

일반 순서 통계량을 이용한 소프트웨어 신뢰확률 중첩모형에 관한 베이지안 접근에 관한 연구

이 병 수[†] · 김 회 철^{††} · 백 수 기^{†††} · 정 관 회^{†††} · 윤 주 용^{††††}

요 약

소프트웨어 시스템이 복잡해지면 고장의 원인이 하나의 강도함수에 의해서만 일어나지 않고 여러 원인이 중첩되어 발생할 수 있다. 이러한 복잡한 시스템에 의한 우도함수의 계산상의 어려움 때문에 반복표본을 이용하는 깁스 샘플링 기법이 고려되었다. 관찰된 고장시점은 중첩모형으로 표현이 가능한 잠재(latent)변수들을 이용하여 깁스 알고리즘을 적용 하였다. 단순모형과 중첩모형의 비교를 위해 사후베이지요인과 상대오차의 합을 이용하여 모형선택을 시도하였다.

수치적인 예에서 GOS 속성을 가진 Goel-Okumoto 모형과 Weibull 모형을 선택하고 NHPP의 자료는 Lewis와 Shedler[25]에 의해 제시된 Thinning 알고리즘을 이용하여 발생한 자료를 이용하고 사전분포는 상대적으로 확산분포(diffuse priors)를 이용한 모수추정과 사후베이지요인과 상대오차를 이용한 모형선택을 한 결과 단순모형들 보다 중첩모형이 좋은 모형으로 간주할 수 있음을 보여 주었다.

A Study on Bayesian Approach of Software Stochastic Reliability Superposition Model using General Order Statistics

Byoung-Soo Lee[†] · Hee-Cheul Kim^{††} · Su-Ki Paik^{†††} ·
Kwan-Hee Chung^{†††} · Ju-Yong Yoon^{††††}

ABSTRACT

The complicate software failure system is defined to the superposition of the points of failure from several component point process. Because the likelihood function is difficulty in computing, we consider Gibbs sampler using iteration sampling based method. For each observed failure epoch, we applied to latent variables that indicates with component of the superposition mode. For model selection, we explored the posterior Bayesian criterion and the sum of relative errors for the comparison simple pattern with superposition model.

A numerical example with NHPP simulated data set applies the thinning method proposed by Lewis and Shedler[25] is given, we consider Goel-Okumoto model and Weibull model with GOS, inference of parameter is studied. Using the posterior Bayesian criterion and the sum of relative errors, as we would expect, the superposition model is best on model under diffuse priors .

† 종신회원 : 시립인천대학교 전자계산학과 교수
†† 정 회 원 : 시립인천대학교 전자계산학과 강사
††† 정 회 원 : 경기대학교 전자계산학과 교수
†††† 정 회 원 : 한국도로공사 정보처 정보운영부장
논문접수 : 1999년 1월 6일, 심사완료 : 1999년 7월 7일

1. 서 론

소프트웨어 고장들로 인한 컴퓨터 시스템의 고장은 우리 사회에 엄청난 손실을 초래할 수도 있다. 따라서 소프트웨어 신뢰도 엔지니어링에서의 연구 활동은 지난 20년 넘게 행해졌고 많은 소프트웨어 신뢰도 모델들이 소프트웨어에 남아 있는 고장들의 수와 소프트웨어 신뢰도의 추정을 위해서 제안되었다. 제품이 고장이 나면 수리를 해서 사용할 수 있는 수리계 시스템(repairable system)에서의 신뢰성에 대한 연구 역시 중요한 문제이다.

시스템이 고장이 나면 고장이 난 원인을 찾아 필요할 경우 새로운 디자인을 개발하거나 새로운 기술을 도입하게 된다. 따라서 시간이 지남에 따라 신뢰도의 증가가 기대되어 진다. 이런 모형을 소프트웨어 신뢰 모형(software reliability model)이라고 한다. 소프트웨어 신뢰도의 예측문제에 대한 많은 경험기법들은 그들의 정도에 있어서 많은 차이를 보이고 있다. 확실한 것은 모든 환경 하에서 신뢰성 있는 결과를 만족시켜주는 기법은 아주 어렵고 전통적으로 어느 하나의 데이터에서 다른 데이터로 이동함에 따라 즉, 어느 하나의 예측형태에서 다른 형태로 변환함에 따라 다양하게 변할 수 있으나 비효율적인 고장이나 새로운 결함을 완화할 수 있다면 소프트웨어의 신뢰도가 높아질 수 있다. 소프트웨어 테스트 단계에서 소프트웨어 오류수와 고장간격시간에 의해 소프트웨어 고장현상을 수리적으로 모형화를 하면 소프트웨어에 대한 평가를 쉽게 할 수 있으며, 소프트웨어 오류수, 소프트웨어 고장발생간격시간, 소프트웨어 신뢰도 및 고장률 등의 신뢰성 평가측도들이 추정되어 예측할 수 있다.

소프트웨어 신뢰도에서 발견된 고장의 수를 모형화하는 것은 비동질적 포아송 과정(nonhomogenous Poisson process, NHPP)으로 널리 사용하여 왔다[17, 19]. $M(t)$ 를 시간 $(0, t]$ 사이에 발견된 고장의 수라 정의하면 $M(t)$ 는 평균값함수(mean value function) $m(t) = EM(t)$ 인 NHPP에 의해 모형화 될 수 있고 $m(t) = \theta F(t)$ 와 테스트 시점에서 결함의 총수를 N 이라고 가정함으로써 NHPP으로부터 일반순서통계량(general order statistic, GOS) 모형을 유도할 수 있다.

지금까지 연구들은 강도함수가 하나인 Jelinski-Moranda 모형이나 Goel-Okumoto 모형 같은 단순모형

을 개선하는 수정모형이 연구되어 왔다. 그러나 소프트웨어 시스템이 복잡해지면 고장의 원인이 하나의 강도함수에 의해서만 일어나지 않고 여러원인이 중첩되어 발생 할 수 있다. 이러한 복잡한 시스템에 의한 우도함수의 계산상의 어려움 때문에 반복표본을 이용하는 깃스 알고리즘을 적용하였다. NHPP 모형의 중첩에 대하여 잠재(latent)변수들을 이용하여 깃스알고리즘을 적용하였고 단순모형과 중첩모형의 비교를 위해 사후 베이즈요인을 이용하여 모형선택을 할 수 있다.

본 논문의 2절에서는 제안된 중첩모형을 유도하는데 관련된 NHPP와 GOS 그리고 Jelinski-Moranda 모형과 Goel-Okumoto 모형의 관계를 요약하였고 3절에서는 중첩모형에 대한 일반적인 깃스샘플링방법을 설명하였고 GOS속성을 가진 Goel-Okumoto와 Weibull의 중첩과정을 제시하였으며 4절에서는 모수의 추론과 모형선택에 대한 방법을 소개하고 5절에서는 수치적인 예가 나열되었고 끝으로 6절에서는 결론 및 연구과제가 서술되었다.

2. 관련연구

2.1 NHPP

소프트웨어 신뢰도에서 발견된 고장의 수를 모형화하는 것은 비동질적 포아송 과정(nonhomogenous Poisson process, NHPP)으로 널리 사용하여 왔다[17,18]. $M(t)$ 를 시간 $(0, t]$ 사이에 발견된 고장의 수라 정의하면 $M(t)$ 는 평균값함수(mean value function) $m(t) = EM(t)$ 인 NHPP에 의해 다음과 같이 모형화 될 수 있다.

$$P(M(t) = n) = \frac{m(t)^n}{n!} e^{-m(t)} \quad (2.1)$$

단, $n = 0, 1, 2, \dots$.

$m(t)$ 가 t 에 대한 비감소함수(nondecreasing function) 추세를 가진 미분가능함수이면 강도함수(intensity function)는 $\lambda(t) = m'(t)$ 가 됨이 알려져 있다. $\lambda(t)$ 가 상수($m(t)$ 가 선형(linear) 추세)이면 동질적 포아송 과정(homogeneous Poisson process)이고, t 에 대한 함수형태이면 NHPP가 된다, 예를 들어 동질적 포아송 과정, Musa-Okumoto(1984) 과정, Weibull(1983) 과정 그리고 Cox-Lewis[6] 과정은 각각 NHPP의 강도함수가 상수, 시간에 관한 부분(fraction)함수, 시간에 관한

력(power)함수, 시간에 관한 대수선형(log-linear)함수이다.

2.2 일반 순서 통계량(GOS)

$m(t)$ 는 미지의 모수 θ, β 에 의하여 결정된다고 가정하고 ($m(t) = m(t|\theta, \beta)$), θ 는 소프트웨어에 존재하는 총오류의 수이고 β 는 결함탐색률(the fault detection rate)을 의미한다.

따라서 시간 t 까지 테스트하기 위한 시간절단(time truncated) 모형이 주어지면, n 번의 고장시점의 순서화된 자료를 x_1, x_2, \dots, x_n 이라고 하면 데이터 집합 $D_t = \{n, x_1, x_2, \dots, x_n; t\}$ 으로 표현되고 n 번째까지 고장시점이 관찰된 고장절단(failure truncated) 모형일 경우 데이터 집합 D_{x_n} 은 (x_1, x_2, \dots, x_n) 으로 구성된다. 주어진 시간절단모형에서 구간 $(0, t]$ 에서 관측된 고장의 수는 확률변수인 n 이 된다. 따라서 주어진 $M(t) = n$ 에서 구간 $(0, t]$ 인 $x_1 < \dots < x_n < t$ 의 분포는 동일한 분포

$$F_w(w) = m(w)/m(t), 0 \leq w \leq t$$

로부터 얻은 확률변수의 순서통계량(GOS)과 같다[20]. 따라서 주어진 자료에 대한 시간절단모형에 대한 GOS 모형의 우도 함수는 다음과 같다[17].

$$L_{NHPP}(\theta, \beta|D_t) = [\prod_{i=1}^n \lambda(x_i)] \exp(-m(t)) \tag{2.2}$$

단 $m(t) = \int_0^t \lambda(u) du$ 이고, $0 < x_1 < x_2 < \dots < x_n < t$, β 은 미지의 모수벡터를 의미한다.

NHPP에서 $\lim_{t \rightarrow \infty} m(t) = \theta < \infty$ 이고, N 은 테스트 시점에서의 결함의 수라고 가정하면 시간절단모형에 대한 데이터 D_t 가 주어졌을 때 θ, β, N 에 대한 우도함수는 다음과 같다[21].

$$L_{NHPP}(\theta, \beta, N|D_t) = \frac{N!}{(N-n)!} \left[\prod_{i=1}^n \frac{\lambda(x_i)}{\theta} \right] \left(1 - \frac{m(t)}{\theta} \right)^{N-n} \tag{2.3}$$

여기서 $p(N|\theta)$ 을 평균이 θ 인 포아송 확률질량함수하면 다음과 같이 표현할 수 있다.

$$\sum_{N=n}^{\infty} L_{NHPP}(\theta, \beta, N|D_t) \times p(N|\theta) = L_{NHPP}(\theta, \beta|D_t) \tag{2.4}$$

그러므로 (2.2)식의 우도함수는 N 이 포아송분포를 따른다는 가정을 이용하면 주변우도함수가 된다. 고장절단 모형에서는 t 을 x_n 으로 대치하면 (2.2)식과 유사한 형태가 된다. GOS 모형에서의 고장시점은 특정한 형태의 분포함수 $F(S|\beta)$ 을 가진 $f(S|\beta)$ 에 따라 분포되는 독립이고 동일한 분포의 관찰값 N 으로부터 추출된 n 개의 순서통계량이라고 가정된다. 그러므로 고장절단모형의 우도함수는 다음과 같다.

$$L_{GOS}(N, \beta|D_{x_n}) = \frac{N!}{(N-n)!} \left[\prod_{i=1}^n f(x_i|\beta) \right] [1 - F(x_n|\beta)]^{N-n} \tag{2.5}$$

시간절단모형에서는 우도함수 $L_{GOS}(N, \beta|D_t)$ 은 F 에 있는 x_n 을 t 로 대치하면 (2.5)식과 유사한 모형이 된다. $m(t) = \theta F(t)$ (즉, $\lambda(t) = \theta f(t)$)라고 하면 $L_{GOS}(N, \beta|D_t)$ 와 (2.3)식의 조건부 우도함수와 동일하다. 유사한 결과로 인해서 고장절단모형의 테스트에 적용할 수 있다.

결과적으로 $m(t) = \theta F(t)$ 와 테스트시점에서 결함의 총수를 N 이라고 가정함으로써 NHPP으로부터 GOS 모형을 유도할 수 있다[20]. 역으로 생각하면 고장시점이 GOS 모형이고 평균이 θ 인 포아송 분포를 따르는 N 을 가정하면 $M(t)$ 는 다음과 같이 평균값함 $m(t) = \theta F(t|\beta)$ 인 NHPP가 성립함을 보일 수 있다.

GOS 모형으로부터 N 이 주어진 $M(t) = n$ 의 조건부 확률이 다음과 같이 표현 할 수 있다.

$$p(M(t) = n|N) = \binom{N}{n} F(t|\beta)^n [1 - F(t|\beta)]^{N-n}$$

이 경우 N 이 포아송분포를 따른다는 가정을 이용하면 다음의 결과를 도출 할 수 있다.

$$p(M(t) = n) = \sum_{N=n}^{\infty} \binom{N}{n} F(t|\beta)^n [1 - F(t|\beta)]^{N-n} \frac{\theta^N e^{-\theta}}{N!} = e^{-\theta F(t|\beta)} \frac{[\theta F(t|\beta)]^n}{n!} \tag{2.6}$$

그러므로 (2.2)식과 비교하면 $M(t)$ 는 평균값함수 $m(t) = \theta F(t; \beta)$ 인 NHPP가 된다.

2.3 Jelinski-Moranda 모형과 Goel-Okumoto 모형

$\lim_{t \rightarrow \infty} m(t)$ 가 무한(infinite)할 때 (2.2)식과 관련하여 평균값함수 $m(t) = -\ln(1 - F(t))$ 이고 강도함수가 위험함수(hazard function) $\lambda(t) = m'(t) = f(t)/(1 - F(t))$ 가 성립함을 증명하였다(Resnick,1987; Kuo 와 Yang, 1996).

여기서 $F(t)$ 는 밀도함수 $f(t)$ 를 따르는 누적분포함수이고 지수누적분포함수 $F(t) = (1 - e^{-\beta t})$ 인 경우에는 Jelinski-Moranda 모형이 된다. 그러나 GOS 모형에서는 $m(t) = \theta F(t)$ 이므로 $m(t) = \theta (1 - e^{-\beta t})$ 를 가진 NHPP가 되는데 이것을 Goel-Okumoto 과정이라고 한다. Jelinski-Moranda 모형은 $f(t) = \beta \exp(-\beta t)$ 을 가진 GOS 모형의 특수한 경우가 된다. Jelinski-Moranda 모형과 Goel-Okumoto 과정의 관계는 Langberg와 Singpurwalla[16] 의해서 재조명되었다. Goel-Okumoto 과정과 관련하여 NHPP중에서 $F(t)$ 가 이증지수, 고펜퍼츠(Gompertz), 랄리(Rayleigh), 감마, 그리고 와이불(Weibull)일 때 이증지수 NHPP, 고펜퍼츠 NHPP, 랄리 NHPP, 감마 NHPP, 와이불 NHPP 등 여러모형을 제시 할 수 있다.

2.4 깃스 샘플링 알고리즘과 메트로 폴리스 알고리즘

깃스 샘플링은 다차원 이상의 적분을 계산하기 위한 것으로서 신경망(Neural network), 전문가 시스템(Expert system)과 같은 척도가 큰(large scale) 복잡한 모형에 적용되어 사용되어 왔다. 본 절에서는 Gelfand와 Rubin[10]이 제시한 다중열(multiple sequence)을 이용한 반복적 시뮬레이션 기법을 설명하고자 한다.

깃스 샘플링의 가정은 추출된 표본은 서로 독립이고 충분한 수의 반복이 이루어지면 깃스 알고리즘으로부터 추출된 표본은 안정된 특정분포로 수렴한다. 과산포분포(over-dispersed distribution)에서 유도된 초기점을 가지는 각각 길이 $2n$ 인 독립열 $m \geq 2$ 인 열을 시뮬레이트하고, 초기분포의 의존성을 작게하기 위해 각 열에 대해 전반부 n 번 반복을 제외하고 후반부 n 번 반복만 고려하게 된다.

확률변수 U_1, U_2, \dots, U_p 의 결합분포가 주어졌을 때 각각의 주변확률밀도가 이용가능하다고 가정하고

초기값을 $(u_1^{(0)}, u_2^{(0)}, \dots, u_p^{(0)})$ 라고 정의하면 다음과 같은 조건밀도로부터 변량이 추출된다.

$$\begin{aligned} u_1^{(1)} &\sim f(U_1 | U_2^{(0)}, U_3^{(0)}, \dots, U_p^{(0)}, D_n), \\ u_2^{(1)} &\sim f(U_2 | U_1^{(1)}, U_3^{(0)}, \dots, U_p^{(0)}, D_n), \\ &\vdots \\ u_p^{(1)} &\sim f(U_p | U_1^{(1)}, U_2^{(1)}, \dots, U_{p-1}^{(1)}, D_n) \end{aligned} \tag{2.7}$$

각각의 변수들은 최신값들로 갱신되면서 p 개의 변량을 발생시킬 수 있고 $2n$ 번의 반복이후에 다음과 같은 변량을 얻을 수 있다.

$$(U_1^{(2n)}, U_2^{(2n)}, \dots, U_p^{(2n)})$$

그러므로 후반부 n 번 반복(iteration)의 m 번 적용(replication)하는 깃스 샘플링은 다음과 같이 mn 개를 발생시킨다.

$$(U_{ij}^{(l)}, \dots, U_{pj}^{(l)}) (j=1, 2, \dots, m, l=n+1, \dots, 2n)$$

위 식에서 U_1, \dots, U_p 는 경우에 따라서 벡터가 될 수 있다. 사후밀도의 값을 계산하기 위해서 본 논문은 다음과 같이 Rao-Blackwell 개념을 사용하였다[9,10].

$$f(\bar{U}_s) \approx (mn)^{-1} \sum_{j=1}^m \sum_{l=n+1}^{2n} f(U_s | U_{rs}^{(l)}, r \neq s) \tag{2.8}$$

조건부 밀도가 쉽게 구체화되지 않을 때나, 공액사전분포(conjugate prior distribution)가 없는 상황에서는 메트로 폴리스(Metropolis) 알고리즘을 사용할 수 있다[11]. 이 기법은 다음과 같이 요약 할 수 있다.

$$p(u_1 | u_2, \dots, u_p, D_t) = \frac{p(D_t | u_1, u_2, \dots, u_p) \pi(u_1, u_2, \dots, u_p)}{\int_{-\infty}^{\infty} p(D_t | u_1, u_2, \dots, u_p) \pi(u_1, u_2, \dots, u_p) du_1} \tag{2.9}$$

식(2.9)와 같은 밀도로부터 표본을 추출한다고 가정하자. 표기법을 단순하게 하기 위해 위 조건부 밀도 (2.9)를 $f(u_1)$ 이라고 표현하자. u_1 을 X 에 연관(map)시키기 위해 추이 커널(kernel) 밀도 $q(u_1, X)$ 을 정의하자.

u_1 는 범위가 $(-\infty, \infty)$ 에서의 실수 변수이고 $X \leftarrow u_1 + \sigma Z$ 를 만족한다. 여기서 Z 는 표준화 정규 분포의

변량이고 σ 은 (2.9)에서 u_1 의 조건부 분산을 반영한다. u_1 이 (a, b) 사이의 유계(bound) 변수라면 커널 밀도 $q(u_1, X)$ 는 (a, b) 사이에 균등분포가 되고[11], 감마분포의 분산을 크게하면 커널 밀도로 사용할 수 있다[14]. 따라서 메트로폴리스 알고리즘은 다음과 같은 순서로 시행한다.

순서1> 임의의 점 $u_1^{(0)}$ 을 초기값으로 하고, 단계 지시 변수 $j=0$ 으로 놓자.

순서2> 추이 커널 $q(u_1^{(j)}, X)$ 에 따라서 하나의 점 X 을 발생시킨다.

순서3> $u_1^{(j)}$ 의 값에 대하여 $\min\left(1, \frac{f(X)}{f(u_1^{(j)})}\right)$ 을 계산한 값이 1이 되지 않으면 $u_1^{(j+1)} = X$ 로 놓는다.

반복 시행 하다가 $1 - \min\left(1, \frac{f(X)}{f(u_1^{(j)})}\right)$ 의 값이 0을 만족하면 $u_1^{(j)}$ 는 멈춘다.

순서4> 순서2, 3을 반복하여 정상 분포가 나올 때까지 계속한다.

결국은 (2.9)식에 있는 2개의 값들을 사용하고 있으므로 표준화 상수(normalizing constant)가 필요한 것이 아니라 우도와 사전분포의 합수적 형태가 필요한 것이다. 따라서 깃스 알고리즘 단계에서 메트로폴리스 알고리즘을 혼합 사용하면 보다 효율적인 사후분포를 계산 할 수 있다[11,14].

3. 중첩과정에 대한 깃스샘플링

3.1 중첩과정

소프트웨어 시스템이 복잡해지면 고장의 원인이 하나의 강도함수에 의해서만 일어나지 않고 여러원인이 중첩되어 발생 할 수 있다. 이런 경우에 강도함수와 평균값 함수를 알고 있으면 모형화가 가능하다.

$M_j(t)$ 을 시간 (0, t] 사이에서 강도함수가 $\lambda_j(t|\beta)$ 을 가진 j 번째 요소로부터 고장이 발생하는 NHPP라고 표현하고 $\lambda_j(t|\beta_j)$ 은 미지의 모수 β_j 을 가진 알고 있는 값이고 j 번째 요소에 의해 발생된 고장의 수 $M_j(t)$ ($j=1, 2, \dots, J$)는 독립이라고 가정하면 $M(t) = \sum_{j=1}^J M_j(t)$. 즉, 중첩과정에서 (0, t]의 구간에서 발생된 고장의 총수는 강도함수와 평균값함수는 각각 다

음과 같은 NHPP가 된다[5]. 단, $\beta = \{\beta_1, \beta_2, \dots, \beta_J\}$.

$$\begin{aligned} \lambda(x_i|\beta) &= \lambda_1(x_i|\beta_1) + \lambda_2(x_i|\beta_2) + \dots + \lambda_J(x_i|\beta_J) \\ m(t|\beta) &= m_1(t|\beta_1) + m_2(t|\beta_2) + \dots + m_J(t|\beta_J) \end{aligned} \quad (3.1)$$

따라서 시간절단모형의 GOS속성을 가진 중첩과정에 대한 우도함수는 다음과 같다.

$$L_{NHPP}(\theta, \beta|D_t) = \left[\prod_{i=1}^n \lambda(x_i|\beta) \right] \exp[-m(t|\beta)] \quad (3.2)$$

3.2 깃스샘플링

깃스추출법은 적분이 복잡하거나 난해한 경우에 반복표본을 이용하여 정보를 얻는 알고리즘이다. 이 기법은 마코브체인몬테카를로(MCMC)의 기법중 하나이며 마코브 연쇄의 전이분포는 여러 개의 조건부밀도로서 이루어진다. 이러한 조건부분포의 반복표본추출을 이용하여 깃스추출법을 시행한다. 결국 마코브연쇄의 정상분포가 원하는 사후분포가 된다. 마코브연쇄의 다중열(multiple sequences)을 얻기 위하여 독립 초기점(starting points)을 가진 연쇄를 반복한다. 이러한 기법은 Tanner와 Wong[24], Gelfand & Smith[9], Casella & George[4], 그리고 Kuo & Yang[14]등 많은 학자들에 의해서 제시되고 있다. 본 논문에서 θ 는 소프트웨어에 존재하는 총오류의 수이고 β 는 결함탐색률(the fault detection rate)을 의미하며 시간 t까지 테스트하기 위한 시간절단(time truncated) 모형이 주어지면, n번의 고장시점의 순서화된 자료를 x_1, x_2, \dots, x_n 이라고 하면 데이터 집합 $D_t = \{n, x_1, x_2, \dots, x_n; t\}$ 으로 표현 할 수 있다. 중첩과정에서 θ 와 β 의 사후결합분포는 (3.2)의 우도함수와 사전분포를 베이즈정리에 의해 다음과 같이 나타낼 수 있다.

$$\begin{aligned} f(\theta, \beta|D_t) &\propto \left\{ \prod_{i=1}^n [\lambda_1(x_i|\beta_1) + \lambda_2(x_i|\beta_2) + \dots \right. \\ &\quad \left. + \lambda_J(x_i|\beta_J)] \right\} \cdot \exp\{-[m_1(t|\beta_1) + \dots \\ &\quad + m_J(t|\beta_J)]\} \cdot \prod_{j=1}^J \pi_j(\beta_j) \cdot \pi(\theta) \end{aligned} \quad (3.3)$$

단, β 는 모수벡터이고, π_j 는 β_j 에 대한 사전분포, π 는 θ 에 대한 사전분포를 의미하고 β_j 와 θ 은 독립이라고 가정하면 식(3.3)에 있는 $\prod_{i=1}^n [\lambda_1(x_i|\beta_1) + \lambda_2$

$(x_i|\beta_2) + \dots + \lambda_j(x_i|\beta_j)$ 의 표현 때문에 데이터 집합 D_i 의 조건하에서 β 와 θ 의 조건부분포의 형태를 구체화되기 어렵다. 이러한 어려움을 해소하기 위하여 잠재변수(latent variables)를 사용한다. I 를 잠재변수들의 집합(collection)이라고 표시하고 이 I 을 사용하여 β 와 θ 의 사후분포를 계산한다. 각 $i(i=1, \dots, n)$ 에 대해 잠재변수 $I_i=(I_{i1}, \dots, I_{ij})$ 을 설정하자. j 번째 요소에 의해서 i 번째 고장의 원인이 발생했을 때를 $I_{ij}=1$ 라고 하자. 그 외에는 $I_{ij}=0$ 이라하면, $i=1, \dots, n$, $\sum_{j=1}^2 I_{ij}=1$ 의 조건을 만족한다.

β, θ 와 D_i 의 조건에서의 I_i 의 조건부밀도는 다항분포(multinomial distribution, MN)가 되는데 이 분포는 모수가 1이고, 셀확률 (p_{i1}, \dots, p_{ij}) 을 가진다. 단,

$$p_{ij} = \lambda_j(x_i) / [\lambda_1(x_i) + \lambda_2(x_i) + \dots + \lambda_j(x_i)] \quad (34)$$

$I=(I_1, \dots, I_n)^T$ 라고 표시하면 i 번째 고장발생의 요인은 $n \times J$ 행렬이 된다. 여기서 시뮬레이션된 I_1, \dots, I_n 은 독립이고 다음과 같다.

$$I_i \sim MN(1, (p_{i1}, \dots, p_{ij})), \quad i=1, \dots, n. \quad (35)$$

따라서 I 와 D_i 의 조건하에서 β, θ 의 사후결합밀도는 다음과 같이 표현할 수 있다.

$$\begin{aligned} f(\theta, \beta | I, D_i) &\propto L(\theta, \beta | D_i) \cdot \prod_{i=1}^n P(I_i | \theta, \beta, D_i) \cdot \prod_{i=1}^n \pi_j(\beta_j) \cdot \pi(\theta) \\ &\propto \left[\prod_{j=1}^2 \prod_{i: I_{ij}=1} \lambda_j(x_i | \beta_j, \theta) \right] \cdot \prod_{i=1}^n \exp(-m_i(\theta)) \cdot \prod_{j=1}^2 \pi_j(\beta_j) \cdot \pi(\theta). \end{aligned} \quad (36)$$

조건부밀도는 I 와 D_i 의 조건하에서 β_j 의 독립 사후밀도들로 구성된 결합밀도가 된다. 그러므로 깃스샘플링은 반복을 통해서 $f(I|\theta, \beta, D_i)$ 로부터 I 을 추출하고 $f(\theta|\beta, I, D_i)$ 로부터 θ 를 추출하고 $f(\beta|I, \theta, D_i)$ 로부터 β 를 추출하는 반복형식을 통해 독립적으로 최종표본을 얻을 수 있다.

3.3. Goel-Okumoto와 Weibull의 중첩과정

3.2절에서 유도된 강도함수가 각 각 다음과 같다

Goel-Okumoto와 Weibull의 중첩과정을 제시하고자 한다.

$$\begin{aligned} \lambda_1(t) &= \theta \beta_1 \exp(-\beta_1 t), \\ \lambda_2(t) &= \theta \beta_2 \alpha t^{\alpha-1} \exp(-\beta_2 t^\alpha) \end{aligned} \quad (37)$$

단, $\alpha_1 > 0, \beta_i > 0 (i=1, 2), \theta > 0$ 을 만족하며 식(3.2)에서 강도함수가 들인 중첩과정에 대한 우도함수는 다음과 같다.

$$\begin{aligned} L(\beta, \theta | D_i) &= \left(\prod_{i=1}^n [\theta \beta_1 \exp(-\beta_1 x_i) + \theta \beta_2 \alpha x_i^{\alpha-1} \exp(-\beta_2 x_i^\alpha)] \right) \cdot \exp[-\theta(1-e^{-\beta_1 t} - (1-e^{-\beta_2 t^\alpha})] \end{aligned} \quad (38)$$

단, $\Gamma(a, b)$ 는 평균이 a/b 인 감마 분포를 표시하고, 4개의 독립 요소에 대한 사전분포는 다음과 같다.

$$\theta \sim \Gamma(a_1, b_1); \beta_1 \sim \pi_1(\beta_1); \beta_2 \sim \pi_2(\beta_2); \alpha \sim \pi_3(\alpha).$$

여기서, π_1, π_2 와 π_3 는 $\beta_i > 0 (i=1, 2), \alpha > 0$ 을 만족하는 임의의 사전분포이고 β, θ 와 D_i 의 조건하에서의 I_i 의 조건부밀도는 다항분포가 되는데 이 분포는 모수가 1이고 셀확률 (p_{i1}, p_{i2}) 을 가진다. 단, p_{ij} 는 $\lambda_j(x_i) / [\lambda_1(x_i) + \lambda_2(x_i)]$, $i=1, \dots, n, j=1, 2$ 이 된다. 이 중첩과정의 경우는 β, θ 와 D_i 의 조건하에서의 I_n 는 $i=1, \dots, n$ 에 대한 p_{i1} 의 모수를 가진 베르누이 분포로부터 발생된다. 단,

$$p_{i1} = \frac{\theta \beta_1 e^{-\beta_1 x_i}}{\theta \beta_1 e^{-\beta_1 x_i} + \theta \beta_2 \alpha x_i^{\alpha-1} e^{-\beta_2 x_i^\alpha}} \quad (39)$$

이고 $I_{i2} = 1 - I_{i1}$ 이 된다.

따라서 I 와 D_i 의 조건하에서 β, θ 의 사후결합밀도는 (3.6)식에 의해서 다음과 같다.

$$\begin{aligned} f(\theta, \beta | I, D_i) &\propto \prod_{j=1}^2 \prod_{i: I_{ij}=1} [\lambda_j(x_i)] \cdot \exp[-\theta(1-e^{-\beta_1 t} - (1-e^{-\beta_2 t^\alpha})] \\ &\quad - \theta(1-e^{-\beta_2 t^\alpha})] \frac{b_1^{\alpha_1} \theta^{\alpha_1} e^{-b_1 \theta}}{\Gamma(\alpha_1)} \\ &\quad \cdot \pi_1(\beta_1) \cdot \pi_2(\beta_2) \cdot \pi_3(\alpha) \end{aligned} \quad (310)$$

(3.10)식을 베이즈정리와 장애모수(nuisance parameter)의 개념을 사용하여 깃스추출법 알고리즘을 이용하기 위한 모든 조건부밀도들은 다음과 같은 4개의 조건식이 필요하다.

$$(1) \theta | \beta_1, \beta_2, \alpha, I, D_i \propto \exp[-\theta(1 - e^{-\beta_1 t} - \theta(1 - e^{-\beta_2 t}))] \cdot e^{-\beta_1 \theta} \cdot \theta^n \cdot \theta^{a_i} \sim \Gamma(a_1 + n, \beta_1 + 2 - e^{-\beta_1 t} - e^{-\beta_2 t}) \quad (3.11)$$

$$(2) \beta_1 | \theta, \beta_2, \alpha, I, D_i \propto \beta_1^{\sum_{i=1}^{I-1} I_i} \left(\prod_{i=1}^{I-1} \exp(-\beta_1 x_i) \right) \cdot \{\exp(\theta e^{-\beta_1 t})\} \cdot \pi_1(\beta_1) \quad (3.12)$$

$$(3) \beta_2 | \theta, \beta_1, \alpha, I, D_i \propto \beta_2^{\sum_{i=1}^{I-1} I_i} \left(\prod_{i=1}^{I-1} \exp(-\beta_2 x_i^a) \right) \cdot \{\exp(\theta e^{-\beta_2 t})\} \cdot \pi_2(\beta_2) \quad (3.13)$$

$$(4) \alpha | \theta, \beta_2, \beta_1, I, D_i \propto \alpha^{\sum_{i=1}^{I-1} I_i} \left(\prod_{i=1}^{I-1} x_i^a \exp(-\beta_2 x_i^a) \right) \cdot \{\exp(\theta e^{-\beta_2 t})\} \cdot \pi_3(\alpha) \quad (3.14)$$

깃스샘플링 알고리즘을 시행하기 위해서는 각 모수의 사전분포의 초기값을 대입하여 위 식의 $\theta | \beta_1, \beta_2, \alpha, I, D_i$ 의 분포로부터 θ 을 추출하고 갱신된(추출된) θ 와 나머지 초기값을 대입하여 $\beta_1 | \theta, \beta_2, \alpha, I, D_i$ 의 분포로부터 β_1 를 추출한다. 그리고 갱신된 θ, β_1 와 나머지 초기값을 대입하여 $\beta_2 | \theta, \beta_1, \alpha, I, D_i$ 로부터 β_2 를 추출하고 갱신된 θ, β_1, β_2 와 나머지 초기값을 대입하여 $\alpha | \theta, \beta_2, \beta_1, I, D_i$ 로부터 α 를 추출하는 반복형식을 통해 독립적으로 최종표본을 얻을 수 있다. 반복과정에서 θ 는 감마분포에서 표본이 발생하여 사용할 수 있지만 β_1, β_2, α 는 일정한 분포를 알지 못하므로 메트로폴리스 알고리즘[11]에 의해서 발생된다. 즉 깃스단계에서 메트로폴리스와 깃스샘플링을 혼합하여 사용하는 방법은 Kuo 와 Yang[15]에 의해 제시되었다

4. 베이지안 추론과 모형선택

베이지안 예측은 고장절단상황에서 고려하는 것이 일반적이기 때문에 n 번째 고장까지 조사하여 t 를 x_n

으로 대치함으로써 해결할 수 있다[14]. 만일 시점 x_i 로부터 x 만큼 떨어진 예측패턴은 미래생존함수(future survival function)의 통해서 해결할 수 있고 다음과 같은 방법을 통하여 추론이 이루어 진다[5].

$$\begin{aligned} E(S(x)|D_x) &= E[E(P(X_{n+1} - x_n > x) | \beta, D_x) | D_x] \\ &= E[\exp\{-m(x_n + x) + m(x_n | \beta)\} | D_x] \\ &= \int \dots \int \exp\{-m(x_n + x | \beta) + m(x_n | \beta)\} f(\beta | D_x) d\beta \end{aligned} \quad (4.1)$$

식(4.1)은 Gelman과 Rubin[10]이 제시한 깃스추출방법을 사용하여 다음과 같이 적용을 하였다.

$$\widehat{S}(x | D_x) = \frac{2}{RI} \sum_{r=1}^R \sum_{i=\frac{I}{2}+1}^I \exp[-m(x_n + x | \beta^{(i,r)}) + m(x_n | \beta^{(i,r)})] \quad (4.2)$$

(4.2)식에서 $\beta^{(i,r)}$ 은 깃스반복알고리즘을 사용하여 i 번 반복 후 r 번 적용을 통해 β 를 발생시킨 깃스 표본추출을 의미하고 $i = I/2 + 1, \dots, I, r = 1, \dots, R$ 이며 충분히 큰 R 과 짝수인 I 를 대입한 깃스표본추출을 사용하면 식(4.2)의 추정치를 유도 할 수 있다. 예를들어 앞 절에서 제시한 Goel-Okumoto와 Weibull의 중첩과정의 경우에는 다음과 같이 계산 할 수 있다.

$$\begin{aligned} \widehat{S}(x | D_x) &= \frac{2}{RI} \sum_{r=1}^R \sum_{i=\frac{I}{2}+1}^I \exp[-\theta^{(i,r)} \{2 - \exp(\beta_1^{(i,r)} x_{i+1}) - \exp(\beta_2^{(i,r)} x_i^{a^{(i,r)}})\} + \theta^{(i,r)} \{2 - \exp(\beta_1^{(i,r)} x_i) - \exp(\beta_2^{(i,r)} x_i^{a^{(i,r)}})\}] \end{aligned} \quad (4.3)$$

NHPP는 여러 가지 모형이 존재할 수 있으며 어떤 모형이 적당한가는 모형의 정도(adequacy)와 모형선택(model selection)으로 평가할 수 있다 NHPP는 여러 과정들이 존재할 수 있으므로 최적과정 선택은 Dawid [7]와 Kuo와 Yang[15]등이 제안한 PCPO(Prequencial Conditional Probability Ordinate)를 사용하여 해결할 수 있다. 따라서 미래시점 x_{i+1} 에 대한 PCPO는 (x_1, \dots, x_i) 이 주어진 미래의 관찰시점 X_{i+1} 의 조건부밀도 $c_{i+1} = p(x_{i+1} | D_x), i \geq 1$ 에 의해 정의된다. 이 PCPO는 과거의 데이터가 주어진 상태에서 X_{i+1} 의 값을 예측하기 때문에 과정선택을 하는데도 적절한 도

구가 된다. 계열 $\{X_i\}_{i \geq 1}$ 이 주어지면 PCPO는 다음과 같은 식을 통하여 계산된다.

$$\begin{aligned} p(X_{i+1} | D_x) &= \int \dots \int p(X_{i+1} | \beta, D_x) p(\beta | D_x) d\beta \\ &= \int \dots \int \lambda(X_{i+1}) \exp(-m(X_{i+1}) \\ &\quad + m(x_i)) \times p(\beta | D_x) d\beta \end{aligned} \quad (4.4)$$

(4.4)식을 이용하여 앞 절에서 제시한 Goel-Okumoto와 Weibull의 중첩과정의 경우에는 다음과 같이 계산할 수 있다.

$$\begin{aligned} p(X_{i+1} | D_x) &= \frac{2}{IR} \sum_{i=1}^R \sum_{j=\frac{i}{2}+1}^i [\theta^{(i,s)} \beta_1^{(i,s)} \exp(-\beta_1^{(i,s)} x_{i+1}) + \\ &\quad + \theta^{(i,s)} \beta_2^{(i,s)} a^{(i,n)} x_{i+1}^{a^{(i,n)}-1} \exp(-\beta_2^{(i,s)} x_{i+1}^{a^{(i,n)}})] \\ &\quad \times \exp[-\theta^{(i,n)} \{ 2 - \exp(\beta_1^{(i,n)} x_{i+1}) - \exp(\beta_1^{(i,n)} \\ &\quad \quad x_{i+1}^{a^{(i,n)}}) + \theta^{(i,n)} \{ 2 - \exp(\beta_2^{(i,n)} x_i) \\ &\quad \quad - \exp(\beta_2^{(i,n)} x_i^{a^{(i,n)}}) \}] \end{aligned} \quad (4.5)$$

모형 선택을 위하여 사전자료를 사용하여 $c_{i+1} = p(x_{i+1} | D_x, \cdot), i \geq 1$ 을 각 시점에 대하여 계산하여 곱하면 베이지안 예측우도기준은 예측밀도가 되고 다음과 같이 표현 할 수 있다[23].

$$Pd = \prod_{i=1}^n C_{i+1} = p(x_1, x_2, \dots, x_n) \quad (4.6)$$

이 예측밀도를 최대화하는 모형이 더 좋은 모형으로 간주된다[7]. 모형선택의 대안으로서 본 논문에서는 상대오차의 합(the sum of relative errors)을 제시하고자 한다. 상대오차의 합은 다음과 같이 정의 할 수 있다.

$$RE(a) = \sum_{i=1}^n \left| \frac{\text{참값모형} - \text{추정값모형}}{\text{참값모형}} \right|$$

단, 참값모형은 각각의 모형에 관찰된 자료와 진의(true)의 모수값(본 논문에서 혼합모형의 경우 θ, β_1, β_2)을 대입했을 때의 값이고 추정값모형은 각각의 모형에 관찰된 자료와 깁스 알고리즘에 의해 구해진 모수 추정값을 대입했을 때의 값이고 $RE(a)$ 의 값이 작으면 보다 좋은 모형이라고 할 수 있을 것이다.

따라서 다음과 같은 강도함수를 가지는 부분모형

(submodels)을 고려하여 모형선택을 할 수 있다.

$$\lambda_1(t), \lambda_2(t), \lambda_1(t) + \lambda_2(t)$$

5. 수치적인 예

모수추정과 모형선택을 위해 다음과 같은 GOS의 강도함수를 가진 NHPP를 가진 세 과정을 이용하고 자 한다. 즉, 부분함수 $\lambda_1(t) = \theta \beta_1 \exp(-\beta_1 t)$ 와, $\lambda_2(t) = 2\theta \beta_2 t \exp\{-\beta_2 t^2\}$, 그리고 중첩모형 $\lambda_1(t) + \lambda_2(t)$.

단, $\theta > 0, \beta_i > 0 (i = 1, 2)$, λ_1 은 Goel-Okumoto 과정의 강도함수이고, λ_2 는 Weibull과정의 특수한 강도함수($\alpha = 2$)를 의미한다. 데이터들은 일반적으로 NHPP에 대한 자료생성에 이용되는 Lewis와 Shedler[25]에 의해 제시된 Thinning알고리즘을 이용하여 IMSL RANPP($t = 2$) 루틴에 의하여 모수 $\theta = 50, \beta_1 = \beta_2 = 0.0005$ 을 입력하여 시뮬레이트되었고 이 루틴에 의한 자료들은 다음과 같다.

0.036, 0.083, 0.127, 0.179, 0.231, 0.266, 0.339, 0.380, 0.450, 0.505, 0.533, 0.573, 0.624, 0.664, 0.697, 0.743, 0.769, 0.803, 0.858, 0.900, 0.936, 0.977, 1.032, 1.105, 1.146, 1.186, 1.243, 1.288, 1.319, 1.370, 1.407, 1.441, 1.479, 1.538, 1.607, 1.658, 1.772, 1.836, 1.877, 1.900,

깁스추출법을 시행하기 위하여 각 모수들에 대한 사전분포는 상대적으로 확산분포(diffuse priors)를 제시하여 경험적 결과를 유도하였다. 사전분포는 $\theta = \Gamma(150, 3)$ 을 선택 이용하였고, β_1, β_2 에 대한 비정보사전밀도(noninformative prior density)는 각각 $\pi_1(\beta_1) = 1/\beta_1, \beta_1 > 0, \pi_2(\beta_2) = 1/\beta_2, \beta_2 > 0$ 을 가정하였다. 이 세 과정에 대한 확산사전분포는 <표 1>에 요약되었다. 반복이 얼마나 필요한지를 결정하기 위하여 Gelman & Rubin[10]이 제시한 방법을 사용하여 깁스추출법의 수렴성을 고려하여 제안된 사전분포를 가지고 FORTRAN IMSL언어를 이용하여 충분한 1000의 반복을 시행하였다. 모든 경험적 결과들의 깁스추출은 1000회의 반복에 50의 적용을 시행하였다.

3절에서 제시된 Goel-Okumoto와 Weibull의 중첩 과정에 대한 조건부 밀도를 이용한 깁스 알고리즘 시행 단계는 다음과 같이 이루어진다.

(0단계)

$\Gamma(150, 3)$ 와 $\Gamma(50, 0.1), \Gamma(50, 0.01)$ 의 분포에서

데이터를 랜덤추출하여 초기값을 각각 $\theta^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}$ 를 정하고 $I_{i1} (i=1, \dots, 40) \sim \text{Bernoulli}(1, p_{i1})$ 을 계산한다. 단,

$$p_{i1} = \frac{\theta^{(0)} \beta_1^{(0)} e^{-\beta_1^{(0)} x_i}}{\theta^{(0)} \beta_1^{(0)} e^{-\beta_1^{(0)} x_i} + 2 \theta^{(0)} \beta_2^{(0)} x_i e^{-\beta_2^{(0)} x_i^2}}$$

이고 $I_{i2} = 1 - I_{i1}$ 이 된다.

<표 1> 사전분포

| 강도 함수 $\lambda(t)$ | Goel-Okamoto과정 $\lambda_1 = \theta \beta_1$ $\exp(-\beta_1 t)$ | 와이블 과정 $\lambda_2 = \theta \beta_2 \alpha t$ $\exp(-\beta_2 t^\alpha)$ | 중첩과정 $\lambda_1 + \lambda_2$ |
|--------------------|--|--|--|
| 사전 분포 | $\theta = \Gamma(150, 3)$ $\beta_1 = \Gamma(50, 0.1)$ | $\theta = \Gamma(150, 3)$ $\beta_2 = \Gamma(50, 0.01)$ | $\theta = \Gamma(150, 3)$ $\beta_1 = \Gamma(50, 0.1)$ $\beta_2 = \Gamma(50, 0.01)$ |

(1-1단계)

$\beta_1 = \beta_1^{(0)}, \beta_2 = \beta_2^{(0)}$ 로 고정시켰을 경우 (3.10)에 대입하여 생성된 랜덤포본 하나를 $\theta^{(1)}$ 이라 한다. 즉

$$\theta^{(1)} \sim \Gamma(50 + 40, b_1 + 2 - e^{-\beta_1^{(0)} 1.9} - e^{-\beta_2^{(0)} 1.9^2})$$

(1-2단계)

$\theta = \theta^{(1)}, \beta_1 = \beta_1^{(0)}$ 로 고정시켰을 경우 (3.11)에 대입하여 생성된 다음과 같은 과정 이후에 생성된 랜덤포본 하나를 $\beta_1^{(1)}$ 이라 한다.

$$f(\beta_1) \propto \beta_1^{(0) \sum_{i=1}^{40} I_{i1}} \left(\prod_{i=1}^{40} \exp(-\beta_1^{(0)} x_i) \right) \cdot \{ \exp(\theta^{(1)} e^{-\beta_1^{(0)} 1.9}) \} \cdot (1/\beta_1^{(0)})$$

분산을 보다 크게 한 감마분포 $\Gamma(1, 10^{-4})$ 에서 β_1^w 을 발생하고 균등분포 (0,1]에서 확률변량을 w 라고 하면 $\log f(w) \leq \log f(\beta^w) - \log f(\beta_1)$ 을 만족하면 β^w 을 β_1 으로 간주되고 만족하지 않으면 β_1 을 β^w 으로 대치 되면서 만족할 때까지 계속 반복(iteration)된다.

(1-3단계)

$\theta = \theta^{(1)}, \beta_2 = \beta_2^{(0)}$ 로 고정시켰을 경우 (3.11)에 대

입하여 생성된 다음과 같은 과정 이후에 생성된 랜덤포본 하나를 $\beta_2^{(1)}$ 이라 한다.

$$f(\beta_2) \propto \beta_2^{(0) \sum_{i=1}^{40} I_{i2}} \left(\prod_{i=1}^{40} \exp(-\beta_2^{(0)} x_i^2) \right) \cdot \{ \exp(\theta^{(1)} e^{-\beta_2^{(0)} 1.9^2}) \} \cdot (1/\beta_2^{(0)})$$

분산을 보다 크게 한 감마분포 $\Gamma(1, 10^{-4})$ 에서 β_2^w 을 발생하고 균등분포 (0,1]에서 확률변량을 w 라고 하면 $\log f(w) \leq \log f(\beta^w) - \log f(\beta_2)$ 을 만족하면 β^w 을 β_2 으로 간주되고 만족하지 않으면 β_2 을 β^w 으로 대치 되면서 만족할 때까지 계속 반복(iteration)된다.

(2단계)

(1-1단계), (1-2단계), (1-3단계)로부터 고정시킨 θ, β_1, β_2 의 값을 가장 최근에 생성된 랜덤포본의 값으로 대체하면서 (1-1단계), (1-2단계), (1-3단계)를 충분히 큰 수 (본 논문에서는 1000번) 만큼 반복 수행한다. 이렇게 하여 얻은 최종포본을 $(\theta_1^{(1000)}, \beta_1^{(1000)}, \beta_2^{(1000)})$ 이라 한다 여기서 수렴성을 고려하여 앞의 500번 반복은 제거하고 뒷 부분 500번의 반복만 정보를 얻는데 사용된다..

(3단계)

(1-1단계), (1-2단계), (1-3단계)를 다시 (50)번 반복 적용하면 총 2500개의 랜덤포본이 얻어진다

$$(\theta_j^{(i)}, \beta_{1j}^{(i)}, \beta_{2j}^{(i)}) \text{ 단, } i=1, \dots, 500, j=1, \dots, 50.$$

(4단계)

최종적인 결과에 의해 θ, β_1, β_2 의 추정은 다음과 같이 사후평균을 계산 할 수 있다.

$$\hat{\theta}_{Cibbs} = \frac{1}{2500} \sum_{i=1}^{500} \sum_{j=1}^{50} \theta_j^{(i)},$$

$$\hat{\beta}_{1Cibbs} = \frac{1}{2500} \sum_{i=1}^{500} \sum_{j=1}^{50} \beta_{1j}^{(i)}, \quad \hat{\beta}_{2Cibbs} = \frac{1}{2500} \sum_{i=1}^{500} \sum_{j=1}^{50} \beta_{2j}^{(i)}$$

위와 같은 방법으로 계산된 사후평균은 <표 2>에 요약하였고, 중첩과정에서 $\{\hat{\theta}, \hat{\beta}_1, \hat{\beta}_2\}$ 의 베이지안 추

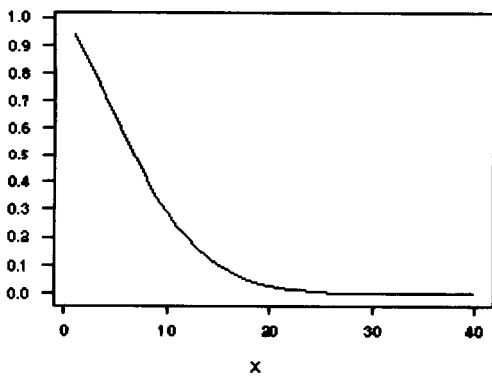
정은 여러모형 들과 비교되었다. <표 2>는 이 세 과정에 대한 PCPO의 값(Pd)도 위와 같은 유사한 알고리즘을 통해 계산된 것이다. 그 결과 예측된 것처럼 단순과정보다 중첩과정이 Pd의 값이 제일 크므로 중첩과정이 더 좋은 과정으로 간주할 수 있다 <그림 1>은 시간의 흐름에 따라 비증가함수(nonincreasing function)의 패턴을 가진 미래생존함수 $\hat{S}(x|D_{x_n})$ 을 나타낸 것이고, <그림 2>는 시간에 따라 중첩모형에 대한 진의(true)의 강도함수와 베이스 추정치에 의한 강도함수가 비교하여 본 결과 거의 밀접한 추세를 보이고 있으므로 수렴하고 있다고 간주할 수 있다.

<표 2> 사후 평균과 log(Pd)

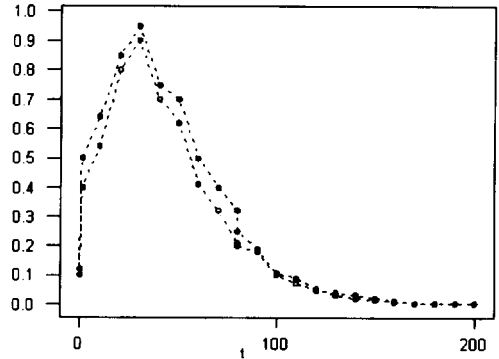
| 추 세 | Goel-Okumoto 과 정 | 와이블과정 | 중첩과정 |
|-------------------|---|---|--|
| Posterior mean | $\hat{\theta} = 52.15$ $\hat{\beta}_1 = 0.00524$ | $\hat{\theta} = 53.53$ $\hat{\beta}_2 = 0.00024$ | $\hat{\theta} = 52.83$ $\hat{\beta}_1 = 0.00582$ $\hat{\beta}_2 = 0.00039$ |
| log(Pd) | -67.53 | -65.78 | -63.01 |

<표 3> 상대오차값의 합

| 모 형 | n | 상대오차의 합 |
|----------------|----|----------|
| Goel-Okumoto과정 | 40 | 9.278453 |
| 와이블과정 | 40 | 6.876537 |
| 중첩과정 | 40 | 2.576232 |



(그림 1) 시간(x)에 대한 중첩모형에 대한 예측생존함수 $\hat{S}(x|D_{x_n})$



(그림 2) 시간(t)에 대한 진의(true)의 강도함수 (o)와 베이스 추정치에 의한 강도함수(●)의 비교

6. 결 론

소프트웨어 시스템이 복잡해지면 고장의 원인이 하나의 강도함수에 의해서만 일어나지 않고 여러원인이 중첩되어 발생 할 수 있다. 이러한 복잡한 시스템에 의한 우도함수가 적분하기 난해하므로 반복표본을 이용하는 수치해석적 방법인 깁스 알고리즘이 제안되었다.

깁스 알고리즘은 Markov Chain Monte Carlo기법으로 바람직한 사후분포로서 정상성분포를 가지는 마코프 체인에 따라 표본들이 변화하는 알고리즘이다. 이 마코프체인의 추이 추정치는 일반적으로 조건부밀도함수의 곱으로 이루어진다. NHPP모형의 중첩에 대하여 한 변수는 소프트웨어에 남아있는 수많은 오류들이며, 다른 변수는 포아송 확률들 중에서 척도모수(scale parameter)의 변화에 따라서 어떤 요소가 도움을 주는지를 나타내는 잠재변수들을 이용하여 깁스알고리즘을 적용하였고 단순모형과 중첩모형의 비교를 위해 사후베이지요인을 이용하여 모형선택을 하였다. 추가적으로 수치적인 예에서 Goel-Okumoto 모형과 Weibull 모형을 선택하고 NHPP의 자료는 Lewis와 Shedler[25]에 의해 제시된 Thinning 알고리즘을 이용하여 발생된 자료를 이용하고 사전분포는 상대적으로 확산분포를 이용한 모수추정과 사후베이지요인을 이용한 모형선택을 한 결과 단순모형들 보다 중첩모형이 좋은 모형으로 간주할 수 있음을 보여 주었다.

끝으로 본 논문에서의 내용을 참고로 하여 소프트웨어 중첩신뢰모형에 대한 극한분포 이론과 응용에 관한 연구가 기대된다.

참 고 문 헌

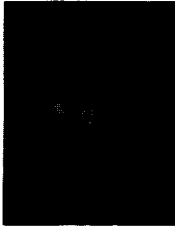
- [1] Aitkin, M., "Posterior Bayes Factors," *Journal of the Royal Statistical Society B*, pp.111-142, 1991.
- [2] Berger, J. O. and Sun, D., "Bayesian Analysis For The Poly-Weibull Distribution," *Journal of the American Statistical Association*, 88, pp. 1412-1418, 1993.
- [3] Box, G., "Sampling and Bayes' Inference in Scientific Modeling and Robustness (with discussion)", *Journal of the Royal Statistical Society, Ser. A*, 143, pp.382-430, 1980.
- [4] Casella, G. and George, E. I., "Explaining the Gibbs Sampler," *The American Statistician*, 46, pp. 167-174, 1992.
- [5] Cinlar, E., '*Introduction to Stochastic Process*', New Jersey: Prentice-Hall, 1975.
- [6] Cox, D. R. and Lewis, P. A., '*Statistical Analysis of Series of Events*,' London: Methuen, 1966.
- [7] Dawid, A. P., "Statistical Theory: The Prequential Approach," *Journal of the Royal Statistical Society, Ser. A* 147, pp.278-292, 1984.
- [8] Geisser, S., and Eddy, W., "A Predictive Approach to Model Selection," *Journal of the American Statistical Association*, 74, pp.153-160, 1979.
- [9] Gelfand, A. E. and Smith, A. F. M., "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, pp.398-409, 1990.
- [10] Gelman, A. E., and Rubin D., "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, pp.457-472.
- [11] Greenberg, E. and Chib, S., "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49, pp.327-335, 1995
- [12] Jelinski, Z., and Moranda, P. B., "*Software Reliability Research*, in *Statistical Computer Performance Evaluation*," ed. W. Freiberger, New York : Academic Press, pp.465-497, 1972.
- [13] Joe, H., "Statistical Inference for General Order Statistics and Nonhomogeneous Poisson process Software Reliability Models," *IEEE Transactions on Software Engineering*, 15, pp.1485-1490, 1989
- [14] Kuo, L., and Yang, T. Y., "Bayesian Computation of Software Reliability," *Journal of Computational and Graphical Statistics*, pp.65-82, 1995.
- [15] Kuo, L., and Yang, T. Y., "Bayesian Computation for Nonhomogeneous Poisson process in Software Reliability", *Journal of the American Statistical Association*, 91, pp.763-773, 1996.
- [16] Langberg, N., and Singpurwalla, N. D., A Unification of Some Software Reliability Models, *SIAM Journal on Scientific and Statistical Computing*, 6, pp.781-790, 1985.
- [17] Lawless, J. F., "*Statistical Models and Methods for lifetime Data*," New York: John Wiley & Sons, 1982.
- [18] Musa, J. D. and, Iannino, A., and Okumoto, K., "*Software Reliability: Measurement, Prediction, Application*," New York: McGraw Hill, 1987.
- [19] Musa, J. D., and Okumoto, K., A, "Logarithmic Poisson Execution Time Model for Software Reliability Measurement", in *Proceedings Seventh International Conference on Software Engineering Orlando*, pp.230-238, 1984.
- [20] Parzen, E., '*Stochastic Process*,' San Francisco: Holden-Day, 1962.
- [21] Raftery, A. E., "Inference and Prediction for a General Order Statistic Model with Unknown Population Size," *Journal of the American Statistical Association*, 92, pp.1195-1212, 1997.
- [22] Pesnick, S. I., '*Extreme Values, Regular Variation, and Point Process*,' Berlin: Springer-Verlag, 1987.
- [23] Shih, D. and Day, D. K., "Semiparametric Bayesian Analysis of Survival Data", *Journal of the American Statistical Association*, 81, pp.82-86, 1987.
- [24] Tanner, M. and Wong, W., "The Calculation of Posterior Distributions by Data Augmentation" (with discussion), *Journal of the American Statistical Association*, 81, pp.82-86, 1987.
- [25] "*USER'S MANUAL STAT/LIBRARY FORTRAN*

Subroutines for statistical analysis," LMSL,
Volume 3, pp.1050-1054, 1987.



이 병 수

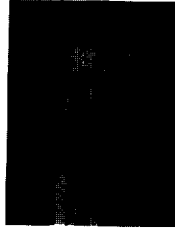
e-mail : bsi@lion.inchon.ac.kr
1999년 8월 경기대학교 대학원 전
자계산학과 졸업(이학박사)
1981년~현재 시립 인천대학교
전자계산학과 교수
관심분야 : 소프트웨어공학, 의사결
정지원시스템, IT.



김 희 철

e-mail : kim1458@cakra.dongguk.ac.kr
1997년 2월 동국대학교 대학원 통
계학과 졸업(이학박사)
1993년~현재 인천대학교 전자계
산학과 강사
1996년~현재 청주대학교 응용통
계학과 강사

1998~현 재 제주정보기술연구소 소프트웨어 개발 책
임 연구원
관심분야 : 소프트웨어 신뢰성공학, 전산통계, 전산수
학 등



백 수 기

e-mail : skDaik@kuic.kyonggi.ac.kr
1992년 8월 동국대학교 대학원 전
산통계학과 졸업(이학박사)
1981년 3월~현재 경기대학교 전
자계산학과 교수
관심분야 : 컴퓨터 통신, 자료처리



정 관 희

e-mail : KHCHUNG@kuic.kyonggi.ac.kr
1992년 8월 동국대학교 대학원 전
산통계학과 졸업(이학박사)
1980년 3월~현재 경기대학교 전
자계산학과 교수
관심분야 : 전산통계, 컴퓨터 시뮬
레이션



윤 주 용

e-mail : yooNJy@free way.co.kr
1999년 현재 경기대학교 대학원 전
자계산학과 박사과정중
1999년 현재 한국도로공사 정보처
정보운영부장
관심분야 : 위기관리시스템, 컴퓨
터 통신, OS