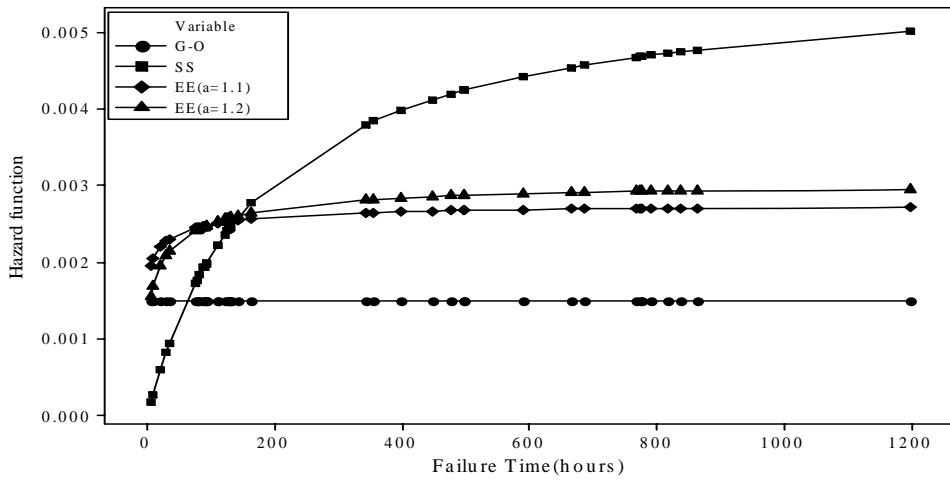
Kolmogorov-test -- plot



단, G-O: 지수(Goel-Okumoto) 커버리지 모형 , SS: S-형(Yamada-Ohba-Osaki)커버리지모형
 EE: 지수화 지수 커버리지 모형.

<그림 2> Kolmogorov 검정

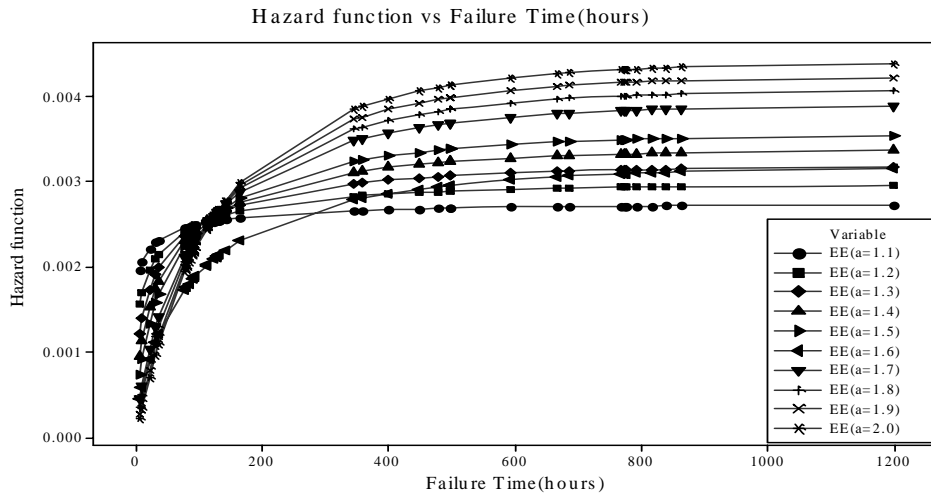
Hazard function vs Failure Time(hours)



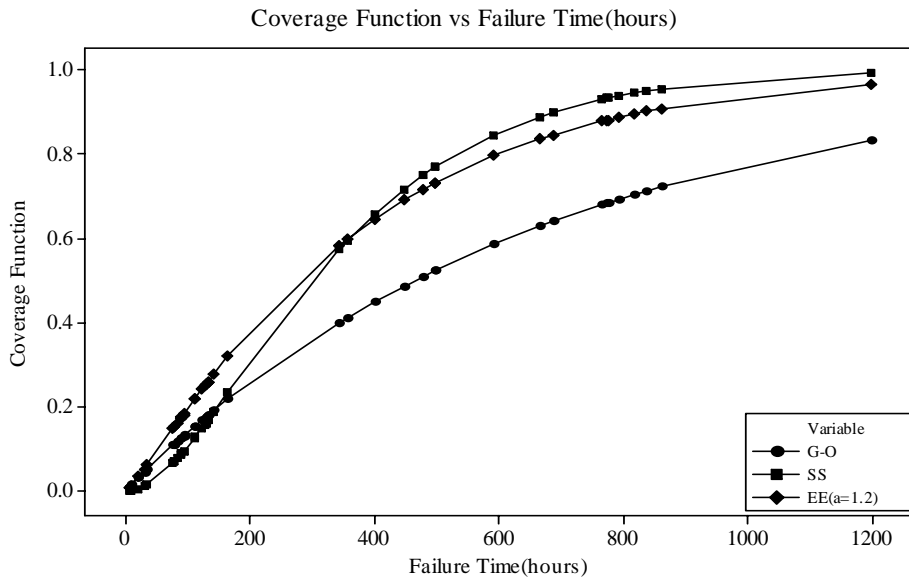
단, G-O: 지수(Goel-Okumoto) 커버리지 모형 , SS: S-형(Yamada-Ohba-Osaki)커버리지모형
 EE: 지수화 지수 커버리지 모형.

<그림 3> 각 모형에 대한 위험함수

임무시간(Mission time)에 대한 신뢰도 즉, 소프트웨어의 고장이 일어나지 않을 확률인 신뢰도(Reliability) $\hat{R}(t | x_n)$ 는 (2.8)식을 이용하면 <그림 7>과 같다. 이 그림에서도 지수화지수 분포 커버리지 모형($a=1.2$)이 상대적으로 효율적인 신뢰도를 나타내고 있다.

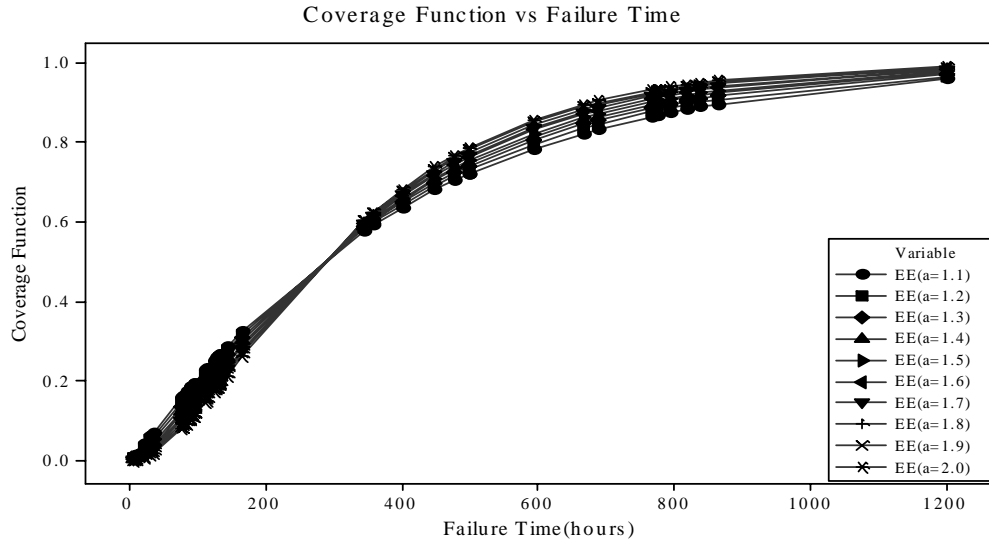


<그림 4> 형상모수 a 에 따른 위험함수 패턴

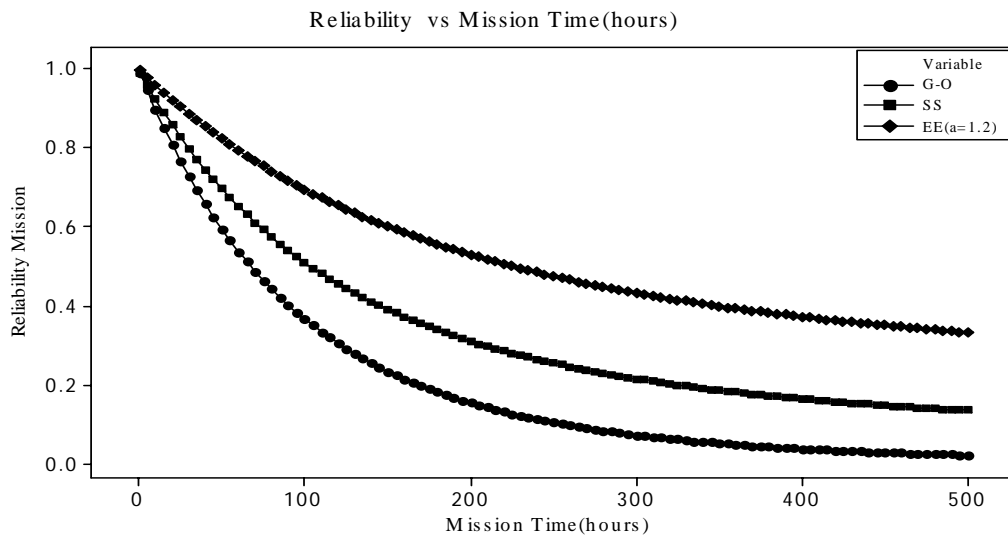


단, G-O: 지수(Goel-Okumoto) 커버리지 모형, SS: S-형(Yamada-Ohba-Osaki)커버리지모형
EE: 지수화 지수 커버리지 모형.

<그림 5> 각 모형에 대한 커버리지함수



<그림 6> 형상모수 a 에 따른 커버리지함수 패턴



단, G-O: 지수(Goel-Okumoto) 커버리지 모형 , SS: S-형(Yamada-Ohba-Osaki)커버리지모형
 EE: 지수화 지수 커버리지 모형.

<그림 7> 각 모형에 대한 신뢰도

VI. 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[21].

본 논문에서는 기존에 존재하는 NHPP모형에서 테스트 커버리지를 포함하는 모형인 ENHPP 모형에 대하여 연구하였다. 잔존 결함 1개당 고장 발생률이 증가추세를 가진 지수화 지수분포를 이용한 지수화 지수 커버리지 모형을 제안하였다. 고장 간격시간으로 구성된 자료를 이용하여 기존의 모형과 지수화 지수 모형에 대하여 최우추정법을 이용하여 모수 추정을 실시하였다. 소프트웨어 고장분석 자료는 S27 자료를 통하여 분석하였다. 효율적인 모형 비교를 위한 편차자승합의 결과는 지수화 지수 커버리지 모형이 기존에 잘 알려진 지수 커버리지 모형이나 S-커버리지 모형보다 우수함을 보이고 있고 콜모고로프 거리의 결과도 지수화 지수 커버리지 모형이 상대적으로 정상성에 가까운 모형으로 간주할 수 있다. 임무시간 비교분석한 결과도 지수화 지수 커버리지 모형 시행이 지수나 S-형 커버리지 모형보다 우수함을 보이고 있다. 분석된 자료에 대한 평가에서도 산술평균 검정과 라플라스 추세 검정을 실시한 결과도 신뢰성장이 되고 있음을 나타내고 있다.

따라서 지수화지수 분포를 이용한 ENHPP 모형도 이 분야에서 효율적으로 이용 할 수 있는 모형이 됨을 알 수 있었다. 향후 이러한 지수화지수 분포를 이용한 베이지안적 접근 방법이나 수리적인 추정과 검정 부분에 대한 수리적인 접근이 기대 된다.

참고문헌

- [1] 김희철 · 최유순 · 박종구, “어랑분포를 이용한 NHPP 소프트웨어 신뢰성장 모형에 관한 연구”, 『한국해양정보통신학회논문지』, 제10권 제1호, 2006, pp.7-14.
- [2] 김희철, “지수화 지수분포에 의존한 NHPP 소프트웨어 신뢰성장 모형에 관한 연구”, 『한국컴퓨터정보학회논문지』, 제11권 제5호, 2006, pp.9-18.
- [3] S. S. Gokhale and K. S. Trivedi, A time/structure based software reliability model, *Annals of Software Engineering*, vol.8, 1999, pp.85-121.

- [4] 김희철, “Burr 커버리지 함수에 기초한 ENHPP 소프트웨어 신뢰성장모형에 관한 연구”, 『한국컴퓨터 정보학회논문지』, 제12권 제4호, 2007, pp.33-42.
- [5] S. S. Gokhale and T. Philip and P. N. Marinos and k. S. Trivedi, Unification of Finite Failure Non-Homogeneous Poisson Process Models through Test Coverage, In Proc. of Int. *Symposium on software Reliability Enginnering*, White Plains. NY, 1996.
- [6] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York, 1981.
- [7] 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), 1979, pp.206-211.
- [8] S. Yamada, M. Ohba and S. Osaki. S-Shaped Reliability Growth Modeling for Software Error Detection, *IEEE Trans. on Reliability*, R-32(5), 1983, pp.475-485.
- [9] R. D. Gupt and D. Kundu, Exponentiated exponential family; an alternative to gamma and Weibull, *Biometrical Journal*, vol.43, 2001, pp.117-130.
- [10] R. D. Gupt and D. Kundu, Discriminating between gamma and generalized exponential distributionl, *Journal of Statistical Computation and Simulation*, vol.74, No.2, 2004, pp.107-121.
- [11] K. Kanoun, J. C. Laprie. *Handbook ofSoftware Reliability Engineering*, M.R.Lyu, Editor, chapter Trend Analysis. McGraw-Hill New York, 1996, pp.401-437.
- [12] 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, 1999, pp.169-175.
- [13] J. F. Lawless. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, New York, 1981.
- [14] S. Selvin. *Moden Applied Biostatistical Methods Using S-Plus*, pages 141-184. Oxford University Press, New York, 1998.