

Inference of Parameters for Superposition with Goel-Okumoto model and Weibull model Using Gibbs Sampler

Kiheon Choi¹⁾ and Heecheul Kim²⁾

Abstract

A Markov Chain Monte Carlo method with development of computation is used to be the software system reliability probability model. For Bayesian estimator, considering computational problem and theoretical justification, we studies relation Markov Chain with Gibbs sampling. Special case of GOS with Superposition for Goel-Okumoto and Weibull models using Gibbs sampling and Metropolis algorithm considered. In this paper discuss Bayesian computation and model selection using posterior predictive likelihood criterion. We consider in this paper data using method by Cox-Lewis. A numerical example with a simulated data set is given.

1. 서 론

소프트웨어는 만들어지거나 또는 새로운 기능을 추가 할 때에 고장이 일어 날 수가 있다. 시스템의 수리가 가능할 때에 구간 $(0, t]$ 에서 GOS(general order statistics)모형을 적용하기 위하여 소프트웨어 테스트 시점에서 미지의 N 개의 결함 수가 있다고 가정하자. 관찰된 고장 시점은 N 으로 부터 추출된 처음 n 개의 순서통계량은 R^+ (양의 실수의 집합)에서 만들어진(supported) 밀도함수 f 을 가진 독립이고 동일한 분포(*i.i.d.*)의 관찰 값으로 모형화 될 수 있다. 만일 f 가 지수분포인 경우에는 Jelinski-Moranda 모형으로 알려져 있다. 따라서 f 는 여러 가지 분포를 따른다. 예를 들면 와이블, 파레토, 감마분포인 경우를 고려 할 수 있다. 이러한 경우의 모형을 와이블, 파레토, 감마 순서통계량 모형으로 제시 할 수 있다.

GOS 모형(Raftery(1987))에서 N 은 평균 θ 을 가진 포아송분포를 한다는 가정이 필요하다. 즉 $M(t)$ 는

$$m(t) = \theta F(t) \quad (1.1)$$

을 가진 *NHPP*(nonhomogeneous Poission process)가 된다(Kuo & Yang(1996)). 단, F 는 밀도함수 f 를 따르는 누적분포함수를 의미한다. 이 때 $\lim_{t \rightarrow \infty} m(t)$ 가 유한 (finite) 할 때의 과정을

1) Professor, Department of Statistics, Duksung Woman's University, Seoul, 132-714, Korea

2) Lecturer, Department of Statistics, Dongguk University, Seoul, 100-715, Korea

$NHPP(1)$ 이라고 표기하자. 특히 Jelinski-Moranda 모형에서 $F(t) = (1 - e^{-\beta t})$ 일 때 $M(t)$ 는 평균함수 $m(t) = \theta(1 - e^{-\beta t})$ 을 가진 $NHPP$ 가 된다. 이것을 Goel-Okumoto 과정(1979)이라고 알려져 있다. 반면에 평균함수가 $m(t) = \theta[1 - (1 + \beta t)e^{-\beta t}]$ 이면 Ohba-Yamada 과정이 된다.

$M_j(t)$ 을 $(0, t]$ 사이에서 강도함수가 $\lambda_j(t|\beta_j)$ 을 가진 j 번째 요소로부터 고장이 발생하는 $NHPP$ 라고 표현하자. 단, $\lambda_j(t|\beta_j)$ 은 벡터도 될 수 있는 미지의 모수 β_j 로 알고 있는 값이다. 추가적으로 j 번째 요소에 의해 발생한 고장의 수 $M_j(t)$ ($j = 1, 2, \dots, J$)는 독립이라고 가정하자.

따라서 중첩과정으로 알려진 $(0, t]$ 의 구간에서 발생한 고장의 총수 $M(t) = \sum_{j=1}^J M_j(t)$ 도

강도함수가

$$\lambda(t|\beta) = \lambda_1(t|\beta_1) + \lambda_2(t|\beta_2) + \dots + \lambda_J(t|\beta_J) \quad (1.2)$$

을 가지는 비동질적 포아송과정이 됨이 알려져 있다. (Cinlar(1975)). 단, $\beta = \{\beta_1, \beta_2, \dots, \beta_J\}$.

이런 단순한 강도함수들의 합으로 중첩된 강도함수를 가진 $NHPP$ 에 대한 베이저안 추론의 어려움도 깃스추출법(Gibbs sampling)을 사용하면 해소 할 수 있다.

D_t 을 t 시간까지 관찰된 데이터 시점이 n 개 즉, (x_1, x_2, \dots, x_n) 이라고 나타내면 이 데이터 집합에 순서를 부여하여 $0 < x_1 < x_2 < \dots < x_n < t$ 을 만족하는 데이터 집합으로 변환한다. 따라서, 순서화된 시점에 대한 점과정(point process)은 중첩과정을 가지고 부분 시스템에 대한 고장의 원인은 알려져 있지 않고, β_j ($j = 1, \dots, J$)에 대한 사전분포는 독립이라고 가정하자. 본 논문은 이러한 단조성의 강도함수의 가정을 완화하기 위하여 단순한 강도함수들을 가진 여러 개의 독립적인 $NHPP$ 들이 서로 겹쳐진 중첩(superposition)과정에 초점을 두고자 한다.

본 논문은 신뢰성이론에 많이 사용하는 Goel-Okumoto 패턴과 이것과 다른 소프트웨어 신뢰도와 고장시스템을 가진 Weibull 패턴과의 중첩과정을 제시하고, 또, 모형선택을 하고자 한다. 결국 중첩과정도 $NHPP$ 이기 때문에 이 이론에 적용할 수 있다.

본 논문의 2절에서는 중첩과정에 대한 깃스추출법을 3절에는 Goel-Okumoto와 Weibull 패턴의 중첩과정에 대한 분석을 한다. 그리고 4절에는 베이저안 추론과 모형선택을 시행하였고, 5절에는 수치적인 예를 제시하였다.

2. 중첩과정에 대한 깃스추출법

조건부 분포를 이용하여 깃스추출법을 적용할 수 있다 이 깃스추출법은 마코브체인몬테칼로(MCMC)의 기법중 하나이다. 마코브 연쇄의 전이분포는 여러 개의 조건부 밀도로서 이루어진다. 마코브 연쇄의 정상분포가 원하는 사후분포가 된다. 마코브 연쇄의 다중 열(multiple sequences)을 얻기 위하여 독립 초기점(starting points)을 가진 연쇄를 반복한다. 이러한 기법은 Tanner & Wong(1987), Gelfand & Smith(1990), Casella & George(1992), 그리고 Kuo & Yang(1995)등 많

은 학자들에 의해서 제시되고 있다. 어떤 시스템이 시간 t 까지 가동된다고 가정하고 관찰된 n 번의 고장시간을 x_1, x_2, \dots, x_n 이라고 표시하면 데이터 집합 D_t 는 $\{n, x_1, x_2, \dots, x_n; t\}$ 으로 구성된다. 만약 n 번째까지 관찰된 고장절단(failure truncated)과정이 주어진다면 데이터 집합 D_t 는 $\{x_1, x_2, \dots, x_n\}$ 으로 된다. 그리고 시간절단(time truncated)과정에 대한 데이터가 주어졌을 때 $NHPP$ 의 우도함수는 다음과 같다.

$$L_{NHPP}(\theta, \beta | D_t) = \left(\prod_{i=1}^n \lambda(x_i) \right) \exp(-m(t)) \tag{2.1}$$

단, $m(t) = \int_0^t \lambda(u) du$ 이고, $0 < x_1 < x_2 < \dots < x_n < t$, β 은 미지의 모수벡터를 의미한다. 이 절단과정에 대한 우도함수는 Lawless(1982)등이 이미 제시되었다. 고장절단과정은 t 대신에 x_n 을 대입하면 식(2.1)과 유사한 표현이 된다.

중첩 과정에서 θ, β 의 사후 밀도가 관찰하기 위해서 다음을 유도한다.

$$f(\theta, \beta | D_t) \propto \left\{ \prod_{i=1}^n [\lambda_1(x_i | \beta_1) + \lambda_2(x_i | \beta_2) + \dots + \lambda_j(x_i | \beta_j)] \right\} \cdot \exp\{-[m_1(t | \beta_1) + \dots + m_j(t | \beta_j)]\} \cdot \prod_{j=1}^j \pi_j(\beta_j) \cdot \pi(\theta). \tag{2.2}$$

단, β 는 모수벡터이고, $\pi_j(\beta_j)$ 는 β_j 에 대한 사전밀도, $\pi(\theta)$ 는 θ 에 대한 사전밀도를 의미하고 β_j 와 θ 은 독립이라고 가정하자.

식(2.2)에 있는

$$\prod_{i=1}^n [\lambda_1(x_i | \beta_1) + \lambda_2(x_i | \beta_2) + \dots + \lambda_j(x_i | \beta_j)] \tag{2.3}$$

의 표현 때문에 D_t 의 조건하에서 β 와 θ 의 조건부 밀도에 대한 형태를 구체화하기 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. 즉, 반복을 통해서 $f(I | \theta, \beta, D_t)$ 로부터 I 을 추출하고 $f(\theta | \beta, I, D_t)$ 로부터 θ 을 추출하고 $f(\beta | I, \theta, D_t)$ 로

부터 β 을 추출하는 반복 형식을 통해 사후분포를 얻게 된다. 각 $i(i=1, \dots, n)$ 에 대해 잠재변

수 $I_i = (I_{i1}, \dots, I_{ij})$ 을 설정하자.

j 번째 요소에 의해서 i 번째 고장의 원인이 발생했을 때를 $I_{ij} = 1$ 이라고 하자. 그 외에는 $I_{ij} = 0$ 이라 하고, $i = 1, \dots, n$, $\sum_{j=1}^J I_{ij} = 1$ 의 조건을 만족한다.

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

단,

$$p_{ij} = \lambda_j(x_i | \beta_j) / [\lambda_1(x_i | \beta_1) + \lambda_2(x_i | \beta_2) + \dots + \lambda_J(x_i | \beta_J)], \quad (2.4)$$

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

$$I_i \sim MN(1, (p_{i1}, \dots, p_{ij})), \quad i = 1, \dots, n. \quad (2.5)$$

I 와 D_i 의 조건하에서 β, θ 의 사후밀도는 다음과 같다.

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

조건부밀도는 I 와 D_i 의 조건하에서 β_j 의 독립 사후밀도들로 구성된 결합밀도가 된다. 그러므로 β_j 로부터 독립적으로 최종표본을 얻을 수 있다. 따라서 다음절에 중첩과정에 대한 베이지안 추정 기법을 제시한다.

3. Goel-Okumoto와 Weibull의 중첩과정

이 장에서는 Musa-Okumoto와 Weibull의 중첩과정을 제시하고자 한다. 즉, 각각의 강도함수는 다음과 같다.

$$\lambda_1(t) = \theta \beta_1 \exp\{-\beta_1 t\}, \quad \lambda_2(t) = \theta \beta_2 \alpha t^{\alpha-1} \exp\{-\beta_2 t^\alpha\} \quad (3.1)$$

단, $\alpha > 0$, $\beta_i > 0$ ($i = 1, 2$), $\theta > 0$ 을 만족한다. 식(2.1)을 이용하여 중첩과정에 대한 우도함수는 다

음과 같다.

$$L(\beta, \theta | D_t) = \left(\prod_{i=1}^n [\theta \beta_1 \exp(-\beta_1 x_i) + \theta \beta_2 \alpha x_i^{\alpha-1} \exp(-\beta_2 x_i^\alpha)] \right) \cdot \exp[-\theta(1-e^{-\beta_1 t}) - \theta(1-e^{-\beta_2 t^\alpha})] \quad (3.2)$$

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 $\pi_1(\beta_1), \pi_2(\beta_2)$ 와 $\pi_3(\alpha)$ 는 $\beta_i > 0 (i=1, 2), \alpha > 0$ 을 만족하는 임의의 사전분포이다.

β, θ 와 D_t 의 조건하에서의 I_i 의 조건부밀도는 다항분포가 되는데 이 분포는 모수가 1이고 셀 확률 (p_{i1}, p_{i2}) 을 가진다. 단, p_{ij} 는 $\lambda_j(x_i) / [\lambda_1(x_i) + \lambda_2(x_i)]$, $i=1, \dots, n, j=1, 2$ 이다. 이 중첩과정의 경우는 β, θ 와 D_t 의 조건하에서의 I_i 는 $i=1, \dots, n$ 에 대한 p_{i1} 의 모수를 가진 베르누이 분포로부터 발생된다. 단,

$$p_{i1} = \frac{\theta \beta_1 e^{-\beta_1 x_i}}{\theta \beta_1 e^{-\beta_1 x_i} + \theta \beta_2 \alpha x_i^{\alpha-1} e^{-\beta_2 x_i^\alpha}} \quad (3.3)$$

이고 $I_{i2} = 1 - I_{i1}$ 이 된다.

깁스추출법 알고리즘을 이용하기 위해서는 다음과 같은 4개의 조건식이 필요하다.

$$\begin{aligned} (i) \quad & \theta | \beta_1, \beta_2, \alpha, I, D_t \sim \Gamma(a_1 + n, b_1 + 2 - e^{-\beta_1 t} - e^{-\beta_2 t^\alpha}); \\ (ii) \quad & \beta_1 | \theta, \beta_2, \alpha, I, D_t \propto \beta_1^{\sum_{i=1}^n I_{i1}} \left(\prod_{i: I_{i1}=1}^n \exp(-\beta_1 x_i) \right) \cdot \{ \exp(\theta e^{-\beta_1 t}) \} \cdot \pi_1(\beta_1); \\ (iii) \quad & \beta_2 | \theta, \beta_1, \alpha, I, D_t \propto \beta_2^{\sum_{i=1}^n I_{i2}} \left(\prod_{i: I_{i2}=1}^n \exp(-\beta_2 x_i^\alpha) \right) \cdot \{ \exp(\theta e^{-\beta_2 t^\alpha}) \} \cdot \pi_2(\beta_2); \\ (vi) \quad & \alpha | \theta, \beta_1, \beta_2, \alpha, I, D_t \propto \alpha^{\sum_{i=1}^n I_{i2}} \left(\prod_{i: I_{i2}=1}^n x_i^\alpha \exp(-\beta_2 x_i^\alpha) \right) \cdot \{ \exp(\theta e^{-\beta_2 t^\alpha}) \} \cdot \pi_3(\alpha); \end{aligned} \quad (3.4)$$

여기서 β_1, β_2, α 는 메트로폴리스(Metropolis) 알고리즘(Greenberg & Chib 1985)에 의해서 발생된다. 둘 이상의 중첩과정을 가진 깁스 알고리즘도 유사한 방법으로 적용할 수 있다.

4. 베이지안 추론과 모수 선택

β 와 같은 모수를 추정하기 위해서는 깃스추출법의 경험적 측도(empirical measure)를 사용할 수 있다. 미래예측은 고장절단 상황에서 고려하는 것이 보다 쉽다. 즉, n 번째 고장까지 조사한다. 모든 사후분포는 앞절에서와 같이 t 를 x_n 으로 대치함으로써 수정할 수 있다.

x_i 로부터 거리 x 만큼 떨어진 시점에서의 미래 생존함수(future survival function)의 추론은 다음과 같은 식을 통해 얻을 수 있다. (Cinlar 1975, p.97 참조).

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

식(4.1)은 깃스추출법을 이용하거나 Rao-Blackwell정리를 사용하여 평가할 수 있다. 이것을 Gelman & Rubin(1992)이 제시한 MCMC방법을 적용하고자 한다.

*NHPP*는 여러 가지 모형이 존재 할 수 있다. 어떤 모형이 적당한가는 모형의 정도(adequacy)와 모형 선택(model selection)으로 평가 할 수 있다.

데이터 집합 (x_1, \dots, x_i) 가 주어진 상황에서 미래고장시점 X_{i+1} 의 조건부 밀도함수(conditional predictive density)는 다음과 같은 식을 통해서 계산 할 수 있다.

$$\begin{aligned} p(X_{i+1} | D_{x_i}) &= \int \cdots \int p(X_{i+1} | \beta, D_{x_i}) p(\beta | D_{x_i}) d\beta \\ &= \int \cdots \int \lambda(X_{i+1}) \exp\{-m(X_{i+1}) + m(x_i)\} \times p(\beta | D_{x_i}) d\beta. \end{aligned} \quad (4.2)$$

그러므로 (4.2)에 깃스추출법을 사용하여 X_{i+1} 의 $1-\alpha$ 의 신뢰구간을 설정할 수 있다. 예를 들어 와이블분포의 경우는 (4.2)을 이용하여 다음과 같이 유도된다.

$$p(X_{i+1} | D_{x_i}) = \int \int [\alpha \beta X_{i+1}^{\alpha-1} \exp(-\beta X_{i+1}^{\alpha} + \beta x_i^{\alpha})] \times p(\alpha, \beta | D_{x_i}) d\alpha d\beta.$$

Gelman & Rubin은 X_{i+1} 에 대한 95% 예측 신뢰구간은 $\{x_{i+1}^{(i,s)}\}_{s=1}^S$, $i = I/2+1$ 에서 발생된 표본에서 하한과 상한은 각각 2.5%와 97.5%의 분위수(quantiles)로 설정 할 수 있다고 하였다.

따라서, 과거자료 $D_{x_{i-1}}$ 을 정보로 하여 X_i ($i = 2, \dots, n$)의 예측구간을 설정할 수 있다. 모형의 정도는 실제 관측 자료 x_i 가 95%의 신뢰구간안에 포함되는가에 의해서 판단 할 수 있다.

$\beta^{(i,s)}$ 은 MCMC의 반복 알고리즘을 사용하여 i 번 반복 후 s 번 적용을 통해 β 을 발생시킨 깃

스 표본추출을 의미한다. 여기서, $i=I/2+1, \dots, I, s=1, \dots, S$ 을 의미하며 충분히 큰 S 과 짝수인 I 을 대입한 깁스 표본추출을 사용하여 Goel-Okumoto와 Weibull의 중첩과정은 다음과 같이 계산할 수 있다.

$$\begin{aligned}
 p(X_{i+1} | D_{x_i}) = & \\
 \frac{2}{IS} \sum_{s=1}^S \sum_{i=\frac{I}{2}+1}^I [& \theta^{(i,s)} \beta_1^{(i,s)} \exp(-\beta_1^{(i,s)} x_{i+1}) + \theta^{(i,s)} \beta_2^{(i,s)} \alpha^{(i,s)} x_{i+1}^{\alpha^{(i,s)}-1} \exp(-\beta_2^{(i,s)} x_{i+1}^{\alpha^{(i,s)}})] \\
 \exp[-\theta^{(i,s)} \{ & -2 + \exp(\beta_1^{(i,s)} x_{i+1}) + \exp(\beta_1^{(i,s)} x_i) \} + \theta^{(i,s)} \{ -2 + \exp(\beta_1^{(i,s)} x_i) + \exp(\beta_1^{(i,s)} x_i) \}] \quad (4.3)
 \end{aligned}$$

모형 선택을 위하여 각각 다음과 같은 강도 함수를 가지는 3개(= $2^J - 1$)의 부분모형(submodels)을 고려하였다.

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

어떤 모형에서 데이터 집합 D_{x_n} 에 대한 사후 주변우도(posterior marginal likelihood)를 M 으로 표시하면 다음과 같이 표현할 수 있다.

$$V_M = \int L_{NHPP}(\beta_M | D_{x_n}) \pi_M(\beta_M | D_{x_n}) d\beta_M. \quad (4.4)$$

단, β_M 은 모형 M 에서의 미지 모수의 집합이다. 그리고 측도 π_M 은 β_M 의 사후 밀도이다.

사후 베이스요인 기준 (posterior Bayes factor criterion)은 $V_{M_2}/V_{M_1} < 1$ 을 만족 할 때는 M_1 이 M_2 보다 더 좋은 모형이 된다. V_{M_2}, V_{M_1} 은 독립적으로 몬테카를로 기법을 적용하여 근사 계산을 할 수 있다. 즉, $V_{M_i} (i=1, 2)$ 는 깁스추출법을 이용하여 β 들에 의해서 $L_{NHPP}(\beta_M | D_{x_n})$ 에 의해서 계산된다. $2^J - 1$ 들의 모형 선택은 사후 베이스요인 기준 \hat{V} 을 최대로 하는 모형이 더 좋은 모형이 된다. (Aitkin,1991)

5. 수치적인 예

과정선택과 베이지안 추론에서 다음과 같은 특수한 강도함수를 가진 NHPP를 가진 세 과정을 제시하고자 한다. 즉, 부분함수 $\lambda_1(t) = \theta \beta_1 \exp\{-\beta_1 t\}, \lambda_2(t) = \theta \beta_2 \alpha t \exp\{-\beta_2 t^2\}$, 그리고 이들의 합 $\lambda_1(t) + \lambda_2(t)$. 단, $\theta > 0, \beta_i > 0 (i=1, 2)$. λ_1 은 Goel-Okumoto과정의 강도함수이고, λ_2 는 Weibull과정의 특수한 강도함수를 의미한다.

데이터들은

$$\lambda(t) = 50 \cdot 0.0005 [\exp(-0.0005 t) + 2 t \exp(-0.0005 t^2)]$$

을 가정하여 Lewis & Shedler(1962)에 의해 제시된 IMSL RANPP($t=130$ 이라고 가정)이라고 불리는 루틴에 의하여 시뮬레이트 되었다. 이 루틴에 의한 자료들은 다음과 같다.

12.1, 15.4, 20.1, 23.5, 24.2, 27.9, 30.5, 32.7, 35.4, 39.9, 42.3, 43.8, 45.2, 46.4, 49.7, 51.8, 57.1, 60.3, 62.9, 65.5, 69.1, 70.8, 71.2, 73.4, 75.9, 80.2, 85.5, 89.4, 90.1, 91.5, 92.3, 92.9, 93.4, 95.1, 95.9, 96.2, 97.1, 98.4, 98.9, 99.2, 101.1, 101.9, 110.4, 111.3, 119.2, 120.5 125.1 126.7 127.8 128.9
--

상대적으로 사전분포는 확산분포(diffuse prior)를 제시하여 경험적 결과를 유도하였다. 사전분포는 $\theta = \Gamma(150, 3)$ 을 선택 이용하였고, β_1, β_2 에 대한 비정보 사전밀도(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 확산 사전분포는 <표 1>에 요약되었다. 깃스추출법의 수렴성을 고려하여 Gelman & Rubin(1992)이 제시한 방법을 사용하여 제안된 사전분포를 가지고 충분한 50의 반복을 시행하였다. 모든 경험적 결과들은 깃스추출법내에서 50회의 반복에 1,000번의 적용을 시행하였다.

<표 2>는 모수에 대한 사후평균을 요약하였고, 중첩과정에서 $\{\hat{\theta}, \hat{\beta}_1, \hat{\beta}_2\}$ 의 베이저안 추정치는 세 과정과 비교되었다. <그림1>은 시간의 흐름에 따라 비증가함수(nonincreasing function)의 패턴을 가진 미래 생존함수 $\hat{R}(TIME | D_{x_n})$ 을 나타낸 것이고, <표 2>는 이 세 과정에 대한 사후 베이저안 요인의 값을 요약하였다. 그 결과 예측된 것처럼 단순과정보다 중첩과정이 사후 베이스 요인의 값이 제일 크므로 중첩과정이 더 좋은 과정으로 간주 할 수 있다. 모형선택의 대안으로서 상대오차(relative errors)의 합으로 비교할 수 있으며 다음과 같이 정의된다.

$$RE(l) = \sum_{i=1}^{n_i} (n_i - \hat{m}(x_i))^2 / \hat{m}(x_i)$$

단, n_i 는 $(0, x_i]$ 사이에 고장 발생횟수를 의미하고 $\hat{m}(x_i)$ 는 평균값함수의 베이저안 추정량을 의미한다. l 은 인덱스(index)모형을 나타낸다. 따라서 $RE(l)$ 가 작은 값을 가지는 모형이 좋은 모형이라고 간주 할 수 있다. <표 3>는 상대오차의 값의 합을 나타내었고 그 결과 중첩모형이 더 작은 상대오차의 합을 나타내고 있으므로 결국 중첩모형이 효율적 모형이 됨을 알 수 있다..

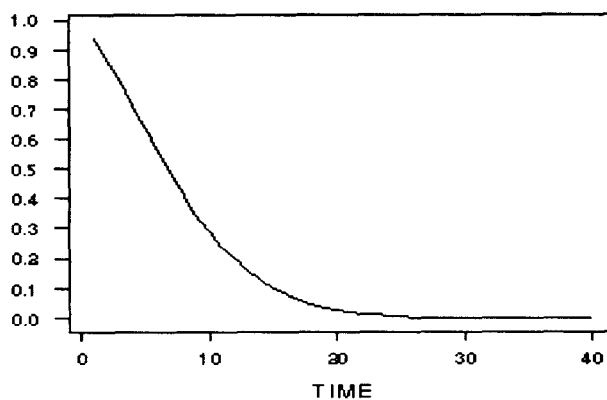
<그림 2>는 중첩모형에 대한 시뮬레이트된 자료와 50회의 반복에 1,000번의 적용을 시행한 베이스 추정치에 의한 자료가 비교되었다. 그 결과 거의 유사한 추세 즉, 수렴하고 있음을 보이고 있다

<표1> 사전분포

강도함수	Goel-Okumoto 과정	와 이 블 과 정	중 첩 과 정
$\lambda(t)$	$\lambda_1 = \theta \beta_1 \exp(-\beta_1 t)$	$\lambda_2 = \theta \beta_2 \alpha t \exp(-\beta_2 t^2)$	$\lambda_1 + \lambda_2$
사전분포	$\theta = \Gamma(150, 3)$	$\theta = \Gamma(150, 3)$	$\theta = \Gamma(150, 3)$
		$\beta_1 = \Gamma(50, 0.0001)$	$\beta_1 = \Gamma(50, 0.0001)$
		$\beta_2 = \Gamma(50, 0.0001)$	$\beta_2 = \Gamma(50, 0.0001)$

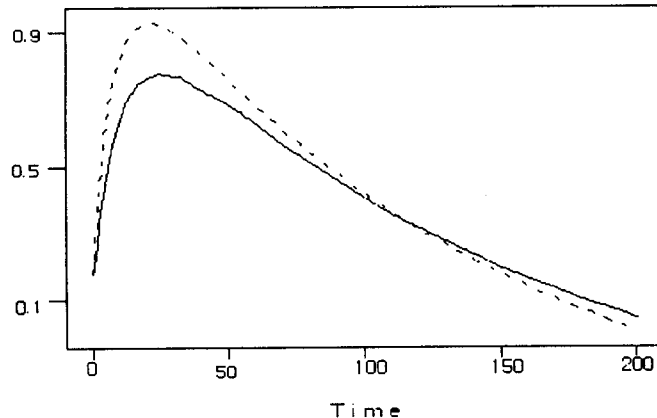
<표2> 사후 평균과 사후 베이지요인

형 태	Goel-Okumoto과정	와 이 블 과 정	중 첩 과 정
사후평균(Posterior mean)	$\hat{\theta} = 51.5307$	$\hat{\theta} = 53.5761$	$\hat{\theta} = 52.2463$
	$\hat{\beta}_1 = 0.009426$	$\hat{\beta}_2 = 0.000342$	$\hat{\beta}_1 = 0.000482$
			$\hat{\beta}_2 = 0.000132$
\ln (사후 베이지요인)	-66.15	-63.45	-60.17



<그림 1> 예측생존함수 $\hat{R}(TIME | D_n)$

<그림 2> 시간에 대한 $(\lambda(t) = 50 \cdot 0.0005 \cdot [\exp(-0.0005t) + 2t \exp(-0.0005t^2)] (\dots)$ 와



베이지 추정치에 의한 중첩과정(—)에 대한 강도함수

<표3> 상대오차의 합

형 태	Goel-Okumoto 과정	와이블 과정	중첩 과정
$RE(l)$	27.7966	13.6438	9.19660

References

- [1] Aitkin, M.(1991), "Posterior Bayes Factors," *Journal of the Royal Statistical Society*, B 111-142.
- [2] Blumenthal, S. and Marcus, R.(1975), "Estimating Population Size with Exponential Failure," *Journal of the American Statistical Association*, Vol. 70, 913-922.
- [3] Berger, J. O. and Sun, D.(1993), "Bayesian Analysis For The Poly-Weibull Distribution," *Journal of the American Statistical Association*, Vol. 88, 1412-1418.
- [4] Box, G.(1980), "Sampling and Bayes' Inference in Scientific Modeling and Robustness" (with discussion), *Journal of the Royal Statistical Society, Ser. A*, Vol. 143, 382-430.
- [5] Casella, G. and George, E. I. (1992), "Explaining the Gibbs Sampler," *The American Statistician*, Vol. 46, 167-174.
- [6] Cinlar, E.(1975), *Introduction To Stochastic Process*, New Jersey: Prentice-Hall.
- [7] Chib, S and Greenberg, E.(1995) Understanding the Metropolis-Hastings Algorithm, *The American Statistician*, Vol. 49, 327-335.
- [8] Geisser, S., and Eddy, W. (1979), "A Predictive Approach to Model Selection," *Journal of the American Statistical Association*, Vol. 74, 153-160.
- [9] Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, Vol. 85, 398-409.

- [10] Gelman, A. E., and Rubin D. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, Vol. 7, 457-472.
- [11] Hastings, W. K.(1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, Vol. 87 ,97-109.
- [12] Kuo, L., and Yang, T. Y. (1995), "Bayesian Computation of Software Reliability", *Journal of Computational and Graphical Statistics*, 65-82.
- [13] Kuo, L., and Yang, T. Y. (1996), "Bayesian Computation of Software Reliability", *Journal of the American Statistical Association*, Vol. 91, 763-773.
- [14] Lawless, J. F.(1982), "*Statistical Models and Methods for lifetime Data*," New York Jonn Wiley & Sons.
- [15] Musa,J. D. and., Iannino, A., and Okumoto, K.(1987), "*Software Reliability" Measurement, Prediction, Application*," New York: McGraw Hill.
- [16] Pettit, L. I., and Young, K. D. S. (1990), "Measuring the Effects of Observations on Bayes Factors," *Biometrika*, Vol. 77, 455-466.
- [17] Raftery, A. E. (1987), "Inference and Prediction for a General Order Statistic Model with Unknown Population Size," *Journal of the American Statistical Association*, Vol. 82, 1163-1168.
- [18] Tanner, M. and Wong, W.(1987), "The Calculation of Posterior Distributions by Data Augmentation" (with discussion), *Journal of the American Statistical Association*, Vol. 81, 82-86.
- [19] "*USER'S MANUAL STAT/LIBRARY(1987) FORTRAN Subroutines for statistical analysis*," IMSL, Vol. 3, 1050-1054