

잠재변수를 이용한 NHPP 베이지안 소프트웨어 신뢰성 모형에 관한 연구

김희철* · 신현철**

요 약

본 논문은 소프트웨어 신뢰성장 모형에 대한 베이지안 모수추론과 모형선택 방법이 연구되었다. 소프트웨어 성장 모형은 내재되어 있는 오류와 고장 간격시간으로 모형화하면 소프트웨어 개발 단계에서 유용하게 사용할 수 있다. 본 논문에서는 사후 분포의 정보를 얻기 위한 다중 적분문제에 있어서 일종의 마코브 체인 몬테칼로 방법인 깁스 샘플링을 사용하여 사후 분포의 계산이 이루어졌다. 확산 사전 분포를 가진 소프트웨어 신뢰성에 의존된 일반적 순서 통계량 모형에 대하여 베이지안 모수 추정이 이루어 졌고 효율적인 모형의 선택방법도 시행되었다. 모형 설정과 선택 판단기준은 편차 자승합을 이용한 적합도 검정과 추세 검정이 사용되었다.

본 논문에서 사용된 소프트웨어 고장 자료는 Minitab(version 14) 통계 패키지에 있는 와이블 분포(형상모수가 2이고 척도모수가 5)에서 발생시킨 30개의 난수를 이용한 모의 실험자료를 이용하여 고장자료 분석을 시행하였다.

The NHPP Bayesian Software Reliability Model Using Latent Variables

Hee-Cheul Kim* · Hyun-Cheul Shin**

ABSTRACT

Bayesian inference and model selection method for software reliability growth models are studied. Software reliability growth models are used in testing stages of software development to model the error content and time intervals between software failures. In this paper, could avoid multiple integration using Gibbs sampling, which is a kind of Markov Chain Monte Carlo method to compute the posterior distribution.

Bayesian inference for general order statistics models in software reliability with diffuse prior information and model selection method are studied. For model determination and selection, explored goodness of fit (the error sum of squares), trend tests. The methodology developed in this paper is exemplified with a software reliability random data set introduced by of Weibull distribution(shape 2 & scale 5) of Minitab (version 14) statistical package.

Key words : Gibbs Algorithm, NHPP (Nonhomogeneous Poisson Process), Rayleigh Distribution, Diffuse Prior Information, Software Reliability, SSE (Error Sum of Squares)

* 남서울대학 산업경영공학과

** 백석문화대학 컴퓨터정보학부

1. 서 론

소프트웨어 신뢰성은 일정한 환경조건에서 일정기간동안 고장이 나지 않고 운영할 수 있는 확률이다. 따라서 이러한 소프트웨어 신뢰성은 시스템 신뢰도에 영향을 주는 중요한 요소가 되고 디자인 속성 측면에서는 하드웨어 신뢰성과는 다른 면을 가지고 있으며 소프트웨어의 다양한 기능은 소프트웨어 신뢰성 문제들에 관한 주요한 요인이 된다.

소프트웨어의 신뢰도의 일반적인 정의는 일정한 기간 동안 주어진 환경 하에서 컴퓨터 프로그램을 고장 없이 사용할 수 있는 확률을 의미한다. 어떤 프로그램이 코딩(coding)되었을 때 바로 테스트와 디버깅 작업이 수행된다. 이러한 단계에서 고장이 날 때까지 테스트된다. 테스트 하다가 고장이 발견되면 즉시 고쳐진다.

소프트웨어 테스트 단계에서 소프트웨어 고장수(Number of failure)와 고장간격시간에 의해 소프트웨어 고장현상을 수리적으로 모형화하면 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 통계자료 해석시 자료를 대수변환 이 후에 정규분포로 처리하면 되는 대수정규분포(Lognormal distribution), 이 분포와 유사한 대수로지스틱 분포(Loglogistic distribution)도 이 분야에 사용이 가능하다[2]. 또, 자유도(Degree of freedom; df)에 의존하는 카이제곱(χ^2 distribution) 분포도 적용이 가능하다. 이러한 카이제곱 분포는 변화가 일어날 때 까지의 대기시간을 나타내는데

많이 사용되는 감마분포(Gamma Distribution)의 특수한 경우로도 접근할 수 있기 때문에 소프트웨어 신뢰성 수명분포로 설명할 수 있다[3].

이러한 분야에서 Kuo와 Yang[4] 등은 깁스 알고리즘(Gibbs algorithm)과 잠재변수(latent variable)를 이용한 베이지안 분석방법을 연구하 되기도 하였다[5].

신뢰도 모수의 값을 계산하는데 있어서 일반적인 방법은 최우추정법(Maximum Likelihood estimation; MLE)을 사용한다. 그러나 새로운 자료가 얻어지면 그 자료를 이제까지 얻어던 자료와 결합하여 새로운 결론에 도달 하려는 이론이 베이즈 추정법이라고 알려져 있다. 즉, 베이즈이론은 알려져 있는 사실에 대한 주관적 의견을 경험이나 지식을 바탕으로 사전정보를 만든 다음 실험을 통하여 얻어진 자료와 결합하여 사후 정보를 추출하는 과정이라 할 수 있다. 베이즈 추정법에서 사전확률분포인 수명분포가 다소 복잡해지면 사후정보의 추출이 어려워진다. 이러한 경우 깁스 알고리즘(Gibbs algorithm)을 이용하면 근사값을 추정할 수 있다.

따라서 본 논문에서는 깁스 알고리즘을 이용하여 지수형, 랄리(Rayleigh)형과 멱(power)형태의 특성을 가진 일반 순서통계량에 기초한 소프트웨어 신뢰성 모형에 대하여 사전정보를 포함하는 베이지안적 모수 추정 방법을 이용한 효율적 모형의 선택과 신뢰도의 특성을 알아보려고 한다.

이러한 모형의 적용은 소프트웨어 고장자료에 따른 강도함수(고장발생률) $\lambda(t)$ 형태에 의존된다.

본 논문의 내용은 다음과 같은 내용으로 서술되었다. 2절은 비동질적인 포아송과정에 대하여 알아보았고 3절에서는 일반순서 통계량에 기초한 소프트웨어 신뢰성모형에 대한 베이지안 추론이 제시되었다.

이러한 신뢰성 모형에 대한 모수 추정은 깁스 알고리즘을 이용하였다. 모형 설정 및 선택은 4절에서 서술하였고 5절에서는 수치적인 예로서 모형 비교에 따른 적용사례를 서술하였다.

2. NHPP

소프트웨어 고장이 비동질적인 포아송과정(NHPP; nonhomogeneous Poisson process)에 의존하는 소프트웨어 신뢰성 모형은 시간 영역 모형으로 분류된다. 이러한 확률과정의 모수인 $\lambda(t)$ 는 소프트웨어 수행시점 t 에 관련된 강도함수를 나타낸다.

$N(t)$ 을 수행시점 t 까지의 누적 고장수를 나타내고 $m(t)$ 을 이것의 평균값 함수($m(t) = E[N(t)]$)으로 나타내면 $m(t)$ 와 강도함수 $\lambda(t)$ 은 다음과 같은 관계를 만족한다[6].

$$m(t) = \int_0^t \lambda(s) ds, \quad \frac{dm(t)}{d(t)} = \lambda(t) \quad (1)$$

따라서 $N(t)$ 는 모수 $m(t)$ 을 가진 포아송 확률 밀도함수(probability density function)로 알려져 있다. 즉,

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

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[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. 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 NHPP 모형의 평균값 함수는 다음과 같이 표현 할 수 있다[7].

$$m(t) = \theta F(t) \quad (3)$$

식 (3)으로 부터 고장 강도함수 $\lambda(t) = \theta F'(t)$ 으로 표현된다. 여기에서는 소프트웨어 테스트 시점에서 미지의 N 개의 결함들을 가지고 있고 이 N 개의 결함들로부터 실수영역(R^+)에서 임의의 확률 밀도 함수 f (probability density function)에 따라 발생된 n 개의 순서 통계량이 고장 시점이 된다. 충분한 테스트 시간이 경과함에 따라 평균값 함수 $m(t)$ 가 유한값을 가지는 과정 ($\lim_{t \rightarrow \infty} m(t) = \theta < \infty$)을 NHPP(fin)라고 표시하면 식 (3)으로부터,

$$F(t) = 1 - e^{-\beta t}$$

을 가지면 Goel-Okumoto [8]과정(지수 순서 통계량)이 되고, $F(t) = 1 - e^{-\beta t}$ 이면 Goel[9]과정(와이블 순서 통계량)이 되며,

$$F(t) = 1 - (1 + \beta t) e^{-\beta t}$$

이면 Yamada, Ohba-Osaki[6] 과정이 된다. 이러한 모형은 각 수리 시점에서 새로운 결함이 발생하지 않는다는 가정을 한다. 따라서 디버깅 과정에서 새로운 결함(faults)이 발생하면 일반 순서 통계량 모형을 기록값 통계량 모형으로 대체된다. 이러한 상황에서 $t \rightarrow \infty$ 에 따라 $m(t) \rightarrow \infty$ 로 되는 경우를 NHPP(inf) [4]로 표시하면 $m(t) = -\ln(1 - F(t))$ 가 된다. 이러한 관계로부터 $\lambda(t) = \alpha/(t + \beta)$ 가 되면 Musa-Okumoto 과정이 되고 $\lambda(t) = \alpha \beta t^{\alpha-1}$ 가 되면 the Duane 과정, $\lambda(t) = \exp(\alpha + \beta t)$ 이면 Cox and Lewis 과정이 된다[9].

3. 순서 통계량 및 잠재변수에 기초한 소프트웨어 신뢰성 모형에 대한 베이지안 모수 추론

순서통계량에 기초한 모형은 미지의 모수 θ 와

β 을 포함한 평균값 함수 $m(t)$ 을 가진 NHPP의 형태이다. t 시간까지 테스트 되는 시간 절단(time truncated)모형은 n 번 까지 관찰된 고장 시점을 x_1, x_2, \dots, x_n 로 표시하면 데이터 집합 D_t 는 $\{n; x_1, x_2, \dots, x_n, t\}$ 로 표시할 수 있다. 이러한 경우의 우도함수는 다음과 같이 됨이 알려져 있다[11].

$$L_{NHPP}(\theta, \beta | D_t) = \left(\prod_{i=1}^n \lambda(x_i) \right) e^{-m(t)} \quad (4)$$

n 번 고장까지 관찰된 고장 절단모형을 고려하면 집합 D_{x_n} 는 $\{x_1, x_2, \dots, x_n\}$ 로 표시할 수 있다.

감마족 순서통계량 모형에 대한 베이지안 추론에 대하여 깁스 알고리즘(Gibbs algorithm)을 고려할 수 있다[11].

우도함수인 식 (4)에 포함된 $m(t) = \theta F(t)$ 의 표현 때문에 깁스 알고리즘에 사용되는 데이터 집합 D_t 가 주어졌을 때 θ 와 β 에 대한 조건부밀도함수를 구체화하기가 어렵기 때문에 모수 $\theta[1 - F(t|\beta)]$ 를 가진 포아송분포를 가지는 잠재변수(latent variable) $N' = N - n$ 를 도입하면 해결될 수 있다. 이러한 방법을 이용하여 사후분포 $p(\theta, \beta | D_t)$ 는 결합밀도 $p(\theta, \beta, N' | D_t)$ 을 주변밀도화(marginalization)함으로써 얻을 수 있다. 즉, 결합사후 밀도 $p(\theta, \beta, N' | D_t)$ 은 다음과 같은 조건부 밀도들을 고려한 깁스 샘플링을 이용하면 얻을 수 있다[4, 15].

$$p(N' | \theta, \beta, D_t), p(\theta | N', \beta, D_t), p(\beta | N', \theta, D_t).$$

3.1 지수 모형(Goel-Okumoto)에 대한 깁스 알고리즘

식 (3)에 있는 $F(t)$ 가 지수 분포일 경우에는 다음과 같은 평균값 함수로 나타낼 수 있다.

$$m(t) = \theta F(t) = \theta(1 - e^{-\beta t}) \quad (5)$$

단, θ 는 소프트웨어에 잠재해 있는 고장의 기대수를 의미하고 β 는 결합 탐색비율을 의미한다. 따라서 강도함수는 다음과 같이 표현된다.

$$m'(t) = \lambda(t) = \theta \beta e^{-\beta t}.$$

식 (4)를 이용하여 지수 순서 통계량 모형을 고려하면 θ 와 β 에 대한 우도함수는 다음과 같이 표현된다.

$$L_{NHPP}(\theta, \beta | D_t) = \left(\prod_{i=1}^n \theta \beta e^{-\beta x_i} \right) e^{-\theta(1 - e^{-n})} \quad (6)$$

베이지안 추정을 하기 위하여 잠재변수 $N' = N - n$ 을 도입하여 N', θ, β 에 대한 사전분포(밀도)는 다음과 같이 가정하였다.

- (1) $N' \sim P(\theta(1 - F(t|\beta))) = P(\theta e^{-\beta t})$;
- (2) $\theta \sim \Gamma(a, b)$; $a(>0), b(>0)$: known
- (3) $\beta \sim \Gamma(c, d)$; $c(>0), d(>0)$: known
- (4) $\theta \perp \beta$.

$P(k)$ 는 모수 k 를 가진 포아송 분포를 의미하고 $\Gamma(a, b)$ 는 평균이 a/b 이고 분산이 a/b^2 인 감마분포를 의미하며 $\theta \perp \beta$ 는 θ 와 β 가 독립임을 나타내고 결합 사후 밀도는 다음과 같이 유도된다[4, 13].

$$\begin{aligned} f(N', \theta, \beta | D_t) &\propto \frac{(\theta e^{-\beta t})^{N'}}{N'!} e^{-\theta e^{-n}} \theta^{n+a-1} \\ &\cdot \beta^{c+n-1} e^{-b\theta} e^{-(d + \sum_{i=1}^n x_i)\beta} \cdot e^{-\theta(1 - e^{-n})} \quad (7) \\ &\propto \frac{1}{N'!} \theta^{N'+n+a-1} \beta^{c+n+1} e^{-(b+1)\theta} e^{-(d+tN' + \sum_{i=1}^n x_i)\beta} \end{aligned}$$

고장 절단 모형에서는 식 (6)과 식 (7)에 있는 표현에서 t 을 x_n 으로 대체하면 된다. 따라서 깁스 알고리즘에 사용되는 사후 밀도함수들은 다음과 같이 표현된다.

- (1) $N' | \theta, \beta, D_t \sim P(\theta e^{-\beta t})$;
- (2) $\theta | N', \beta, D_t \sim \Gamma(a+n+N', b+1)$;
- (3) $\beta | N', \theta, D_t \sim \Gamma(c+n, d+tN' + \sum_{i=1}^n x_i)$;

3.2 랄리 모형에 대한 깃스 알고리즘

식 (3)에 있는 $F(t)$ 가 랄리(Rayleigh) 분포일 경우에는 다음과 같은 평균값 함수로 나타낼 수 있다.

$$m(t) = \theta F(t) = \theta(1 - e^{-\beta t^2})$$

따라서 강도함수 $\lambda(t) = 2\theta\beta t e^{-\beta t^2}$ 으로 유도된다.

식 (4)를 이용하여 랄리 순서 통계량 모형을 고려하면 θ 와 β 에 대한 우도함수는 다음과 같이 표현된다.

$$L_{NHPP}(\theta, \beta | D_t) = \left(\prod_{i=1}^n 2\theta x_i \beta e^{-\beta x_i^2} \right) e^{-\theta(1 - e^{-t^2})}$$

베이지안 추정을 하기 위하여 잠재변수 $N' = N - n$ 을 도입하여 N', θ, β 에 대한 사전분포(밀도)는 다음과 같이 가정하였다.

- (1) $N' \sim P(\theta(1 - F(t | \beta))) = P(\theta e^{-\beta t^2})$
- (2) $\theta \sim \Gamma(a, b)$; $a(>0), b(>0)$: known
- (3) $\beta \sim \Gamma(c, d)$; $c(>0), d(>0)$: known
- (4) $\theta \perp \beta$.

따라서 사후 결합밀도함수는 다음과 같이 유도된다.

$$f(N', \theta, \beta | D_t) \quad (8)$$

$$\propto \frac{1}{N'!} \theta^{N'+n+a-1} \beta^{c+n-1} \cdot e^{-(b+1)\theta} e^{-(d+t^2N'+\sum_{i=1}^n x_i^2)\beta}$$

고장 절단 모형에서는 식 (8)에 있는 표현에서 t 을 x_n 으로 대체하면 된다. 따라서 깃스 알고리즘

에 사용되는 사후 밀도함수들은 다음과 같이 표현된다.

- (1) $N' | \theta, \beta, D_t \sim P(\theta e^{-\beta t^2})$;
- (2) $\theta | N', \beta, D_t \sim \Gamma(a+n+N', b+1)$;
- (3) $\beta | N', \theta, D_t \sim \Gamma(c+n, d+t^2N' + \sum_{i=1}^n x_i^2)$;

3.3 멱 모형에 대한 깃스 알고리즘

식 (3)에 있는 $F(t)$ 가 멱(power) 분포일 경우 일 경우에는 다음과 같은 평균값 함수로 나타낼 수 있다.

$$m(t) = \theta F(t) = \theta(\lambda t)^2$$

따라서 강도함수 $\lambda(t) = 2\theta\lambda^2 t$ 으로 유도된다.

$m(t)$ 와 $\lambda(t)$ 표현에서 λ^2 을 β 으로 대체하면 유사한 표현으로 수정할 수 있다.

식 (4)를 이용하여 멱 순서 통계량 모형을 고려하면 θ 와 β 에 대한 우도함수는 다음과 같이 표현된다.

$$L_{NHPP}(\theta, \beta | D_t) = \left(\prod_{i=1}^n 2\theta x_i \beta \right) \exp(-\theta\beta t^2)$$

베이지안 추정을 하기 위하여 θ, β 에 대한 사전 분포(밀도)는 다음과 같이 가정하였다.

- (1) $\theta \sim \Gamma(a, b)$; $a(>0), b(>0)$: known
- (2) $\beta \sim \Gamma(c, d)$; $c(>0), d(>0)$: known
- (3) $\theta \perp \beta$.

따라서 사후밀도함수는 다음과 같이 유도할 수 있다.

$$f(\theta | D_t) \propto \theta^{n+a+1} \beta^{n+c-1} e^{-2\theta\beta t^2} e^{-b\theta} e^{-\beta d} \quad (9)$$

고장 절단 모형에서는 식 (9)에 있는 표현에서 t 를 x_n 으로 대체하면 된다. 따라서 깃스 알고리즘에 사용되는 사후 밀도함수들은 다음과 같이 표현된다.

- (1) $\theta | \beta, D_t \sim \Gamma(a+n, b+\beta t^2)$;
- (2) $\beta | \theta, D_t \sim \Gamma(c+n, d+2\theta t^2)$;

4. 모형 설정 및 선택

이 절에서 소프트웨어 고장 자료의 분석에 사용될 기법에 대하여 설명하고자 한다. 소프트웨어 신뢰성의 연구는 일반적으로 신뢰성장 모형에 적용할 척도를 개발하는 일이다. 사용될 신뢰성장 모형을 결정하기 위하여 추세검정은 매우 유용하다. 추세 검정은 산술평균 검정(arithmetic mean test)과 라플라스 검정(Laplace test)[2, 14]이 있다.

4.1 산술평균 검정

산술평균 검정은 관측된 고장 간격시간 ($t_j, j = 1, 2, \dots$)에 대한 산술평균을 계산하면 된다.

$$\text{즉, } \tau(i) = \frac{\sum_{j=1}^i t_j}{i}. \quad (10)$$

$\tau(i)$ 의 값이 증가추세를 가지면 신뢰성장이 되지만 감소추세를 가지면 신뢰성 감소가 된다.

4.2 라플라스 검정

라플라스 검정(Laplace test)은 최적고장간격시간과의 차이에 대한 검정을 말한다. 이 검정은 NHPP에 대한 가정[2]이 만족했을 때 $N(t)$ 를 이용한다. $N(t)$ 가 (0, t)시간에서 탐색된 결함들의 수라

고 할 때 감소(decreasing)패턴을 가지면 신뢰성장 속성을 가지고 있고 감소패턴을 가지면 신뢰성 감소 속성을 가진다. 그리고 상수이면 안정된 신뢰 속성(stable reliability)을 가지며 라플라스 검정은 다음과 같은 라플라스 요인[14]을 계산함으로써 시행할 수 있다.

$$\mu(i) = \frac{\frac{1}{i-1} \sum_{n=1}^{i-1} \sum_{j=1}^n t_j - \frac{\sum_{j=1}^i t_j}{2}}{\sum_{j=1}^i t_j \sqrt{\frac{1}{12(i-1)}}} \quad (11)$$

계산된 라플라스 요인에 따라서 몇 가지 기준으로 해석할 수 있는데 음수 값을 가지면 강도(intensity)가 감소형태를 가지고 양수 값을 가지면 증가추세를 나타내는 신뢰 속성을 가지며 -2와 2사의 값을 가지는 라플라스 요인의 값은 안정된 신뢰성을 가진다.

4.3 적합도 검정

관측된 고장자료는 모형에 대한 모수를 추정하는데 사용된다. 따라서 평균값 함수가 계산되고 그림으로 그려보면 전체적인 윤곽을 알 수 있다. 여기에서 편차 제곱합(error sum of squares)은 적합도 검정을 평가하기 위한 도구[6]로 사용되어 질 수 있다. 모형 적합도 검정에서 편차 제곱합이 작으면 보다 적합한 모형이라는 결론을 내릴 수 있다. 편차 제곱합(C_{SSE})은 다음과 같이 정의된다.

$$C_{SSE} = \sum_{i=1}^n (n_i(x_i) - \hat{m}(x_i))^2 \quad (12)$$

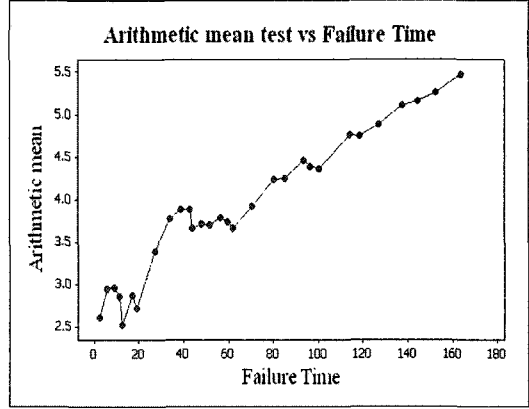
단, $n_i(x_i)$ 은 시간 x_i 시점까지 탐색된 결함의 수이고 $\hat{m}(x_i)$ 은 시간 x_i 시점까지 탐색된 결함의 기대값을 의미한다.

5. 수치적인 예

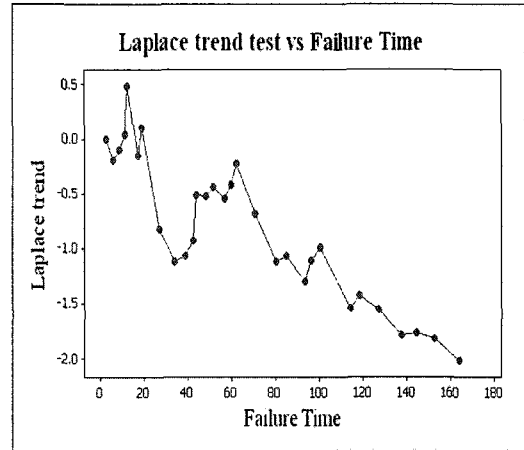
데이터집합은 Minitab(version 14) 통계 패키지에 있는 와이블분포(형상모수가 2이고 척도모수가 5)에서 발생시킨 30개의 난수를 이용한 모의 실험자료를 이용하여 고장자료 분석을 시행하였다. 데이터의 요약은 <표 1>에 요약하였다. 이 자료를 이용하여 산술평균 검정과 라플라스 검정을 실시한 결과가 (그림 1)에 요약되었다. 이 그림에서 산술평균 검정은 시간의 흐름에 따라 증가패턴을 가지고 있으므로 신뢰성장 추세를 보이고 있는 자료로 간주되고 라플라스 검정의 결과는 시간의 흐름에 따라 양에서 음수 형태로 변화가는 추세를 보이고 있으므로 고장 강도(intensity)가 증가하다가 감소하는 신뢰속성을 가지고 있다. 이러한 검정에서 보여 주듯이 <표 1>에 있는 자료는 신뢰성장을 나타내기 때문에 신뢰성장 모형을 이 자료에 적용시키는 것이 가능하다.

<표 1> 모의 고장 자료

Failure number	Failure Interval Time	Failure Time
1	2.62	2.62
2	3.28	5.90
3	2.99	8.89
4	2.55	11.44
5	1.22	12.66
6	4.63	17.29
7	1.83	19.12
8	7.96	27.08
9	6.89	33.97
10	4.89	38.86
11	3.90	42.76
12	1.26	44.02
13	4.25	48.28
14	3.48	51.76
15	5.06	56.82
16	3.05	59.87
17	2.35	62.22
18	8.51	70.73
19	9.71	80.44
20	4.62	85.06
21	8.51	93.57
22	3.07	96.63
23	3.81	100.44
24	13.80	114.25
25	4.37	118.62
26	8.45	127.07
27	10.72	137.78
28	6.60	144.39
29	8.12	152.51
30	11.53	164.05



(그림 1-A) 산술평균 검정



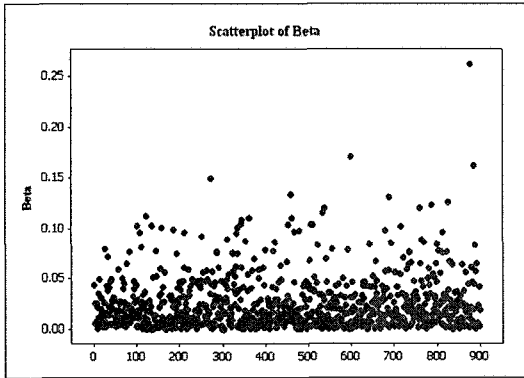
(그림 1-B) 라플라스 검정

따라서 이 고장시간 자료에 대한 정보는 $n=30$ 이고 $x_n = x_{30} = 164.05$ 가 된다.

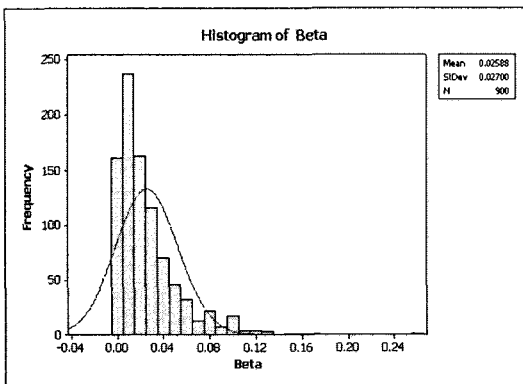
따라서 고장절단 모형에서는 x_n 을 $x_{30} = 164.05$ 으로 가정되고 β 에 대한 사전분포는 $\Gamma(1, 10^{-3})$, θ 에 대한 사전정보는 $\Gamma(60, 2)$ 을 주었다.

N' , β 와 θ 에 대한 주변 사후 밀도들(marginal posterior densities)을 계산하기 위하여 10개의 분리된 깁스 체인들(Gibbs chain)에 대하여 1000번의 반복시행을 하였다. 반복정지 시점을 결정하기 위하여 분산분석 기술(analysis of variance technique)을 사용하는 Gelman & Rubin[16]의 방법을 이용

하여 깃스 샘플링의 수렴성을 모니터링(monitoring) 하였다. 즉, 각 모수에 대하여 10개의 깃스 체인에 대하여 1000번을 반복하여 110번째, 120번째, ..., 1000번째 값을 고려하면 표본의 크기는 900개가 된다. 지수모형에 대한 모수 β 에 대한 사후분포 밀도는 (그림 2)에 나타나 있고 이 사후 표본에 의한 히스토그램은 (그림 3)에 나타나 있다. 이러한 그림들로부터 사후 평균 추정값은 0.025884임을 알 수 있다.



(그림 2) 지수모형에 대한 모수 β 에 대한 사후분포 밀도



(그림 3) 지수모형에 대한 모수 β 에 대한 사후분포 히스토그램

유사한 방법으로 다음과 같은 사후 평균 추정값을 정보를 얻을 수 있다.

<표 2> 각 모형에 대한 사후 밀도에 대한 정보

지수 모형 (Exponential model)	평균(Mean)	중위수 (Median)	표준편차 (S.D)
N'	4.198735	2.00000	1.234724
θ	34.13292	18.87567	3.871356
β	0.025884	0.017771	0.027002

랄리 모형 (Rayleigh model)	평균(Mean)	중위수 (Median)	표준편차 (S.D)
N'	2.099316	1.233451	3.232456
θ	31.352787	16.023412	2.345672
β	5.181×10^{-5}	2.438×10^{-5}	1.348×10^{-6}

멱 모형 (Power mode)	평균(Mean)	중위수 (Median)	표준편차 (S.D)
θ	89.6784	86.4529	2.3691
β	6.091×10^{-3}	3.165×10^{-3}	1.469×10^{-3}

<표 3> 편차제곱합 $C_{SSE} = \sum_{i=1}^{30} [n_i(x_i) - \hat{m}(x_i)]^2$

지수 모형	2058.569	랄리 모형	2039.765	멱모형	2189.275
-------	----------	-------	----------	-----	----------

<표 2>는 S-plus 소프트웨어[17]를 이용하여 계산된 사후 밀도의 정보를 나타낸 표이고 <표 3>은 각모형에 대한 편차제곱합을 계산한 표이다. 이 표에서 Rayleigh 모형이 편차제곱합이 제일 작기 때문에 지수 모형이나 멱 모형에 비해 우수한 모형이라고 간주할 수 있다.

6. 결 론

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, 발생한 고장 수와 고장 발생시간과의 관계를 효율적으로 관리함으로써 소프트웨어 신뢰도를 성장시킬 수 있다. 이러한 과정을 소프트웨어 성장 과정이라고 볼 수 있다.

본 논문에서는 각 모수에 대한 사전정보가 있는 비동질적 포아송 과정(NHPP)에 기초한 소프트웨어 에러현상에 대한 베이지안적 모수 추정을 위하여 잠재변수(latent variable)와 깁스 샘플링을 사용하였다. 지수분포, 레일리분포, 멱 분포 모형을 적용하고 효율적 모형의 선택을 위하여 편차제곱합을 이용하였다. 수치적인 예에서는 와이블분포(형상모수가 2이고 척도모수가 5)에서 발생시킨 30개의 난수를 이용한 모의 실험자료를 이용하였다. 그 결과 지수나 멱 분포를 이용한 모형보다는 레일리 분포 모형이 효율적인 모형으로 판단된다. 향후 실제적인 자료를 이용하거나 특수한 분포 예를 들면 보다 넓은 범위에 분포되는 Burr분포, 극값분포 등을 이용하면 효율적인 모형선택이 될 것으로 사료된다. 이러한 이론을 바탕으로 베이지안 분석과 모수 추정에 대한 수리적인 분포이론과 응용에 관한 연구가 기대된다.

참 고 문 헌

- [1] 김희철, “일반화감마분포를 이용한 NHPP 소프트웨어 신뢰도 모형에 관한 연구”, 한국컴퓨터정보학회논문지, 제10권 제6호, pp. 27-35, 2005.
- [2] Gokhale, S. S. and Trivedi. K. S. “A time/structure based software reliability model”, Annals of Software Engineering, 8, pp. 85-121. 1999.
- [3] 김희철, “카이제곱 NHPP에 의한 소프트웨어 신뢰성 모형에 관한 연구”, 한국컴퓨터정보학회논문지, 제11권 제1호, pp. 45-53, 2006.
- [4] Kuo, L. and Yang, T. Y. “Bayesian Computation of Software Reliability”, Journal of the American Statistical Association, Vol. 91, pp. 763-773, 1996.
- [5] 김희철, 이상식, 송영재, “NHPP 소프트웨어 신뢰도 모형에 대한 모수 추정 비교”, 한국정보처리학회논문지D, 제11-D권 제6호, pp. 1269-1276. 2004.
- [6] Yamada, S., Ohba, M. and Osaki. S. “S-Shaped Reliability Growth Modeling for Software Error Detection”, IEEE Trans. on Reliability. R-32 (5) : pp. 475-485, Dec. 1983.
- [7] Pham, H., Nordmann, L. and Zhang, X. “A General Imperfect-Software-Debugging Model with S-Shaped Fault-Detection Rate”, IEEE Trans. on reliability, Vol. 48, No. 2, 1999.
- [8] Goel, A. L. and Okumoto, K. “Time Dependent Error Detection Rate Model for Software Reliability and Other Performance Measures”, IEEE Transactions on Reliability, Vol. 28, pp. 206-211, 1979.
- [9] Lawless, J. F. “Statistical Models and Methods for Lifetime Data”, John Wiley and Sons, New York, 1981.
- [10] Duane, J. F. “Learning curve approach to reliability monitoring”, IEEE Transaction on Aerospace, AS-2(2), pp. 563-566, 1964.
- [11] Musa, J. D., Iannino, A. and Okumoto, K. “Software Reliability : Measurement, Prediction, Application”, New York : McGraw Hill, 1987.
- [12] Casella, G. and George, E. I. “Explaining the Gibbs Sampler”, The American Statistician, Vol. 46, pp. 167-174, 1992.
- [13] 강기훈 외 5인 공저, “베이지안 통계학”, 자유아카데미, pp. 129-133, 2005.

- [14] Kanoun, K. and Laprie, J. C. "Handbook of Software Reliability Engineering, M.R. Lyu, Editor, chapter Trend Analysis", pp. 401-437. MacGraw-Hill, New York, 1996.
- [15] Casella, G. and George, E. I. "Explaining the Gibbs Sampler", The American Statistician, Vol. 46, pp. 167-174, 1992.
- [16] Gelman, A. E. and Rubin D. "Inference from Iterative Simulation Using Multiple Sequences", Statistical Science, Vol. 7, pp. 457-472, 1992.
- [17] Selvin, S. "Modern Applied Biostatistical Methods Using S-Plus", Oxford University Press, New York, 1998.

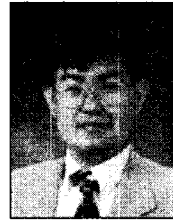


김희철

1992년 동국대학교 통계학과
(이학석사)

1998년 동국대학교 통계학과
(이학박사)

2005년 ~ 현재 남서울대학교 산업
경영공학과 전임강사



신현철

2002년 원광대학교 컴퓨터공학
(공학박사)

1994년 ~ 현재 백석문화대학
컴퓨터정보학부 부교수

2004년 (주) 아이비루션 자문위원