

# 기록값 통계량에 기초한 무한고장 NHPP 소프트웨어 혼합 신뢰성장 모형에 관한 연구

김희철\* · 신현철\*\* · 김경수\*\*

## 요 약

무한고장수를 가진 비동질적인 포아송 과정에 기초한 모형들에서 잔존 결함 1개당 고장 발생률은 일반적으로 상수, 혹은 단조증가 및 단조 감소 추세를 가지고 있다. 본 논문에서는 소프트웨어 신뢰성 모형인 지수분포모형과 어랑 분포 모형을 재조명하고 보다 현실적인 혼합분포모형을 제안 하였다 고장 간격시간으로 구성된 자료를 이용한 모수추정 방법은 최우추정법과 일반적인 수치해석 방법인 이분법을 사용하여 모수 추정을 실시하고 효율적인 모형 선택은 편차자승합(SSE) 및 콜모고로프 거리를 적용하여 모형들에 대한 효율성 입증방법을 설명하였다. 소프트웨어 고장 자료 분석에서는 41개의 고장 수를 가진 S27[12] 자료를 통하여 분석하였다. 이 자료들에서 지수분포 모형과 어랑분포 모형 및 혼합분포 모형의 비교를 위하여 산술적 및 라플라스 검정, 편의 검정 등을 이용하였다.

## Infinite Failure NHPP Software Mixture Reliability Growth Model Base on Record Value Statistics

Hee-Cheul Kim\* · Hyun-Cheul Shin\*\* · Kyung-Soo Kim\*\*

### ABSTRACT

Infinite failure NHPP models presented in the literature exhibit either constant, monotonic increasing or monotonic decreasing failure occurrence rates per fault. In this paper, exponential distribution and Rayleigh distribution model was reviewed, proposes the mixture reliability model, which made out efficiency substituted for situation for failure time

Algorithm to estimate the parameters used to maximum likelihood estimator and bisection method, model selection based on SSE and Kolmogorov distance, for the sake of efficient model, was employed. Analysis of failure using S27 data set for the sake of proposing shape parameter of the mixture distribution was employed. This analysis of failure data compared with the mixture distribution model and the existing model(using arithmetic and Laplace trend tests, bias tests) is presented.

Key words : Software Reliability Model, Mixture Distribution, Record Value Statistics, SSE

---

\* 백석문화대학 컴퓨터정보학부 교수

\*\* 남서울대학교 산업경영공학과

### 1. 서 론

신뢰도에서 관측시간  $(0, t]$  사이에 발견된 고장 수  $N(t)$ 를 모형화 하는데 비동질적 포아송 과정 (Non-Homogeneous Poission Process ; NHPP)이 널리 사용되어 왔다. 이 과정(Process)에서 강도 함수(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 가정한다[1, 2].

NHPP 모형에서 평균값 함수  $m(t)$  (Mean value function)와 강도 함수  $\lambda(t)$ 는 다음과 같은 관계로 표현할 수 있다[3].

$$m(t) = \int_0^t \lambda(s)ds, \quad \frac{dm(t)}{dt} = \lambda(t) \quad (1.1)$$

따라서  $N(t)$ 는 모수  $m(t)$ 를 가진 포아송 확률 밀도 함수(Probability density function ; Pdf)로 알려져 있다.

$$P\{N(t)=n\} = \frac{[m(t)]^n \cdot e^{-m(t)}}{n!}, \quad n = 0, 1, 2, \dots \infty \quad (1.2)$$

이처럼 시간 관련 모형(Time domain models) 들은 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[4]. 유한 고장 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 NHPP 모형들은 무한 값을 가진다고 가정된다.

즉, 고장 시간을 일반화된 순서 통계량(Generalized Order Statistics ; GOS) 모형을 사용하면 유한 고장 모형이 되고, 기록값 통계량(Record Value Statistics ; RVS)모형을 사용하면 무한 고장 모형

이 됨이 알려져 있다[1, 5]. 일반적으로 GOS를 이용한 모형은 소프트웨어 테스트 시점에서 미지의  $N$ 개의 결함들을 가지고 있고 이  $N$ 개의 결함들로부터 임의의 확률 밀도 함수(Pdf)에 따라 발생한  $n$ 개의 순서 통계량이 고장 시점이 된다. 이 모형은 각 수리 시점에서 새로운 결함이 발생하지 않는다는 가정을 한다. 그러나 실제 상황에서는 수리 시점에서도 고장이 발생할 수도 있다. 이러한 상황을 추가하기 위하여 기록 멈춤 통계량(Record breaking statistics)을 사용하는 RVS 모형을 사용할 수 있다[4]. 유한 고장 NHPP 모형에서 충분한 테스트 시간이 주어졌을 때 탐색되어질 수 있는 결함의 기대값을  $\theta$ 라고 표현하고  $F(t)$ 를 분포함수라고 표현하면 유한 고장 NHPP 모형의 평균값 함수와 순간 고장 강도 함수(Instantaneous failure intensity)  $\lambda(t)$  다음과 같이 표현할 수 있다[2, 3].

$$m(t) = \theta F(t), \quad \lambda(t) = \theta F'(t) = \theta f(x) \quad (1.3)$$

또, 순간 고장 강도 함수(Instantaneous failure intensity)  $\lambda(t)$ 는 다음과 같이 변형하여 표기할 수도 있다.

$$\lambda(t) = [\theta - m(t)] \frac{F'(t)}{1 - F(t)} = [\theta - m(t)] h(t) \quad (1.4)$$

단,  $h(t) = \frac{f(t)}{1 - F(t)}$ 는 위험 함수(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 결함들이 탐색되어 제거되기 때문에 감소성을 가진다. 따라서  $\lambda(t)$ 는  $h(t)$ 의 값에 따라 달라지며 상수, 증가, 감소 혹은 증가하다가 감소하는 패턴을 가질 수 있다. 이 분야의 기본적 모형인 Goel-Okumoto 모형은  $h(t)$ 가 정수 패턴을 가짐으로서 시점  $t$ 에 독립이고 잘 알려진 S 모형

(Yamada, Ohba-Osaki 모형)은 증가 패턴을 가진 다[5, 6]. 반면에 무한 고장 NHPP 모형의 평균값 함수와 간 고장 강도 함수(혹은 ROCOF)는 다음과 같이 알려져 있다[1, 7].

$$m(t) = -\ln(1-F(t)) \tag{1.5}$$

$$\lambda(t) = m'(t) = f(t)/(1-F(t)) = h(t) \tag{1.6}$$

한편, 시간  $(0, t]$ 까지 조사하기 위한 시간 절단(Time truncated) 모형은  $n$  번째까지 고장 시점 자료를

$$x_k = \sum_{i=1}^k t_k \quad (k=1, 2, \dots, n; 0 \leq x_1 \leq x_2 \leq \dots \leq x_n) \tag{1.7}$$

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 모형일 경우에 데이터 집합  $D_{x_n}$ 은  $\{x_1, x_2, \dots, x_n\}$ 으로 구성된다. 이 시간 절단 모형에서의 우도 함수는 다음과 같이 알려져 있다[1, 3, 7].

$$L_{NHPP}(\theta | D_t) = \left( \prod_{i=1}^n \lambda(x_i) \right) \exp(-m(t)) \tag{1.8}$$

단,  $\theta$ 은 미지의 모수 집합을 의미하고 우도 함수 식 (1.8)에서  $t$ 을  $x_n$ 으로 대치하면 유사한 형태의 고장 절단 모형의 우도 함수가 된다[3]. 식 (1.8)과 식 (1.5), 식 (1.6)을 연관하면 무한 NHPP 모형에 대한 우도 함수는 다음과 같은 형태로 표현할 수 있다[1, 4].

$$L_{NHPP_{NFB}}(\theta | D_t) = \prod_{i=1}^n \left( \frac{f(x_i)}{1-F(x_i)} \right) (1-F(x_n)) \tag{1.9}$$

$$= \left( \prod_{i=1}^n h(x_i) \right) (1-F(x_n))$$

따라서 본 연구에서는 무한 고장 NHPP 모형에 대하여 수명 분포가 감마 분포의 특수한 형태인 지수 및 어랑 분포모형과 이들의 혼합모형을 제시하여 신뢰성 척도를 추정하고 이를 바탕으로 소프트웨어 신뢰성에 대한 특성을 알아보고자 한다.

## 2. 관련 연구

### 2.1 기록값 통계량과 무한 고장 NHPP

RVS 모형에서 고장 시점은 밀도 함수  $f(S|\theta)$ 에 의해 동일하고 독립인 분포(Identically Independent Distribution ; IID)에서 발생된 확률변수의 기록값들(Record values)이라고 가정된다. 여기에서  $\theta$ 은 모수의 공간을 의미하고  $S_1, S_2, \dots$ 은 동일하고 독립적인  $f(S|\theta)$ 분포(i.i.d)에서 발생하는 확률 변수라고 표현하면 기록값 열(Sequence of record values)  $\{X_n\}_{n \geq 1}$ 은 다음과 같이 정의된다[1, 4].

$$R_1 = 1$$

$$R_{k+1} = \min \{ i : i > R_k, S_i > S_{R_k} \} \quad \text{단, } k=1, 2, 3, \dots$$

$$X_n = S_{R_n} \quad \text{단, } n \geq 1 \tag{2.1}$$

기록값  $X_1, X_2, \dots$ 의 관측값(Observation)을  $x_1 < x_2 < \dots$ 으로 표시한다. Glick(1978)은 이러한 기록값 열은 무한하게 됨을 보였다. 따라서 고장수가  $t \rightarrow \infty$ 가 됨에 따라 무한(Unbounded)할 경우를 모형화할 수 있다고 하였고[4] RVS 모형과 NHPP 모형을 연관시키는 정리는 Dwass(1964) or Resnick(1987)에 의하여 다음과 같다[1, 4].

$F$ 를  $R^+$  (실수영역) 상에서 정의된 연속 분포 함수라고 했을 때  $(0, t]$  상에 있는 값으로서 만들어진 기록값들은  $(0, t]$  상에 있는 다음과 같은 평균 척도(Mean measure)를 가진 NHPP의 점들(Points)과 같다[1, 4].

$$m(t) = -\ln(1-F(t)) \tag{2.2}$$

따라서 (1.1)과 연관하면 NHPP의 ROCOF  $\lambda(t) = m'(t) = f(t)/(1-F(t))$ . 즉,  $F$ 의 위험함수가 되는 것이다.

### 2.2 지수분포 및 어랑 분포 모형

본 절에서는 지수분포와 어랑분포(Erlang distribution)[1, 8]를 이용한 무한 NHPP 신뢰성 모형에 대하여 최우 추정법(Maximum likelihood estimation)을 이용한 모수 추정에 대하여 살펴보고자 한다. 우선, 수명 분포가 감마 분포  $Gamma(\alpha, \beta)$ 를 적용 할 경우 감마분포의 확률 밀도 함수는

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

이고 분포함수는  $\alpha$  값이 양의 정수일 때

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

가 된다[1]. 따라서 식 (1.9)과 관련하여 무한 고장 NHPP 우도함수는 다음과 같다.

$$L_{Ga}(\alpha, \beta | D_n) = \left( \prod_{i=1}^n \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{x_i^{\alpha-1}}{\sum_{j=0}^{\alpha-1} \frac{(\beta x_i)^j}{j!}} \right) e^{-\beta x_n} \sum_{j=0}^{\alpha-1} \frac{(\beta x_n)^j}{j!} \quad (2.3)$$

최우 추정법을 이용하기 위한  $Gamma(\alpha, \beta)$  모형 로그 우도 함수(Log likelihood function)는 식 (1.11)과 관련하여 다음과 같이 유도된다.

$$\ln L(\alpha, \beta | D_n) = n\alpha \ln \beta - n \ln \Gamma(\alpha) + (\alpha-1) \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \left( \ln \sum_{j=0}^{\alpha-1} \frac{(\beta x_i)^j}{j!} \right) - \beta x_n + \ln \left( \sum_{j=0}^{\alpha-1} \frac{(\beta x_n)^j}{j!} \right) \quad (2.4)$$

본 연구에서는 형상모수  $\alpha=1$ 인 경우인 지수 분포와  $\alpha=2$ 인 어랑 분포(Erlang distribution)[1, 7, 8, 9]를 고려하고자 한다. 이 어랑 분포는 감마 분포의 특수한 경우로서 형상모수  $\alpha$ 가 정수인 경우를 의미한다.

$\alpha=1$ 인 경우에는 식 (2.4)에서  $\beta$ 에 대하여 편미분하여 다음과 같은 식을 만족하는  $\hat{\beta}_{MLE}$ 을 다음과 같이 추정할 수 있다.

$$\frac{\partial \log L(1, \beta | D_n)}{\partial \beta} = n \ln \beta - \beta x_n = 0 \quad (2.5)$$

식 (2.5)를 계산하면  $\hat{\beta}_{MLE} = \frac{n}{x_n}$ 이 된다.

$\alpha=2$ 인 경우에 식 (2.4)에서  $\beta$ 에 대하여 다음과 같은 비선형식을 만족하는  $\hat{\beta}_{MLE}$ 을 수치 해석적 방법으로 계산 할 수 있다.

$$\frac{\partial \log L(2, \beta | D_n)}{\partial \beta} = \frac{2n}{\beta} - \sum_{i=1}^n \left( \frac{x_i}{1 + \beta x_i} \right) - x_n + \frac{x_n}{1 + \beta x_n} = 0 \quad (2.6)$$

### 3. 제안된 혼합모형을 이용한 신뢰성 모형

부품이나 시스템집단의 수명분포가 서로 다른  $k$ 개의 부분집단으로 구성된 경우에 적용가능한 모형이다[9, 10]. 예를 들어 전체집단이 결합부품이나 제조상의 결함을 지닌 초기고장 집단과 정상집단의 두 그룹으로 구성되는 경우도 혼합모형에 해당된다. 이러한 혼합모형은 분포함수와 확률밀도함수는 여러분포의 가중합 형태로 다음과 같이 표현된다.

$$F(t|\theta, p) = \Pr(T < t) = \sum_{i=1}^k p_i F_i(t|\theta_i) \quad (3.1)$$

$$f(t|\theta, p) = \sum_{i=1}^k p_i f_i(t|\theta_i) \quad (3.2)$$

위 식에서  $\theta$ 은 미지의 모수 집합을 의미하고 모집단의  $f_i(t|\theta_i), i=1, \dots, k$ 는 부분집단  $i$ 에 속하는 부품의 고장시간에 대한 확률밀도함수이며 혼합모수(Mixing parameter) $p=(p_1, \dots, p_k)$ 는 다음을 만족한다.

$$\sum_{i=1}^k p_i = 1, \quad 0 \leq p_i \leq 1 \quad (3.3)$$

본 논문에서는  $k=2(p_1=p, p_2=1-p)$ 인 경우를 고려하고자 한다. 따라서 혼합모형의 분포함수와

확률 밀도함수는 다음과 같이 변형된다.

$$F(t|\Theta, p) = p \cdot F_1(t|\theta_1) + (1-p) \cdot F_2(t|\theta_2) \quad (3.4)$$

$$f(t|\Theta, p) = p \cdot f_1(t|\theta_1) + (1-p) \cdot f_2(t|\theta_2) \quad (3.5)$$

무한고장 상황에서의 혼합모형에 대한 평균값 함수는 식 (7)과 식 (8)를 연관하여 평균값 함수  $m_{Mfx}(t|\Theta, p)$ 와 고장발생률  $\lambda_{Mfx}(t|\Theta, p)$ 을 다음과 같이 유도 할 수 있다.

$$m_{Mfx}(t|\Theta, p) = -\ln(1 - F_{Mfx}(t|\Theta, p)) \quad (3.6)$$

$$\lambda_{Mfx}(t|\Theta, p) = \frac{f_{Mfx}(t|\Theta, p)}{1 - F_{Mfx}(t|\Theta, p)} \quad (3.7)$$

즉, 지수분포와 어랑분포의 혼합밀도 함수와 혼합 분포함수는 다음과 같이 유도된다.

$$f_{Mfx}(t|b_1, b_2, p) = p \cdot (b_1 e^{-b_1 t} - b_2^2 t e^{-b_2 t}) + b_2^2 t e^{-b_2 t}$$

$$F_{Mfx}(t|b_1, b_2, p) = p \cdot (e^{-b_2 t} (1 + b_2 t) - e^{-b_1 t}) - e^{-b_2 t} (1 + b_2 t) + 1$$

단,  $p, b_1, b_2, t \geq 0$ . 시간  $(0, t]$ 까지 조사하기 위한 시간 절단(Time truncated)모형은  $n$  번째 까지 고장시점 자료를

$$x_k = \sum_{i=1}^k t_k \quad (k = 1, 2, \dots, n; 0 \leq x_1 \leq x_2 \leq \dots \leq x_n) \quad (3.8)$$

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 식 (3.6), 식 (3.7)를 연관하여 지수분포와 어랑 분포의 혼합밀도함수 이용한 로그우도함수를 정리 하면 다음과 같이 표현된다.

$$\ln L_{Mfx}(D_{x_n}|p, b_1, b_2) = \quad (3.9)$$

$$\sum_{i=1}^n \ln \left[ \frac{p(b_1 e^{-b_1 x_i} - b_2^2 x_i e^{-b_2 x_i}) + b_2^2 e^{-b_2 x_i}}{e^{-b_2 x_i} (1 + b_2 x_i) - p e^{-b_2 x_i} (1 + b_2 x_i) + p e^{-b_1 x_i}} \right]$$

$$+ \ln [e^{-b_2 x_n} (1 + b_2 x_n) - p e^{-b_2 x_n} (1 + b_2 x_n) + p e^{-b_1 x_n}]$$

척도(scale) 모수  $b_1$  값은 상수(사전에 알고 있는 경

우)라고 가정 했을 때 최우추정법을 이용하기 위하여 위 우도함수식을  $b_2$ 와  $p$ 에 대하여 편미분을 하면 다음과 같은 식을 유도 할 수 있다.

$$\frac{\partial \ln L_{Mfx}(D_{x_n}|b_1, b_2, p)}{\partial p} \quad (3.10)$$

$$= \sum_{i=1}^n \frac{b_1 e^{-b_1 x_i} - b_2^2 x_i e^{-b_2 x_i}}{p(b_1 e^{-b_1 x_i} - b_2^2 x_i e^{-b_2 x_i}) + b_2^2 e^{-b_2 x_i}}$$

$$- \sum_{i=1}^n \frac{e^{-b_1 x_i} - e^{-b_2 x_i} (1 + b_2 x_i)}{e^{-b_2 x_i} (1 + b_2 x_i) - p e^{-b_2 x_i} (1 + b_2 x_i) + p e^{-b_1 x_i}}$$

$$+ \frac{e^{-b_1 x_n} - e^{-b_2 x_n} (1 + b_2 x_n)}{e^{-b_2 x_n} (1 + b_2 x_n) - p e^{-b_2 x_n} (1 + b_2 x_n) + p e^{-b_1 x_n}} = 0$$

$$\frac{\partial \ln L_{Mfx}(D_{x_n}|b_1, b_2, p)}{\partial b_2} \quad (3.11)$$

$$= \sum_{i=1}^n \frac{(1-p) b_2 x_i e^{-b_2 x_i} (2 - b_2 x_i)}{p(b_1 e^{-b_1 x_i} - b_2^2 x_i e^{-b_2 x_i}) + b_2^2 e^{-b_2 x_i}}$$

$$- \sum_{i=1}^n \frac{(p-1) e^{-b_2 x_i} x_i^2 b_2}{e^{-b_2 x_i} (1 + b_2 x_i) - p e^{-b_2 x_i} (1 + b_2 x_i) + p e^{-b_1 x_i}}$$

$$+ \frac{(p-1) e^{-b_2 x_n} x_n^2 b_2}{e^{-b_2 x_n} (1 + b_2 x_n) - p e^{-b_2 x_n} (1 + b_2 x_n) + p e^{-b_1 x_n}} = 0$$

식 (3.10)과 식 (3.11)를 비선형 연립방정식(수치 해석적 방법)을 이용하여 풀면 최우추정치  $\hat{p}_{MLE}$ 와  $\hat{b}_{2MLE}$ 의 값을 구 할 수 있다. 본 논문에서는 혼합분포의 모수  $b_1$  값은 0.1, 0.3, 0.5, 0.7, 0.9와 같은 상수인 경우를 본 논문에서는 제시하고자 한다.

#### 4. 소프트웨어 고장 자료 분석

이 장에서 실제 고장자료인 1197.95시간단위에서 41개의 소프트웨어 고장자료 S27[11]를 적용하여 지수분포와 어랑 분포의 모수에 따른 혼합 신뢰도 모형을 분석하고자 한다. 자료는 <표 1>에 요

〈표 1〉 고장 자료

Failure number	Failure Time(hours)	Failure Interval (hours)	failure times $\times 10^{-2}$
1	5.65	2.649	0.0565
2	8.92	3.271	0.0892
3	20.29	11.37	0.2029
4	29.96	9.665	0.2996
5	34.72	4.76	0.3472
6	75.95	41.235	0.7595
7	78.17	2.221	0.7817
8	78.63	0.454	0.4863
9	83.02	4.397	0.8302
10	89.11	6.092	0.8911
11	89.8	0.69	0.898
12	92.86	3.056	0.8386
13	93.66	0.8	0.9366
14	110.66	16.995	1.1066
15	111.99	1.333	1.1199
16	122.55	10.557	1.2255
17	127.05	4.5	1.2705
18	128.71	1.667	1.2871
19	128.99	0.278	1.2899
20	131.77	2.778	1.3177
21	131.83	0.061	1.3183
22	141.71	9.883	1.4171
23	164.21	22.5	1.6421
24	342.85	178.638	3.4285
25	356.14	13.294	3.5614
26	399.14	43	3.9914
27	446.49	47.35	4.4649
28	476.64	30.15	4.7664
29	497.14	20.5	4.9714
30	497.66	0.517	4.9766
31	591.16	93.5	5.9116
32	665.64	74.483	6.6564
33	686.44	20.8	6.8644
34	765.94	79.5	7.6594
35	772.98	7.033	7.7298
36	774.94	1.967	7.7494
37	791.56	16.617	7.9156
38	815.98	24.417	8.1598
39	837.15	21.167	8.3715
40	861.95	24.8	8.6195
41	1197.95	336	11.9795

약되었다. 신뢰모형들을 분석하기 위하여 우선 자료에 대한 추세 검정이 선행 되어야 한다[3, 11]. 추

세 분석에는 산술평균 검정(Arithmetic mean test)과 라플라스 추세 검정(Laplace trend test) 등이 있다. 산술 평균 추세 검정은 다음과 같이 정의 된다.

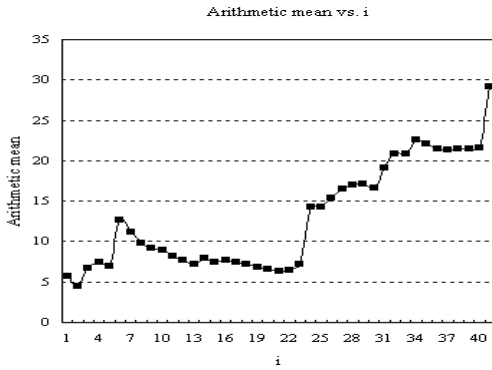
$$\tau(i) = \frac{1}{i} \sum_{j=1}^i t_j \tag{4.1}$$

식 (4.1)에서  $t_j (j = 1, 2, \dots, i)$ 들은 관찰된 고장 간격 시간을 의미하기 때문에  $\tau(i)$ 의 값이 증가하면 신뢰 성장을 의미한다. 또한 라플라스 검정은 식 (4.2)과 같이 계산되며,  $\mu(i)$ 의 값이 감소이면 고장 강도가 감소함으로 인해서 신뢰 성장이 나타난다.

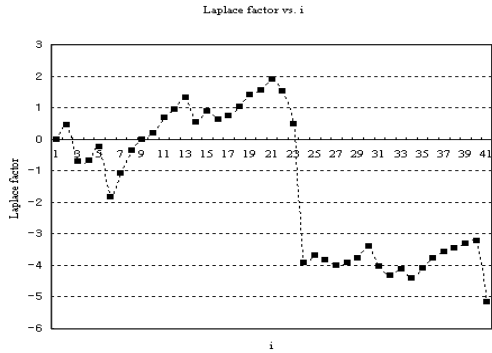
$$\mu(i) = \frac{\frac{1}{i-1} \sum_{n=1}^{i-1} \sum_{j=1}^n t_j - \frac{\sum_{j=1}^i t_j}{2}}{\sum_{j=1}^i t_j \sqrt{\frac{1}{12(i-1)}}} \tag{4.2}$$

이러한 검정을 실시한 결과 (그림 1)에서 산술 평균 검정결과 고장시간이 증가함에 따라 산술 평균이 거의 증가 추세를 보이고 있으므로 신뢰성장(Reliability growth)이 되고 있음을 나타내고 있고 (그림 2)에 나타난 라플라스 추세 검정의 결과도 라플라스 요인(Factor)이 고장시간이 증가함에 따라서 감소하기 때문에 역시 신뢰성장(Reliability growth)이 되고 있음을 나타내고 있다. 따라서 이 자료를 가지고 신뢰성장 모형을 제시하는 것이 효율적임을 시사하고 있다[3, 11]. 본 논문에서는 모수추정을 용이하게 하기 위하여 실제자료에서 수치 변환된 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  $10^{-2}$ 와 10을, 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  $b_1 = 1$  이상인 경우는 비



(그림 1) 산술평균 검정



(그림 2) 라플라스 추세 검정

효율적이고 근이 수렴되지 않아 본 논문에서는 고려대상에서 제외 시켰다. 지수 및 어랑 모형과 혼합 신뢰도 모형에 대한 모수의 추정값들의 결과는 <표 2>에 요약되었다. 이 분야에서는 일반적으로 모형 선택의 하나의 방법으로 다음과 같은 편차자승합(SSE[6])을 이용할 수 있다.

$$\sum_{i=1}^n (n_i(x_i) - \hat{m}(x_i))^2 \quad (4.3)$$

단,  $n_i(x_i)$ 은  $x_i$ 의 시점에서 고장의 누적수를 의미하고,  $\hat{m}(x_i)$ 는 시점  $x_i$ 에서 추정된 누적 고장수를 의미하며, 편차 자승합이 작으면 상대적으로 효율적인 모형이 된다. 본 논문은 효율적 모형의 선

<표 2> 각 모형의 모수 추정값

Model	MLE
Exponential Model	$\hat{\beta}_{MLE} = 0.39812$
Erlang Model	$\hat{\beta}_{MLE} = 4.09341$
Mixture Model( $b_1 = 0.1$ )	$\hat{p}_{MLE} = 0.38351$ $\hat{b}_{2,MLE} = 1.8323$
Mixture Model( $b_1 = 0.3$ )	$\hat{p}_{MLE} = 0.43218$ $\hat{b}_{2,MLE} = 7.36352$
Mixture Model( $b_1 = 0.5$ )	$\hat{p}_{MLE} = 0.65335$ $\hat{b}_{2,MLE} = 0.184233$
Mixture Model( $b_1 = 0.7$ )	$\hat{p}_{MLE} = 0.78653$ $\hat{b}_{2,MLE} = 0.20197$
Mixture Model( $b_1 = 0.9$ )	$\hat{p}_{MLE} = 0.83419$ $\hat{b}_{2,MLE} = 0.05543$

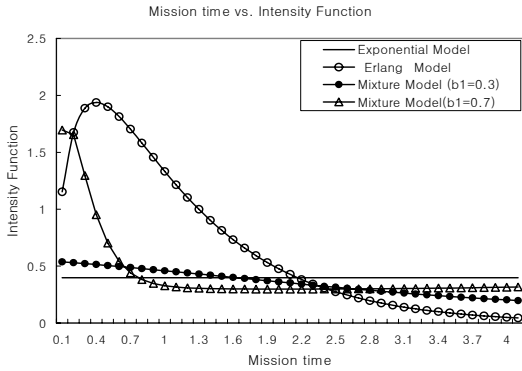
택을 위하여 SSE를 적용하여 효율적인 모형을 선택하고자 한다. 편차 자승합의 값의 결과는 <표 3> 요약 되었으며, 이 표에서 지수모형이나 어랑 모형에 비해 혼합모형( $b_1 = 0.7$ ,  $b_1 = 0.3$ 인 경우)이 상대적으로 효율적 모형으로 나타나고 있지만, 그 외의 경우는 비효율적 모형으로 나타나고 있다.

<표 3> 모형들에 대한 SSE의 값

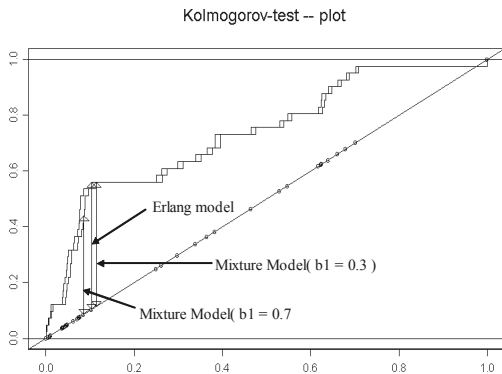
Model	SSE
Exponential Model	20508.19
Erlang Model	<b>4577.912</b>
Mixture Model( $b_1 = 0.1$ )	16671.74
Mixture Model( $b_1 = 0.3$ )	<b>4262.480</b>
Mixture Model( $b_1 = 0.5$ )	19584.82
Mixture Model( $b_1 = 0.7$ )	<b>3480.403</b>
Mixture Model( $b_1 = 0.9$ )	17778.04

(그림 3)에서는 임무시간(Mission Time)에 대한 강도함수(Intensity Function)의 형태를 나타내었다.

이 그림에서 강도함수(위험함수)는 지수모형은 시간에 독립적으로 일정하고 어랑분포나  $b_1 = 0.3$ 인 혼합모형은 시간이 흐름에 따라 급격이 감소하는 형태이지만  $b_1 = 0.7$ 인 혼합모형은 완만한 형태를 보여주고 있다. 예측 오류(Prediction error)의 비정상성(Nonstationarity)에 대한 측도는 Kolmogorov 거리(Distance)[3, 12]로 측정되는데 이 거리가 클수록 상대적으로 비정상성을 내포하고 있다.



(그림 3) 모형에 대한 강도함수의 형태



(그림 4) 모형에 대한 Kolmogorov 검정 그림

(그림 4)는 S-Plus 소프트웨어[13]를 이용하여 Kolmogorov 검정에 대한 그림을 보여주고 있고 이 그림에서도 혼합 신뢰 모형의 모수( $b_1 = 0.7, 0.3$ )에 따른 신뢰혼합모형들이 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 혼합 신뢰 모형의 모수( $b_1 = 0.7, 0.3$ )에 따른 신뢰도혼합모형들이 상대적으로 비정상성 속성이 덜 내포하고 있음을 알 수 있다.

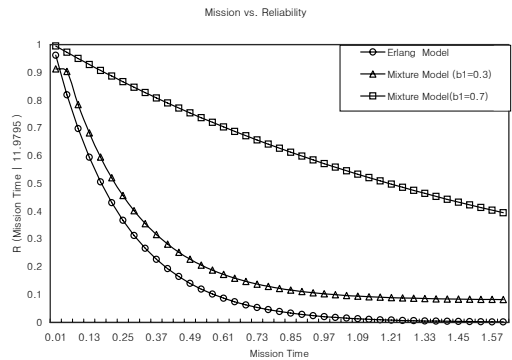
<표 4> 모형에 대한 Kolmogorov 거리

Model	Kolmogorov distance
Erlang Model	0.457326
Mixture Model ( $b_1 = 0.3$ )	0.4463762
Mixture Model ( $b_1 = 0.7$ )	0.228774

임무시간(Mission time)에 대한 즉, 소프트웨어의  $(t, t+x]$ (단,  $x$ 는 임무시간(Mission Time)사이에 고장이 일어나지 않을 확률인 신뢰도(Reliability)  $\hat{R}(t | x_n)$ 는 다음과 같이 정의 된다[3, 7].

$$R(x | x_n) = \exp\{-[m(x_n + x) - m(x_n)]\} \quad (4.4)$$

단,  $x_n \geq 0, x \geq 0$  이러한 신뢰도에 대한 그림은 (그림 5)에 나타내었다. 이 그림에서도 어랑분포나  $b_1 = 0.3$ 인 혼합모형은 시간이 흐름에 따라 급격이 감소하는 형태이지만  $b_1 = 0.7$ 인 혼합모형은 완만한 형태를 보여주고 있다.



(그림 5) 모형에 대한 신뢰도



따라서 혼합 신뢰도 모형( $b_1 = 0.7, 0.3$ )들이 상대적으로 효율적인 신뢰도를 나타내고 있다.

## 5. 결 론

본 논문에서는 기존의 단일 분포보다 더 현실적인 혼합 분포를 이용한 혼합 신뢰도 모형을 제안하였다. 고장 간격시간으로 구성된 41개의 소프트웨어 고장자료 S27[11]를 이용하여 단일 분포인 지수와 어랑분포와 이 분포들의 혼합으로 이루어진 신뢰도 모형에 대하여 최우 추정법을 이용하여 모수 추정을 실시하였다.

효율적인 모형 비교를 위한 편차자승합의 결과는 혼합모형이 지수나 어랑분포보다 더 우수한 모형이 있을 수 있음을 보여주고 있으며 동시에 콜모고로프 거리의 결과도 혼합모형이 상대적으로 정상성에 가까운 모형이 될수 있음을 알 수 있다. 분석된 자료에 대한 평가에서도 산술평균 검정과 라플라스 추세 검정을 실시한 결과도 신뢰성장이 되고 있음을 나타내고 있다. 신뢰도에 대한 평가도 역시 완만하게 감소하는 혼합 모형이 신뢰도가 높음을 알 수 있다. 향후 기존의 사전정보(혼합모형중에서 모수)를 이용한 수리적인 베이저안 분석을 기대한다.

## 참 고 문 헌

- [1] J. F. Lawless, "Statistical Models and Methods for Lifetime Data", John Wiley & Sons, New York, 1981.
- [2] 김희철, "일반화 감마분포를 이용한 NHPP 소프트웨어 신뢰도 모형에 관한 연구", 한국컴퓨터정보학회논문지, 제10권, 제6호, pp. 27-35, 2005.
- [3] S. S. Gokhale and K. S. Trivedi, "A time/structure based software reliability model", Annals of Software Engineering, Vol. 8, pp. 85-121. 1999.
- [4] L. Kuo and T. Y. Yang, "Bayesian Computation of Software Reliability", Journal of the American Statistical Association, Vol. 91, pp. 763-773, 1996.
- [5] S. Yamada, M. Ohba, and S. Osaki, "S-Shaped Reliability Growth Modeling for Software Error Detection", IEEE Trans. on Reliability, Vol. R-32, No. 5, pp. 475-485, Dec. 1983.
- [6] 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.
- [7] J. D. Musa, A. Iannino, and K. Okumoto, "Software Reliability : Measurement, Prediction, Application", McGraw Hill, New York, 1987.
- [8] V. K. Rohatgi, "Statistical inference", JOHN WILEY & SONS, INC, New York, pp. 398-416, 1984.
- [9] 정해성 외 2인, "신뢰성 시험분석평가", 영지문화사, pp. 99-101, 2005.
- [10] D. A. Berry and D. K. Stangl, "Bayesian Biostatistics", MARCEL DEKKER INC, pp. 601-616, 1996.
- [11] K. Kanoun and J. C. Laprie, Handbook of Software Reliability Engineering, M.R.Lyu, Editor, chapter Trend Analysis, pp. 401-437. McGraw-Hill, New York, NY, 1996.
- [12] S. Brocklehurst and B. Littlewood, Handbook of Software Reliability Engineering, M. R. Lyu, Editor, chapter Techniques for Prediction Analysis and Recalibration, pp. 119-166.
- [13] S. Selvin, Modern Applied Biostatistical Methods Using S-Plus, Oxford University Press, New York, pp. 141-184, 1998.



**김희철**

1992년 동국대학교 통계학과  
(이학석사)  
1998년 동국대학교 통계학과  
(이학박사)  
2005년~현재 남서울대학교  
산업경영공학과 전임강사



**김경수**

2001년 순천향대학교 전산학과  
(공학박사)  
2005년~2006년 Virginia  
Common-wealth  
University(객원교수)  
1998년~현재 백석문화대학 컴퓨터정보학부 조교수



**신현철**

2002년 원광대학교 컴퓨터공학과  
(공학박사)  
2004년 (주)아이비루션 자문위원  
2007년 과학기술정보 클러스터  
자문교수(KiSTi)  
1994년~현재 백석문화대학  
컴퓨터정보학부 교수