

Burr 커버리지 함수에 기초한 ENHPP소프트웨어 신뢰성장모형에 관한 연구

김희철*

The Study for ENHPP Software Reliability Growth Model based on Burr Coverage Function

Kim Hee Cheul *

요 약

소프트웨어 제품의 정확한 인도시기를 예측하거나 효율성 및 신뢰성을 예측하기 위해서는 소프트웨어 테스트 과정에서 중요한 요소인 테스트 커버리지를 이용하면 보다 효율적인 테스트 작업을 할 수 있다. 이런 모형을 ENHPP모형이라고 한다. 본 논문에서는 기존의 소프트웨어 신뢰성 모형인 지수 커버리지 모형과 S-커버리지 모형을 재조명하고 이 분야에 적용될수 있는 Burr 분포에 기초한 Burr 커버리지 모형을 제안하였다. 고장 간격 시간으로 구성된 자료를 이용한 모수추정 방법은 최우추정법 과 일반적인 수치해석 방법인 이분법을 사용하여 모수 추정을 실시하고 효율적인 모형 선택은 편차자승합(SSE) 및 콜모고로프 거리를 적용하여 모형들에 대한 효율적인 모형선택도 시도 하였다. NTDS 자료를 사용한 임무시간 비교분석한 결과 Burr 커버리지 모형 시행 이 지수나 S-형 모형보다 우수함을 보이고 있다. 이 자료들에서 기존의 모형과 Burr 커버리지 모형의 비교를 위하여 산술적 및 라플라스 검정, 편의 검정등을 이용하였다.

Abstract

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-shaped model was reviewed, proposes the Kappa coverage model, which made out efficiency application for software reliability. Algorithm to estimate the parameters used to maximum likelihood estimator and bisection method, model selection based on SSE statistics and Kolmogorov distance, for the sake of efficient model, was employed. From the analysis of mission time, the result of this comparative study shows the excellent performance of Burr coverage model rather than exponential coverage and S-shaped model using NTDS data. This

• 제1저자 : 김희철
• 접수일 : 2007. 7.6, 심사일 : 2007. 8.21, 심사완료일 : 2007. 9.20
* 남서울대학교 산업경영공학과 전임강사

analysis of failure data compared with the Kappa coverage model and the existing model(using arithmetic and Laplace trend tests, bias tests) is presented.

- ▶ Keyword : 테스트 커버리지(Test Coverage), 확장된 비동질적인 포아송 과정(Enhanced non-homogeneous poisson process ; ENHPP), Burr 분포(Burr Distribution), 콜모고로프 거리(Kolmogorov Distance)

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 사용되어 왔다. 이러한 NHPP모형은 강도함수(Intensity function)와 평균값함수(Mean value function)에 의존한다.

소프트웨어 고장시간은 수명자료이기 때문에 비음(Nonnegative)의 값을 가진다. 따라서 이 분야에서는 주로 지수분포, 와이블분포, 감마분포 등 일반화 감마 분포(혹은 어랑분포)가 많이 사용되어 왔다[1, 2].

이 범주에서 지금까지 알려진 모형들은 Goel-Okumoto 모형, Weibull 모형 그리고 Cox-Lewis 모형등이 있는데 이 모형들에 대한 강도함수는 각 각 시간에 의존한 함수, 역(Power) 함수, 대수선형(Log-linear) 함수를 가정하였다[3, 4].

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[4, 5].

본 논문에서는 분포함수가 효율적으로 사용할 수 있는 2모수 Burr 분포[6, 7, 8]를 이용한 Burr 커버리지 함수를 제안하여 모형의 효율성과 그 특성을 알아보고자한다.

본 논문의 2장에서는 관련연구로서 테스트 커버리지와 유한 고장 ENHPP 모형에 대하여 서술하였고 3장에서는 Burr 커버리지 함수에 대하여 요약하였고 4장에서는 모수 추정방법에 대하여 나열하였으며 5장에서는 수치적인 예로 실제 고장자료를 이용하여 각 모형에 대한 모수추정 및 모형비교를 실시하였으며 마지막으로 6장에서는 결론을 나열하였다.

II. 관련 연구

2.1 테스트 커버리지

프로그램 검사 능력과 테스트 커버리지(Coverage)는 관련성 있는 개념이다.

검사능력은 하나의 프로그램 상에서 검사할 수 있는 용이성이라면 커버리지는 프로그램 상에서 숨겨진 잠재적인 모든 오류 위치들을(All potential fault-sites) 충분히 찾아낼 수 있는 척도(Measure)를 제공한다. 여기서 잠재적인 오류 사이트들은 구조적으로나 기능적으로 명백히 설명할수 있는 프로그램의 요소라고 정의 할 수 있고 이런 요소들은 합당하게 설계된 테스트를 통해 입증성이나 타당성이 요구되는 보전성(Integrity)을 가지고 있어야 한다[4, 5].

따라서 프로그램 검사 능력과 테스트 커버리지는 다음과 같은 요인에 의해 영향을 받는다.

- (1) 프로그램의 복잡성, (2) 소프트웨어 개발 이론, (3) 사용된 소프트웨어 도구(Tools), (4) 테스트 질(Quality), (5) 테스트의 효율성.

테스트 커버리지에 대한 연구는 이미 여러 학자들에 의하여 연구되었다[9, 10]. 그러나 소프트웨어 개발자나 사용자들은 테스트 커버리지를 측정 할 수 있는 효율적인 평가 도구들을 원하고 있다. 이러한 도구들은 소프트웨어 제품의 인도시기(Release times) 뿐만 아니라 효율적이고 질적인 테스트가 이루어 져야 한다.

테스트 커버리지는 테스트 작업하에서 소프트웨어 제품에 숨겨진 잠재적인 모든 오류 사이트들을 얼마나 찾아낼 수있는지를 평가하는 척도이다. 그러므로 잠재오류 사이트와 오류 사이트는 서로 민감성을 가지고 있다.

따라서 테스트 커버리지는 소프트웨어 제품 과 테스트 지침이 설정되었을 때 고려대상에서 존재하는 잠재적인 총 오류 사이트수에 대한 테스트 하는 동안에 숨겨진 잠재적인 오류 사이트 수의 비(Ratio)라고 정의되어 진다[4, 5].

2.2 소프트웨어 신뢰성에 대한 테스트 커버리지

테스트 커버리지를 유한고장 비동질적 포아송 과정 NHPP에 적용 시킬 수 있다. 이러한 적용 모형을 확장된 비동질적 포아송 과정(ENHPP)이라고 부른다. 이러한 ENHPP 모형의 오류탐색비율은 숨겨진 잠재 오류 사이트 의비와 남아있는 오류함의 기대수와의 곱에 비례한다. 이러한 ENHPP 모형은 다음과 같은 가정을 한다[4].

- (가정 1) 오류들은 모든 잠재 오류 사이트에 대하여 균 일하게(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 오 류는 발생되지 않는다.(이 가정은 유한 고장 NHPP 모형과 유사하다)
- (가정 4) 커버리지는 테스트 시간에 대하여 연속적인 단 조 비감소 함수(Monotonic non-decreasing function)를 따른다.

이러한 모형을 해석학적으로 표시하면 다음과 같이 표현 된다.

$$\frac{dm(t)}{dt} = \tilde{a}K(t) \frac{dc(t)}{dt} \dots\dots\dots (2.1)$$

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

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

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

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

$$m(t) = \tilde{a}Kc(t) \dots\dots\dots (2.3)$$

(2.3)식은 t 시점에서 탐색 될 오류의 기대수는 오류의 총 오류의 기대수와 t 시점에서의 오류 탐색 커버리지의 곱 과 같다.

따라서 $a = \tilde{a}K$ 로 하면 탐색 될 오류의 기대수 $m(t)$ 는 다음과 같이 또 다른 형태로 표현 가능하다.

$$m(t) = a c(t) \dots\dots\dots (2.4)$$

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

$$\frac{dm(t)}{dt} = \lambda(t) = ac'(t) \dots\dots\dots (2.5)$$

고장 강도 함수 $\lambda(t) = [a-m(t)] h(t)$ 임이 됨이 알려져 있으므로[2, 4] (2.5)식을 변형하면 다음과 같이도 표현도 가능하다.

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

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

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

따라서 이러한 ENHPP모형은 테스트가 진행됨에 따라 각각의 오류이 발생할 수 있는 비율은 시간에 따라 변화 하기 때문에 시간에 의존하는 고장 발생 비율이라고 할 수 있다. 또, 이러한 위험함수에 해당하는 분포는 t 시점에서 커버리지 함수를 평가 할 수 있다.

이러한 ENHPP모형은 (불완전한) 커버리지 함수에 대 하여 실제적인 상황($c(\infty) < 1$) 을 적용 시킬 수 있다.

결과적으로 조건부 신뢰도(Conditional reliability) $R(t | s)$ 는 다음과 같이 알려져 있다[4].

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

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

2.3 커버리지 함수를 이용한 NHPP 모형(ENHPP)

고장 시간 영역 모형에서는 고장 강도 함수 $\lambda(t)$ 에 따라 다른 NHPP 모형이 된다. 따라서 커버리지 함수 $c(t)$ 도 다른 형태로 존재 한다. 결국 $c(t)$ 도 여러 가지 형태를 제시 할 수 있다. 본절에서는 잘 알려진 지수 커버리지 함수와 S-형 커버리지 함수를 재조명 한다.

2.3.1 지수 커버리지 함수

잘 알려진 Goel-Okumoto 모형(11)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.

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

$$\lambda(t|a, g) = a g e^{-gt} \dots\dots\dots (2.10)$$

단, a 는 무한하게 테스트 시간이 주어졌을 때 관찰된 고장의 기대수이고 g 는 오류당 고장 발생 비율을 의미한다.

따라서 지수모형에 대한 커버리지 함수 $c(t)$ 수와 위험함수 $h(t)$ 는 다음과 같이 유도 된다.

$$c(t) = 1 - e^{-gt} \dots\dots\dots (2.11)$$

$$h(t) = \frac{c'(t)}{1-c(t)} = g \dots\dots\dots 2.12)$$

2.3.2 S-형 커버리지 함수

S-형 커버리지 함수 모형(12)은 Yamada-Ohba-Osaki 모형으로 알려져 있고 S-형 모형을 NHPP로 접근하면 평균값 함수와 강도 함수는 다음과 같이 알려져 있다.

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

$$\lambda(t|a, g) = a g^2 t e^{-gt} \dots\dots\dots (2.14)$$

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

$$c(t) = 1 - (1+gt)e^{-gt} \dots\dots\dots (2.15)$$

$$h(t) = \frac{c'(t)}{1-c(t)} = \frac{g^2 t}{1+gt} \dots\dots\dots (2.16)$$

III. 제안된 Burr 커버리지 함수

이 절에서 Burr 분포(7, 14)를 이용한 커버리지 함수에 대하여 요약 하고자 한다.

이 분포는 1942년 Burr에 의해 만들어진 분포이고 이 분포의 특징은 왜도(Skewness)와 첨도(Kurtosis) 측면에서 보다 넓은 범위에 분포하기 때문에 어떠한 수명 단봉 자료(Unimodal data)에서도 적용할 수 있는 장점을 가진 분포라고 알려져 있다(14). 이 분포의 확률밀도 함수와 분포함수는 다음과 같이 표현 됨이 알려져 있다.

$$f(t|\tau, b) = \tau b t^{\tau-1} / (1+t^\tau)^{b+1} \dots\dots\dots (3.1)$$

$$F(t|\tau, b) = 1 - (1+t^\tau)^{-b} \dots\dots\dots (3.2)$$

단, $t > 0, \tau > 0, b \geq 1$.

이러한 Burr 분포에 대한 연구는 많은 학자들에 의하여 연구되었는데 그 중에서 Austin(13)은 이 분포를 관리도에 적용하였으며 Dubey(14)는 고장 시간 모형으로서의 Burr 분포의 장점을 고찰 하였고 Tadikamalla(15)는 다른 분포와의 관계를 연구하였고 Upadhyay 와 Javed, Peshwani(8)은 베이지안적 접근 방법을 동원하여 4모수에 대한 모수 추정방법을 연구하기도 하였다.

이러한 Burr 모형을 NHPP로 접근하면 평균값 함수와 강도 함수는 다음과 같이 표현 할 수 있다.

$$m(t|a, \tau, b) = a(1 - (1+t^\tau)^{-b}) \dots\dots\dots (3.3)$$

$$\lambda(t|a, \tau, b) = a \tau b t^{\tau-1} / (1+t^\tau)^{b+1} \dots\dots\dots (3.4)$$

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

$$c(t|\tau, b) = 1 - (1+t^\tau)^{-b} \dots\dots\dots (3.5)$$

$$\begin{aligned}
 h(t|\tau, b) &= \frac{c'(t|\tau, b)}{1-c(t|\tau, b)} \dots\dots\dots (3.6) \\
 &= \frac{\tau b t^{\tau-1} / (1+t^\tau)^{b+1}}{(1+t^\tau)^{-b}} \\
 &= \frac{\tau b t^{\tau-1}}{1+t^\tau}
 \end{aligned}$$

$$\frac{\partial \ln L_{Exp}}{\partial a} = \frac{n}{a} - 1 + e^{-g x_n} = 0,$$

$$\frac{\partial \ln L_{Exp}}{\partial g} = \frac{n}{g} - \sum_{k=1}^n x_k - a x_n e^{-g x_n} = 0$$

각 모수에 대한 최우추정량 \hat{a}_{MLE} 와 \hat{g}_{MLE} 은 다음식을 만족한다.

IV. ENHPP 모형에 대한 모수 추정

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

$$\frac{n}{\hat{a}} = 1 - \exp(-\hat{g}) x_n \dots\dots\dots (3.11)$$

$$x_k = \sum_{i=1}^k t_i \quad (k=1,2,\dots,n; 0 \leq x_1 \leq x_2 \leq \dots \leq x_n) \dots\dots (3.7)$$

$$\frac{n}{\hat{g}} = \sum_{k=1}^n x_k + \hat{a} x_n \exp(-\hat{g} x_n) \dots\dots\dots (3.12)$$

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 구성된다. 이 시간 절단 모형에서의 우도함수는 다음과 같이 알려져 있다[4, 16].

(3.11)과 (3.12)식을 수치적으로 풀어 두 모수를 구할 수 있다.

$$L(\theta | D_t) = \left(\prod_{i=1}^n \lambda(x_i) \right) \exp(-m(t)) \dots\dots\dots (3.8)$$

S-커버리지 모형에 대한 로그 우도 함수 (3.8)식에 (2.10)을 (2.11)에 대입하면 다음과 같다[3, 12].

$$\begin{aligned}
 L_S(a, g | D_{x_n}) &= \left(\prod_{k=1}^n a g^2 x_k e^{-g x_k} \right) \dots\dots\dots (3.13) \\
 &\cdot \exp[-a [1 - (1 + g x_n) e^{-g x_n}]]
 \end{aligned}$$

단, θ 은 미지의 모수 집합을 의미하고 우도함수 (3.8) 식에서 t 을 x_n 으로 대체하면 유사한 형태의 고장 절단 모형의 우도함수가 된다. 본 논문에서 이러한 우도함수한 최우추정법(Maximum Likelihood Estimation: MLE)으로 모수 추정을 실시하고자 한다.

따라서 로그우도함수를 구하면 다음과 같이 표현된다.

지수 커버리지 모형에 대한 로그 우도 함수 는 (3.8)식에 (2.10)을 (3.9)에 대입하면 다음과 같이 나타낼 수 있다..

$$\begin{aligned}
 \ln L_S &= n \ln a + 2n \ln g + \sum_{k=1}^n \ln x_k - g \sum_{k=1}^n x_k \dots\dots (3.14) \\
 &- a [1 - (1 + g x_n) e^{-g x_n}]
 \end{aligned}$$

따라서 고장절단모형에서의 모수 a 와 g에 관한 편미분 식은 다음과 같이 유도할 수 있다.

$$L_{Exp}(a, g | D_{x_n}) = \left(\prod_{k=1}^n a g e^{-g x_k} \right) \exp[-a(1 - e^{-g x_n})] \dots\dots (3.9)$$

$$\frac{\partial \ln L_S}{\partial a} = \frac{n}{a} - 1 + e^{-g x_n} + g x_n e^{-g x_n} = 0,$$

MLE을 이용하기 위한 로그우도함수를 구하면 다음과 같이 유도된다.

$$\frac{\partial \ln L_2}{\partial \beta_S} = \frac{2n}{g} - \sum_{k=1}^n x_k - a g x_n^2 e^{-g x_n} = 0$$

$$\ln L_{Exp} = n \ln a + n \ln g - g \sum_{k=1}^n x_k - a(1 - e^{-g x_n}) \dots\dots (3.10)$$

따라서 각 모수에 대한 최우추정량 \hat{a}_{MLE} 와 \hat{g}_{MLE} 은 다음 식을 만족한다.

따라서 고장절단모형에서의 모수 a 와 g에 관한 편미분 식은 다음과 같이 유도할 수 있다.

$$\frac{n}{\hat{a}} = 1 - \exp(-\hat{g}) - \hat{g} x_n \exp(-\hat{g} x_n) \dots\dots\dots (3.15)$$

$$\frac{2n}{\hat{g}} = \sum_{k=1}^n x_k + \hat{a} \hat{g} x_n^2 \exp(-\hat{g} x_n) \dots\dots\dots (3.16)$$

(3.15)과 (3.16)식을 수치적으로 풀어 두 모수의 MLE 추정량을 구할 수 있다.

유사한 방법으로 (3.4)식과 (3.5)식을 연관하면 우도함수 $L_{Burr}(a, \tau, b | D_{x_n})$ 는 다음과 같이 표현 할 수 있다.

$$L_{Burr}(a, \tau, b | D_{x_n}) = \prod_{k=1}^n a (\tau b x_i^{\tau-1} / (1+x_i^\tau)^{b+1}) \dots\dots\dots (3.17)$$

$$\cdot \exp(-a(1+(1+x_n^\tau)^{-b}))$$

Burr 커버리지 모형에 대한 로그우도함수는 다음과 같이 표현된다[6, 7].

$$\ln L_{Burr}(a, \tau, b | D_{x_n}) = n \ln a + n \ln \tau + n \ln b \dots\dots\dots (3.18)$$

$$+ (\tau-1) \sum_{k=1}^n \ln x_i - (b+1) \sum_{k=1}^n \ln(1+x_i^\tau)$$

$$- a + a(1+x_n^\tau)^{-b}$$

형상 모수 b 값은 상수(사전에 알고 있는 경우)라고 가정했을 때 최우추정법을 이용하기 위하여 (3. 18)식을 a 와 τ 에 대하여 편미분을 하면 다음과 같은 식을 유도 할 수 있다.

$$\frac{\partial \ln L_{Burr}(a, \tau, b)}{\partial a} = \frac{n}{a} - 1 + (1+x_n^\tau)^{-b} = 0 \dots\dots\dots (3.19)$$

$$\frac{\partial \ln L_{Burr}(a, \tau, b)}{\partial \tau} = \frac{n}{\tau} + \sum_{k=1}^n \ln x_i + \dots\dots\dots (3.20)$$

$$(b+1) \sum_{k=1}^n \frac{x_i^\tau \ln x_i}{1+x_i^\tau} - ab(1+x_n^\tau)^{-b-1} x_n^\tau \ln x_n = 0$$

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

$$\frac{n}{\hat{a}} = 1 - (1+x_n^{\hat{\tau}})^{-b} \dots\dots\dots (3.21)$$

$$\frac{n}{\hat{\tau}} = - \sum_{k=1}^n \ln x_i - (b+1) \sum_{k=1}^n \frac{x_i^{\hat{\tau}} \ln x_i}{1+x_i^{\hat{\tau}}} \dots\dots\dots (3.22)$$

$$+ ab(1+x_n^{\hat{\tau}})^{-b-1} x_n^{\hat{\tau}} \ln x_n$$

(3.21) 식과 (3.22) 식을 비선형 연립방정식(수치해석적 방법)을 이용하여 풀면 최우추정치 \hat{a}_{MLE} 와 $\hat{\tau}_{MLE}$ 의 값을 구 할 수 있다.

본 논문에서는 Burr 분포의 형상 모수 b 값은 1, 2, 3 과 같은 상수인 경우를 본 논문에서는 제시하고자 한다. 예를 들어 $b = 1$ 인 경우에는 다음과 같은 연립방정식을 유도 할 수 있으며 다른 경우도 유사한 방법을 이용 할 수 있다.

$$\frac{n}{\hat{a}} = 1 - (1+x_n^{\hat{\tau}})^{-1} \dots\dots\dots (3.23)$$

$$\frac{n}{\hat{\tau}} = - \sum_{k=1}^n \ln x_i - 2 \sum_{k=1}^n \frac{x_i^{\hat{\tau}} \ln x_i}{1+x_i^{\hat{\tau}}} \dots\dots\dots (3.24)$$

$$+ a(1+x_n^{\hat{\tau}})^{-2} x_n^{\hat{\tau}} \ln x_n$$

V. 수치적인 예

이 장에서 실제 고장자료를 이용하여 형상모수에 따른 Burr 모형을 분석하고자 한다. 이 고장자료는 NTDS (Naval Tactical Data System)에 의해 발생된 소프트웨어 고장자료로서 Goel과Okumoto[11], Mazzuchi 와 Soyer[17], Kuo와 Yang[16] 등이 이 고장자료를 이용하여 소프트웨어 모형을 제안 한 바 있다. 본 연구도 이 자료를 이용하고자 한다. 이 자료는 <표 1>에 나열 되어 있고 제시하는 신뢰모형들을 분석하기 위하여 우선 자료에 대한 추세 검정이 선행 되어야 한다[2, 16].

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 고장 시간이 증가함에 따라 산술 평균이 거의 증가 추세를 보이고 있으므로 신뢰성장(Reliability growth) 속성을 가지고 있고 <그림 2>에 나타난 라플라스 추세 검정의 결과도 라플라스 요인(Factor)이 증가하다가 고장시간이 증가함에 따라 음수로서 감소하기 하는 신뢰성장(Reliability growth) 속성을 나타내고 있다. 신뢰성장 모형을 이 자료에 적용시키는 것이 가능하다.[3, 18, 19]. 본 논문에서는 모수추정을 용이하게 하기 위하여 실제자료에서 수치변환된 자료(*failure times* $\times 10^{-2}$)를 이용하였다.

표 1. 고장 간격 자료
Table 1. Failure Interval Time Data

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	386	3.96
30	9	405	4.05
31	135	540	5.4

소프트웨어 신뢰성 모형의 모수 추정에는 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 $b=3$ 이상인 경우는 비효율적이면서 근이 수렴되지 않아 본 논문에서는 고려대상에서 제외 시켰다.

기존의 모형과 Burr 커버리지 모형에 대한 모수의 추정값들의 결과는 <표 2>에 요약되었다.

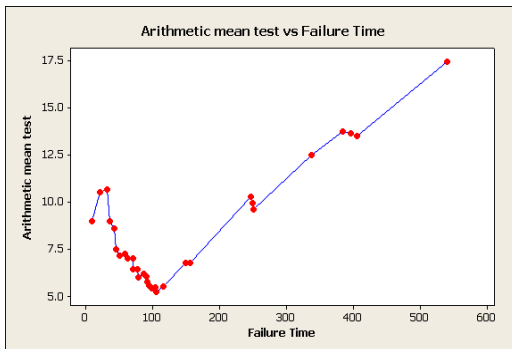


그림 1. 산술평균검정
Figure 1. Arithmetic mean test

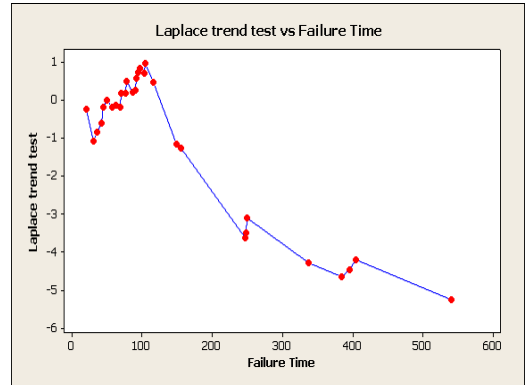


그림 2. 라플라스 추세 검정
Figure 2. Laplace trend test

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

Model	MLE	
<i>exponential coverage</i>	$\hat{g}_{MLE} = 0.39812$	$\hat{a}_{MLE} = 35.073$
<i>S-coverage</i>	$\hat{g}_{MLE} = 0.68351$	$\hat{a}_{MLE} = 34.862$
<i>Burr(b=1) coverage</i>	$\hat{\tau}_{MLE} = 1.71259$	$\hat{a}_{MLE} = 32.726$
<i>Burr(b=2) coverage</i>	$\hat{\tau}_{MLE} = 1.90508$	$\hat{a}_{MLE} = 32.247$
<i>Burr(b=2.5) coverage</i>	$\hat{\tau}_{MLE} = 1.91675$	$\hat{a}_{MLE} = 31.016$

모형 선택의 하나의 방법으로 편차자승합(SSE [1,9])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 Burr 커버리지모형이 기존에 알려진 모형인 S-커버리지모형이나 지수 커버리지모형에 비해 상대적으로 효율적 모형으로 나타나고 있다.

예측오류(Prediction error)의 비정상성(Nonstationarity)에 대한 측도는 Kolmogorov 거리(distance)[2, 4]로 측정되는데 이 거리가 클수록 상대적으로 비정상성을 내포하고 있다. <그림 3>은 S-Plus 소프트웨어[20]를 이용하여 Kolmogorov 검정에 대한 그림을 보여주고 있고 이 그림에서도 전체적으로 Burr모형이 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 Burr 커버리지모형의 형태가 상대적으로 비정상성 속성이 덜 내포하고 있음을 알 수 있다.

표 3. 모형들에 대한 SSE의 값
Table 3. SSE of each model

Model	SSE
<i>exponential coverage</i>	9013.172
<i>S-coverage</i>	5049.463
<i>Burr(b=1) coverage</i>	1022.814
<i>Burr(b=2) coverage</i>	3890.297
<i>Burr(b=2.5) coverage</i>	6876.1376

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

Model	Kolmogorov distance
<i>exponential coverage</i>	0.464987
<i>S-coverage</i>	0.258784
<i>Burr(b=1) coverage</i>	0.196432
<i>Burr(b=2) coverage</i>	0.258784

〈그림 4〉은 각 모형에 대하여 위험함수를 그린 그림이다. 이 그림에서는 지수 커버리지 모형은 일정하고 S-커버리지 모형과 Burr 모형은 증가형태를 나타내고 있으며 〈그림 5〉에서는 커버리지 함수를 그린 그림이다. ENHPP 모형에서 가장한 대로 연속적인 단조 비감소 함수형태로 나타나고 있다.

t 는 임무시간(Mission time)에 대한 즉, 소프트웨어의 고장이 일어나지 않을 확률인 신뢰도(Reliability) $\hat{R}(t | x_n)$ 는 (2.8)식을 이용하면 〈그림 6〉과 같다. 이 그림에서도 Burr 분포 커버리지 모형($b=1, 2$)들이 상대적으로 효율적인 신뢰도를 나타내고 있다.

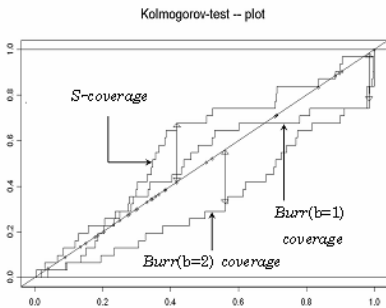


그림 3. 모형에 대한 Kolmogorov 검정
Figure 3. Kolmogorov test of each model

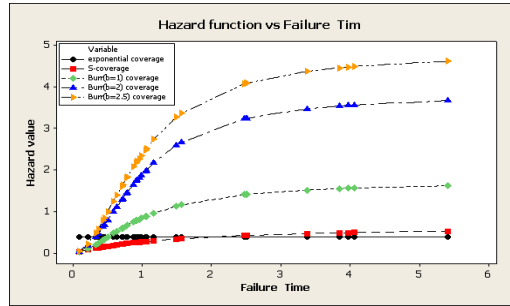


그림 4. 각 모형에 대한 위험함수
Figure 4. Hazard function of each model

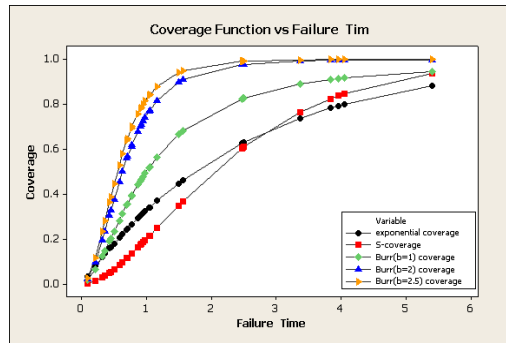


그림 5. 각 모형에 대한 커버리지함수
Figure 5. Coverage function of each model

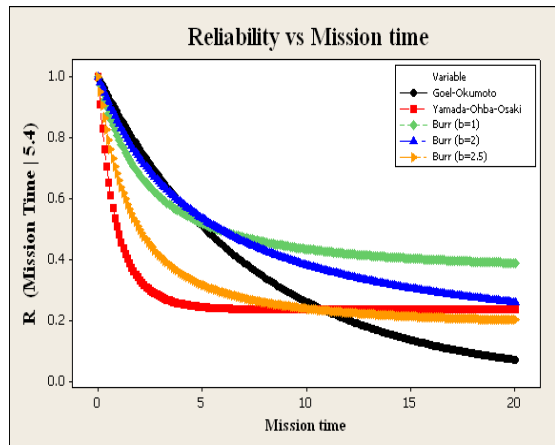


그림 6. 각 모형에 대한 신뢰도
Figure 6. Reliability of each model

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 Burr분포를 이용한 Burr 커버리지모형을 제안하였다. 고장 간격시간으로 구성된 자료를 이용하여 기존의 모형과 Burr 모형에 대하여 최우 추정법을 이용하여 모수 추정을 실시하였다. 소프트웨어 고장분석 자료는 NTDS 자료를 통하여 분석하였다. 효율적인 모형 비교를 위한 편차자승합의 결과는 Burr 커버리지모형이 기존에 잘 알려진 지수 커버리지 모형이나 S-커버리지 모형보다 우수함을 보이고 있고 콜모고로프 거리의 결과도 Burr 커버리지모형이 상대적으로 정상성에 가까운 모형으로 간주할 수 있다. 임무시간 비교 분석한 결과도 Burr 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 Burr 분포를 이용한 ENHPP 모형도 이 분야에서 효율적으로 이용 할 수 있는 모형이 됨을 알 수 있었다. 향후 이러한 Burr 분포를 이용한 베이지안적 접근 방법이나 수리적인 추정과 검정 부분에 대한 수리적인 접근이 기대 된다.

참고문헌

- [1] Lawless, J. F. "Statistical Models and Methods for Lifetime Data". John Wiley & Sons, New York, 1981.
- [2] 김 희철, "일반화감마분포를 이용한 NHPP 소프트웨어 신뢰도 모형에 관한 연구", 한국컴퓨터정보학회 논문지, 10권 6호, pp. 27-35, 2005.
- [3] 김 희철, "지수화 지수분포에 의존한 NHPP 소프트웨어 신뢰성장 모형에 관한 연구", 한국컴퓨터정보 학회 논문지, 11권 5호, pp.9-18, 2006
- [4] S. S. Gokhale and K. S. Trivedi. "A time/structure based software reliability model". Annals of Software Engineering, 8, pp. 85-121, 1999
- [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 Intl. Symposium on software Reliability Engineering, White Plains, NY, 1996.
- [6] 김 희철, 박종구, "Burr분포를 이용한 NHPP 소프트웨어 신뢰성장모형에 관한 연구", 한국해양정보통신 학회 논문지, 11권3호, pp.514-522, 2007.
- [7] I. W. Burr, "Cumulative frequency functions". Ann. Math. Statistic. Vol 13, pp 215-232, 1942.
- [8] S. K. Upadhyay and I. A. Javed and M. Peshwani, "Bayesian analysis of generalized four-parameter Burr distribution via Gibbs sampler, METRON-International Journal of statistics, Vol, LXII, n.1, pp.115-135, 2004.
- [9] J.R. Horgan, S. London, and M.R. Lyu, "Achieving Software Quality with Testing Coverage Measure," IEEE Computer, pp.60-69, Sept. 1994.
- [10] R. Jacoby and K. Masuzawa, "Test Coverage Dependant Software Reliability Estimation by the HGD Model," 3rd Intl. Symposium on Software Reliability Engineering, 1992.
- [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] Yamada, S., Ohba, M. and Osaki, S. "S-Shaped Reliability Growth Modeling for Software

- Error Detection". IEEE Trans. on Reliability, R-32(5):475-485, Dec. 1983.
- [13] J. A. Austin, "Control chart constants for largest and smallest in sampling from a normal distribution using the generalized Burrestimation". Technometrics, Vol. 15, pp. 931-933, 1971.
- [14] S. D. Dubey, "Statistical treatment of certain life testing and reliability problems", ARL TR pp. 73-0155, AD 774537, 1973
- [15] P. K. Tadikamalla, "Alook at the Burr and related distributions", Inter. Statist. Rev., 48, pp. 337-344, 1980
- [16] Kuo, L. and Yang, T. Y. "Bayesian Computation of Software Reliability". Journal of the American Statistical Association, Vol.91, pp.763-773, 1996.
- [17] T. A. Mazzuchi and R. Soyer. "A Bayes Empirical Bayes Model for Software Reliability". IEEE Transactions on Reliability, 37, pp 248-254, 1988
- [18] K. Kanoun and J. C. Laprie. Handbook of Software Reliability Engineering, M.R.Lyu, Editor, chapter Trend Analysis, pp.401-437.
- [19] S. Brocklehurst and B. Littlewood, Handbook of Software Reliability Engineering, M.R.Lyu, Editor, chapter Techniques for Prediction Analysis and Recalibration, pp.119-166.
- [20] S. Selvin. Moden Applied Biostatistical Methods Using S-Plus, pages 141-184. Oxford University Press, New York, 1998.
- [21] 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.

저 자 소개



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