

MUSA-OKUMOTO와 ERLANG(2)의 중첩과정에 대한 베이지안 계산 연구

최기현¹⁾ 김희철²⁾

요약

컴퓨터의 발전에 따른 마코브체인 몬테카를로방법을 소프트웨어 신뢰확률모형에 이용하였다. 베이지안 추론에서 조건부분포를 가지고 사후부분포를 결정하는데 있어서의 계산 문제와 이론적인 정당성을 고려, 마코브연쇄와 메트로폴리스방법의 관계를 고찰하였으며, 특히 Mus-Okumoto와 Erlang(2)의 중첩모형에 대하여 깃스샘플링 알고리즘과 메트로폴리스 알고리즘을 활용하여 베이지안 계산과 예측 우도기준에 의한 모형선택을 제안하고 Cox-Lewis에 의해 제시된 Thing method를 이용한 모의실험자료를 이용하여 수치적인 계산을 시행하고 그 결과가 제시되었다.

1. 서론

수리가 가능한 시스템이나 소프트웨어 신뢰도에서 $(0, t]$ 사이에 발견된 고장의 수 $M(t)$ 을 모형화 하는데 비동질적 포아송 과정(nonhomogenous Poisson process, NHPP)을 널리 사용하여 왔다(Musa, Iannino and Okumoto, 1987). 강도함수(intensity function), 혹은 고장 발생률(rate of occurrence of failures, ROCOF) $\lambda(t) = dE[M(t)]/dt$ 은 흔히 t 에 대한 단조함수로 가정한다. 예를 들어 동질적 포아송 과정, Musa-Okumoto 과정, Weibull 과정 그리고 Cox-Lewis 과정은 각각 ROCOF를 상수, 부분함수, 멱함수, 대수선형함수를 가정하였다. 중첩과정에 대한 고장점(points of failure)들은 여러 개의구성요소의 점과정(component point processes)으로부터 발생된 고장점들의 결합으로 정의할 수 있다. 예를 들어 각각의 부분시스템의 고장은 단순한 ROCOF를 가진 NHPP에 의해서 과정화될 수 있다.

최근에 Kuo 와 Yang(1996)은 일반적 순서통계량 과정(general order statistics model, GOS) 과 기록값통계량 과정(record value statistics model, RVS)을 NHPP와 관련지어 단일화된 이론을 제시하였다. 이들은 GOS와 RVS의 과정이 NHPP가 됨을 보였다.

본 논문에서는 $M_j(t)$ 을 $(0, t]$ 사이에서 ROCOF가 $\lambda_j(t|\beta_j)$ 을 가진 j 번째 요소로부터 고장이 발생하는 NHPP라고 표현하자. 단, $\lambda_j(t|\beta_j)$ 은 벡터도 될 수 있는 미지의 모수 λ_j 을 가진 알고 있는 값이다. j 번째 요소에 의해 발생된 고장의 수 $M_j(t)(j = 1, 2, \dots, J)$ 는 독립이라고 가정하자. 그러면 $M(t) = \sum_{j=1}^J M_j(t)$. 즉, 중첩과정에서 $(0, t]$ 의 구간에서 발생된 고장의 총수는 강도함수가

1) (132-714) 서울시 도봉구 쌍문동 419, 덕성여자대학교, 통계학과

2) (100-715) 서울시 중구 필동 3가 26, 동국대학교, 통계학과

$$\lambda(t|\underline{\beta}) = \lambda(t_1|\beta_1) + \lambda(t_2|\beta_2) + \cdots + \lambda(t_J|\beta_J) \quad (1.1)$$

가 되는 비동질적 포아송 과정이 된다. 단, $\underline{\beta} = \{\beta_1, \beta_2, \dots, \beta_J\}$. 예를 들어, 동질적 포아송, Goel-Okumoto, Musa-Okumoto, Weibull 그리고 Cox-Lewis 과정에 대한 ROCOF들은 각각 β , $\alpha\beta \exp(-\beta t)$, $\alpha/(t + \beta)$, $\alpha\beta t^{\alpha-1}$, $\exp(\alpha + \beta t)$ 가 된다. 이런 과정을 가지고 깁스 추출법(Gibbs sampling)을 사용하여 단순한 ROCOF들의 합으로 중첩된 ROCOF를 가진 NHPP에 대한 베이저안 추론도 가능하다. D_t 을 t 시간까지 관찰된 데이터 시점이 n 개이면 (x_1, x_2, \dots, x_n) 에 순서를 주어 $0 < x_1 < x_2 < \cdots < x_n < t$ 을 만족하는 데이터 집합으로 만든다. 여기서, 순서화된 시점에 대한 점과정(point process)은 중첩 과정을 가지고 부분 시스템에 대한 고장의 원인은 알려져 있지 않고, $\beta_j, j = 1, \dots, J$ 에 대한 사전분포는 독립이라고 가정하자.

본 논문은 이러한 단조성의 ROCOF의 가정을 완화하기 위하여 단순한 ROCOF들을 가진 여러 개의 독립적인 NHPP들이 서로 겹쳐진 중첩(superpositon) 과정에 초점을 두고자 한다. 따라서 신뢰성이론에 많이 사용하는 Musa-Okumoto과정과 이것과 다른 강도함수의 패턴과 고장시스템을 가진 형상모수가 2인 Erlang(2)의 중첩과정을 제시하고자 한다. 결국 중첩 과정도 NHPP이기 때문에 이 이론에 적용할 수 있다. 다음 2절에서는 중첩과정에 대한 깁스추출법을 3절에서는 Musa-Okumoto와 Erlang(2)의 중첩과정에 대한 분석을 하고자 한다. 그리고 4절에는 베이저안 추론과 모형선택을 시행하였고 5절에는 수치적인 예를 제시하였다.

2. 중첩과정에 대한 깁스추출법

조건부 분포를 이용하여 깁스추출법을 적용할 수 있다. 이 깁스추출법은 MCMC (Markov Chain Monte Carlo)의 기법중 하나이다. 마코브 연쇄의 전이분포는 여러 개의 조건부 밀도로서 이루어지고 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 Tanner와 Wong(1987), Gelfand 와 Smith(1990), Casella 와 George(1992), 그리고 Kuo 와Yang(1995)등 많은 학자들에 의해서 제시되고 있다.

어떤 시스템이 시간 t 까지 가동된다고 가정하고 관찰된 n 번의 고장 시간을 x_1, x_2, \dots, x_n 이라고 표시하면 데이터 집합 D_t 는 $\{x_1, x_2, \dots, x_n; t\}$ 으로 구성된다. 만약 n 번째까지 관찰된 고장절단(failure truncated) 과정이 주어진다면 데이터 집합 D_{x_n} 는 $\{x_1, x_2, \dots, x_n\}$ 으로 된다. 그리고 시간절단(time truncated) 과정에 대한 데이터가 주어졌을때 NHPP의 우도함수는 다음 과 같다.

$$L(\underline{\beta}|D_t) = \left(\prod_{i=1}^n \lambda(x_i|\underline{\beta}) \right) \exp(-m(t|\underline{\beta})) \quad (2.1)$$

단, $m(t|\underline{\beta}) = \int_0^t \lambda(t)dt$ 이고, $0 < x_1 < x_2 < \cdots < x_n < t$, $\underline{\beta}$ 은 미지의 모수벡터를 의미한다. 이 절단 과정에 대한 우도함수는 Lawless(1982) 등이 이미 제시되었다. 고장절단 과정

은 t 대신에 x_n 을 대입하면 식 (2.1)과 유사한 표현이 된다. 중첩과정에서 $\underline{\beta}$ 의 사후밀도가 관찰하기 위해서 다음을 유도한다.

$$f(\underline{\beta}|D_t) \propto \left(\prod_{i=1}^n [\lambda_1(x_i|\beta_1) + \lambda_2(x_i|\beta_2) + \dots + \lambda_J(x_i|\beta_J)] \right) \cdot \exp(-[m_1(t|\beta_1) + \dots + m_J(t|\beta_J)]) \times \prod_{j=1}^J \pi_j(\beta_j) \quad (2.2)$$

단, $\underline{\beta}$ 는 모수벡터이고, π_j 는 β_j 에 대한 사전밀도이다. 식 (2.2)에 있는

$$\prod_{i=1}^n [\lambda_1(x_i|\beta_1) + \lambda_2(x_i|\beta_2) + \dots + \lambda_J(x_i|\beta_J)] \quad (2.3)$$

의 표현 때문에 D_t 의 조건하에서 $\underline{\beta}$ 의 조건부밀도에 대한 형태를 구체화하기 어렵다. 이러한 어려움을 해소하기 위하여 잠재변수(latent variables)를 사용한다(Berger와 Sun, 1993 참조). I 를 잠재변수들의 집합(collection)이라고 표시하고 이 I 를 사용하여 $\underline{\beta}$ 의 사후분포를 계산한다. 즉, 반복을 통해서 $f(I|\underline{\beta}, D_t)$ 로부터 I 를 추출하고 $f(\underline{\beta}|I, D_t)$ 로부터 $\underline{\beta}$ 를 추출하는 반복 형식을 통해 사후분포를 얻게 된다. 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 $I_{ij} = 1$ 이라고 하자. 그 외에는 $I_{ij} = 0$ 이라고 하고, $i = 1, \dots, n, \sum_{j=1}^J I_{ij} = 1$ 의 조건을 만족한다. $\underline{\beta}$ 와 D_t 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,

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

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

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

즉, I 와 D_t 의 조건하에서 $\underline{\beta}$ 의 사후밀도는 다음과 같다.

$$f(\underline{\beta}|I, D_t) \propto L(\underline{\beta}|D_t) \times \prod_{i=1}^n p(I_i|\underline{\beta}, D_t) \times \prod_{j=1}^J \pi_j(\beta_j) \quad (2.6)$$

$$\propto \prod_{j=1}^J \prod_{i: I_{ij}=1}^n \lambda_j(x_i) \times \prod_{j=1}^J \exp(-m_j(t)) \times \prod_{j=1}^J \pi_j(\beta_j)$$

조건부밀도는 I 와 D_t 의 조건하에서 β_j 의 독립 사후밀도들로 구성된 결합밀도가 된다. 그러므로 β_j 로부터 독립적으로 최종표본을 얻을 수 있다. 따라서 다음 절에 중첩과정에 대한 베이지안 추정기법을 제시한다.

3. MUSA-OKUMOTO와 ERLANG(2)의 중첩과정

Musa-Okumoto와 형상모수가 2인 Erlang(2)과정의 중첩과정을 제시하고자 한다. 즉, 각각의 강도함수는 다음과 같다.

$$\lambda_1(t) = \frac{\alpha_1}{\alpha_2 + t}, \quad \lambda_2(t) = \frac{\alpha_3^2 t}{\beta_3 t + 1} \quad (3.1)$$

단, $\alpha_i > 0$ ($i = 1, 2, 3$)을 만족한다. 식 (2.1)을 이용하여 중첩과정에 대한 우도함수는 다음과 같다.

$$L(\underline{\beta}|D_t) = \prod_{i=1}^n \left(\frac{\alpha_1}{\alpha_2 + x_i} + \frac{\alpha_3^2 x_i}{\alpha_3 x_i + 1} \right) \cdot \exp \left(-\alpha_1 \ln \left(1 + \frac{t}{\alpha_2} \right) - \alpha_3 t + \ln(1 + \alpha_3 t) \right) \quad (3.2)$$

단, $\Gamma(a, b)$ 는 평균이 a/b 인 감마 분포를 표시하고, 3개의 독립 요소에 대한 사전 분포는 다음과 같다.

$$\alpha_1 \sim \Gamma(a_2, b_2); \quad \alpha_2 \sim \pi_1(\alpha_2); \quad \alpha_3 \sim \pi_2(\alpha_3); \quad (3.3)$$

여기서 π_1 와 π_2 는 $\alpha_2 > 0$, $\alpha_3 > 0$ 을 만족하는 임의의 사전 분포이다. $\underline{\beta}$ 와 D_t 의 조건하에서의 I_i 의 조건부 밀도는 다항분포가 되는데 이 분포는 모수가 1이고 셀확률 (p_{i1}, p_{i2})을 가진다. 단, p_{ij} 는 $\lambda_j(x_i)/[\lambda_1 + \lambda_2(x_i)]$, $i = 1, \dots, n$, $j = 1, 2$ 이다. 이 중첩과정의 경우는 $\underline{\beta}$ 와 D_t 의 조건하에서의 I_i 는 $i = 1, \dots, n$ 에 대한 p_{i1} 의 모수를 가진 베르누이 분포로부터 발생된다. 단,

$$p_{i1} = \frac{\alpha_1}{\alpha_2 + x_i} / \left(\frac{\alpha_1}{\alpha_2 + x_i} + \frac{\alpha_3^2 x_i}{\alpha_3 x_i + 1} \right)$$

이고 $I_{i2} = 1 - I_{i1}$ 이 된다. 깃스추출법 알고리즘을 이용하기 위해서는 다음과 같은 3개의 조건부 분포가 필요하다.

- (1) $p(\alpha_1 | \alpha_2, \alpha_3, I, D_t) \sim \Gamma(a_2 + \sum I_{i1}, \ln(1 + t/\alpha_2) + b_2)$;
- (2) $p(\alpha_2 | \alpha_1, \alpha_3, I, D_t) \propto \frac{1}{\prod_{i: I_{i1}=1} (x_i + \alpha_2)} \times \exp[-\alpha_1 \ln(1 + \frac{t}{\alpha_2})] \cdot \pi_1(\alpha_2)$;
- (3) $p(\alpha_3 | \alpha_1, \alpha_2, I, D_t) \propto \left(\prod_{i: I_{i2}=1} \frac{\alpha_3^2 x_i}{\alpha_3 x_i + 1} \right) \cdot \exp[-\alpha_3 t + \ln(1 + \alpha_3 t)] \cdot \pi_2(\alpha_3)$;

여기서 α_2, α_3 는 메트로폴리스(Metropolis) 알고리즘(Chib and Greenberg, 1985)에 의해서 발생된다. 둘 이상의 중첩과정을 가진 깃스 알고리즘도 마찬가지로이다.

4. 베이지안 추론과 모형선택

$\underline{\beta}$ 와 같은 모수를 추정하기 위해서는 깃스추출법의 경험적 측도(empirical measure)를 사용할 수 있다. 예측은 고장 절단 상황에서 고려하는 것이 보다 쉽다. 즉, n 번째 고장까지 조사 한다. 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).

$$\begin{aligned} E(S(x)|D_{x_i}) &= E[E(P(X_{i+1} > x_i + x)|\underline{\beta}, D_{x_i})|D_{x_i}] \\ &= E[\exp(-m(x_i + x) + m(x_i))|D_{x_i}] \end{aligned} \tag{4.1}$$

식 (4.1)은 깃스추출법을 이용하거나 Rao-Blackwell정리를 사용하여 평가할 수 있다. 이것을 Gelman and Rubin(1992)이 제시한 MCMC방법을 적용하고자 한다. $\alpha^{(r,s)}$ 은 MCMC의 반복 알고리즘을 사용하여 r 번 반복 후 s 번 적용을 통해 α 를 발생시킨 깃스 표본추출을 의미한다. 유사한 방법으로 $\underline{\beta}^{(r,s)}$ 을 발생시킨다. 여기서, $s = S/2 + 1, \dots, S$, $r = 1, \dots, R$ 이며 충분히 큰 R 과 짝수인 S 를 대입한 깃스 표본추출을 사용하면 식 (4.1)의 추정치를 다음과 같이 유도할 수 있다.

$$\hat{S}(x|D_{x_n}) = \frac{2}{RS} \sum_{r=1}^R \sum_{s=\frac{S}{2}+1}^S \exp(-m(x_n + x|\underline{\beta}^{(r,s)}) + m(x_n|\underline{\beta}^{(r,s)})) \tag{4.2}$$

앞에서 제시한 Musa-Okumoto와 형상모수가 2인 Erlang(2)의 중첩과정은 다음과 같다.

$$\begin{aligned} \hat{S}(x|D_{x_n}) &= \frac{2}{RS} \sum_{r=1}^R \sum_{s=\frac{S}{2}+1}^S \exp[-\alpha_1^{(r,s)} \ln \left(1 + \frac{x_n + x}{\alpha_2^{(r,s)}} \right) - \alpha_3^{(r,s)}(x + x_n) \\ &\quad + \ln(1 + \alpha_3^{(r,s)}(x + x_n)) + \alpha_1^{(r,s)} \ln \left(1 + \frac{x_n}{\alpha_2^{(r,s)}} \right) + \alpha_3^{(r,s)} x_n - \ln(1 + \alpha_3^{(r,s)} x_n)]. \end{aligned}$$

미래 시점 x_{i+1} 에 대한 PCPO(Prequential Conditional Predictive Ordinate)는 (x_1, \dots, x_i) 가 주어진 미래의 관찰시점 x_{i+1} 에서 나타나는 확률변수 X_{i+1} 의 조건부밀도 $c_{i+1} = p(x_{i+1}|D_{x_i})$ 으로 정의할 수 있다. 이 PCPO는 과거의 데이터가 주어진 상태에서 X_{i+1} 의 값을 예측하기 때문에 과정선택을 하는데도 적절한 도구가 된다. 계열 $\{X_i\}_{i \geq 1}$ 이 주어지면 $c_{i+1} = p(x_{i+1}|x_i)$ 으로 표기할 수 있다. PCPO는 다음과 같은 식을 통해 계산된다.

$$\begin{aligned} p(X_{i+1}|D_{x_i}) &= \int \cdots \int p(X_{i+1}|\underline{\beta}, D_{x_i})p(\underline{\beta}|D_{x_i})d\underline{\beta} \\ &= \int \cdots \int \lambda(X_{i+1}) \exp(-m(X_{i+1}) + m(x_i)) \times p(\underline{\beta}|D_{x_i})d\underline{\beta} \end{aligned} \tag{4.3}$$

앞에서 제시한 Musa-Okumoto와 형상모수가 2인 Erlang(2)과정의 중첩과정을 고려하면 다음과 같다.

$$p(x_{i+1}|x_i) = \frac{2}{RS} \sum_{r=1}^R \sum_{s=\frac{r}{2}+1}^S \left[\left(\frac{\alpha_1^{(r,s)}}{\alpha_2^{(r,s)} + x_{i+1}} + \frac{\alpha_3^{(r,s)} x_{i+1}}{\alpha_3^{(r,s)} + x_{i+1} + 1} \right) \cdot \exp\left(-\alpha_1^{(r,s)} \ln\left(1 + \frac{x_{i+1}}{\alpha_2^{(r,s)}}\right) - \alpha_3^{(r,s)} x_{i+1} + \ln\left(1 + \alpha_3^{(r,s)} x_{i+1}\right)\right) + \alpha_1^{(r,s)} \ln\left(1 + \frac{x_i}{\alpha_2^{(r,s)}}\right) \alpha_3^{(r,s)} x_i - \ln\left(1 + \alpha_3^{(r,s)} x_i\right) \right]$$

NHPP는 여러 과정들이 존재한다. 최적 과정선택은 베이저안 PCPO를 사용하여 해결할 수 있다. x_{i+1} 에 대한 PCPO는 $c_{i+1} = p(x_{i+1}|x_1, \dots, x_i)$, $i \geq 1$ 에 의해 정의된다. 데이터에 대한 사전분포를 사용하여 각 과정에서 i 에 대한 c_i 를 플롯시킨다. 단, $i = 2, 3, \dots, n$ 이다. 평균적으로 가장 큰 c 를 가진 과정이 더 좋은 과정이 된다. 그러나, 플롯에서 c 의 값이 드물게 작을 때는 이상치가 존재할 수도 있다. 베이저안 예측 우도기준은 예측 밀도

$$C(l) = \prod_{i=1}^n c_i(l) = p(x_1, x_2, \dots, x_n) \quad (4.4)$$

를 최대화시키는 과정이 선택되어진다. 단, l 은 과정들에 대한 지수를 의미한다. 이러한 계산은 식 (4.3)으로 부터 직접 이루어진다. PCPO는 과거의 고장 시점 x_1, x_2, \dots, x_i 의 집합에 조건화를 시킴으로써 x_{i+1} 의 CPO(conditional predictive ordinates)를 정의한다. 이러한 Prequential 접근은 각 고장 시점에서 그 다음 고장 시점을 예측하는 것이 가능하기 때문에 CPO는 현재의 데이터를 알고 있을 때 좋은 과정에 대한 선택을 할 수 있게 한다.

5. 수치적인 예

과정선택과 베이저안 추론에서 다음과 같은 강도함수를 가진 NHPP를 가진 세 과정을 제시하고자 한다. 즉, 부분함수 $\lambda_1(t) = \alpha_1/(\alpha_2 + t)$, $\lambda_2(t) = \alpha_3^2 t/(\alpha_3 t + 1)$, 그리고 이들의 합 $\lambda_1(t) + \lambda_2(t)$. 단, $\alpha_i > 0$, ($i = 1, 2, 3$), λ_1 은 Musa-Okumoto과정의 강도함수이고, λ_2 는 형상모수가 2인 Erlang(2)의 강도함수를 의미한다. 데이터들은 $\alpha_1 = 15$, $\alpha_2 = 200$, $\alpha_3 = 0.05$ 를 가정하여 Lewis와 Shedler(1982)에 의해 제시된 IMSL RANPP ($t = 360$)이라고 불리는 루틴에 의하여 시뮬레이트 되었다. 이 루틴에 의한 자료들은 다음과 같다.

5.08, 10.16, 15.26, 20.34, 25.46, 30.54, 35.65, 40.73, 45.85, 50.96, 56.03, 61.12, 66.21, 71.29, 76.37, 81.44, 86.52, 91.59, 96.67, 101.75, 106.83, 111.90, 117.02, 122.15, 127.21, 132.28, 137.35, 142.46, 147.56, 152.68, 157.75, 162.83, 167.95, 173.02, 178.07, 183.16, 188.22, 193.13, 198.22, 203.30, 208.41, 213.49, 218.60, 223.69, 228.79, 233.90, 238.96, 244.04, 249.13, 254.23, 259.31, 264.38, 269.49, 274.57, 279.69, 284.79, 289.87, 294.96, 300.04, 305.15, 310.23, 315.32, 320.38, 325.50, 330.57, 335.69, 340.81, 345.91, 351.00, 356.08.

상대적으로 사전분포는 확산분포(diffuse priors)를 제시하여 경험적 결과를 유도하였다. 사전분포는 $\alpha_1 = \Gamma(1, 10^{-4})$ 을 선택 이용하였고, α_2, α_3 에 대한 비정보사전밀도

(noninformative prior density)는 각각 $\pi_1(\alpha_2) = 1/\alpha_2, \alpha_2 > 0, \pi_2(\alpha_3) = 1/\alpha_3, \alpha_3 > 0$ 을 가정하였다. 이 세 과정에 대한 확산 사전분포는 표 5.1에 요약되었다. 반복이 얼마나 필요할지를 결정하기 위하여 Gelman 과 Rubin(1992)이 제시한 방법을 사용하였다. German 과 Rnbin은 X_{i+1} 에 대한 95 % 예측 신뢰구간은 $[\{x_{i+1}^{(i,s)}\}_{s=1}^S]_{i=I/2+1}^I$ 에서 발생된 표본에서 하한과 상한은 각각 2.5% 와 97.5%의 분위수(quantiles)로 설정할 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 관측자료 x_i 가 95%의 신뢰구간 안에 포함되는가에 의해서 판단 할 수 있다. 본 논문에서는 수렴성을 고려한 깃스시행은 1000번 반복 중 10개의 분리된 깃스체인(separate Gibbs chains)을 발생하였다 따라서 각 모수에 대하여 깃스체인 10을 가지는 $110^{th}, 120^{th}, \dots, 990^{th}, 1000^{th}$ 반복을 고려하였다. 즉, 크기는 900이 된다. 3장에서 메트로폴리스 알고리즘을 사용하여 α_2 을 추정하는 방법은 다음과 같다.

$$p(\alpha_2 | \alpha_1, \alpha_3, I, D_t) \propto \frac{1}{\prod_{i: I_{i1}=1}^n (x_i + \alpha_2)} \times \exp[-\alpha_1 \ln(1 + \frac{t}{\alpha_2})] \cdot \pi_1(\alpha_2)$$

오른쪽 식 즉, 목적분포를 간편하게 $f(\alpha_2)$ 로 표기하고, 여기서 정규화상수(normalizing constant)는 필요하지 않는다. α_1, α_2 는 앞 단계 즉, 깃스추출법으로 추정된 값이고 $\pi_1(\alpha_2)$ 의 분포는 $\alpha_2 > 0$ 을 만족하는 $1/\alpha_2$ 이라고 가정되었고 추이커널(transitional kernel)은 표 5.1에 가정된 거의 대칭을 이루는 $\Gamma(1, 10^{-4})$ 에서 α' 을 발생하고 균등분포 $(0, 1]$ 에서 확률변량을 w 라고 하면 $\log w \leq \log f(\alpha') - \log f(\alpha_2)$ 을 만족하면 α' 을 α_2 으로 간주되고 만족하지 않으면 α_2 는 α' 으로 대체되면서 계속 반복된다. 표 5.2는 모수에 대한 사후평균의 추정 값을 요약하였고, 중첩과정에서 $\{\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3\}$ 의 베이지안 추정은 여러 과정들과 비교되었다. 그림 5.1은 시간의 흐름에 따라 비증가함수(nonincreasing function)의 패턴 을 가진 미래 생존함수 $\hat{S}(x|D_{x_n})$ 을 나타낸 것이고, 표 5.2는 이 세 과정에 대한 PCPO의 값 $\ln C(l)$ 을 요약하였다. 그 결과 예측된 것처럼 단순과정보다 중첩과정의 PCPO값이 제일 크므로 중첩과정이 더 좋은 과정으로 간주할 수 있다. 그림 5.2는 각 시점에 대한 PCPO를 나타내었다. 그림 5.3에서는 1000번 반복 중 10개의 분리된 깃스체인(separate Gibbs chains)을 발생시켜 얻은 베이즈 강도함수와 시뮬레이트된 강도함수의 비교결과 거의 밀접한 추세를 보이고 있으므로 수렴하고 있다고 간주할 수 있다.

표 5.1: 사전분포

강도함수	Musa-Okumoto과정	Erlang(2) 과정	중첩과정
$\lambda(t)$	$\frac{\alpha_1}{\alpha_2+t}$	$\frac{\alpha_1^2 t}{\alpha_3 t+1}$	$\frac{\alpha_1}{\alpha_2+t} + \frac{\alpha_1^2 t}{\alpha_3 t+1}$
사전분포	$\alpha_1 = \Gamma(1, 10^{-4})$ $\alpha_2 = \Gamma(1, 10^{-1})$	$\alpha_3 = \Gamma(1, 10^{-4})$	$\alpha_1 = \Gamma(1, 10^{-4})$ $\alpha_2 = \Gamma(1, 10^{-1})$ $\alpha_3 = \Gamma(1, 10^{-4})$

표 5.2: 사전평균의 추정값과 PCPO의 값

강도함수	Musa-Okumoto과정	Erlang(2) 과정	중첩과정
$\lambda(t)$	$\frac{\alpha_1}{\alpha_2+t}$	$\frac{\alpha_3^2 t}{\alpha_3 t + 1}$	$\frac{\alpha_1}{\alpha_2+t} + \frac{\alpha_3^2 t}{\alpha_3 t + 1}$
사후평균의 추정값	$\hat{\alpha}_1 = 11.3992$ $\hat{\alpha}_2 = 200.1673$	$\hat{\alpha}_3 = 0.05412$	$\hat{\alpha}_1 = 13.8934$ $\hat{\alpha}_2 = 190.5672$ $\hat{\alpha}_3 = 0.04652$
$\ln C(l)$	-66.4567	-65.7852	-63.2391

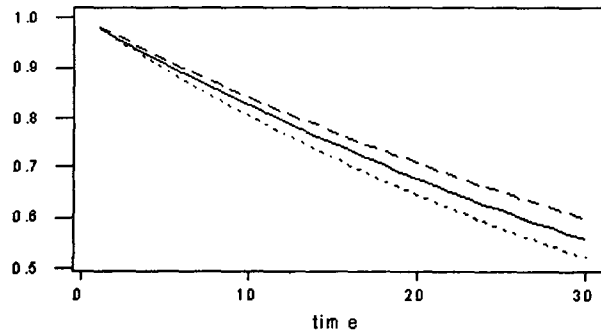


그림 5.1: $\hat{S}(t|D_n)$

(a) Musa-Okumoto과정 (...), (b) Erlang과정 (-), (c) 중첩과정(- -)

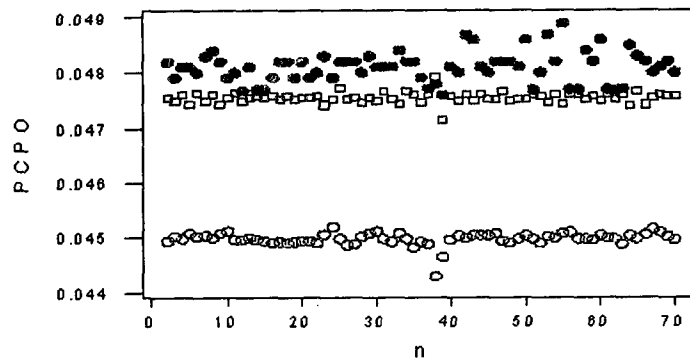


그림 5.2: PCPO

(a) Musa-Okumoto과정(○), (b) Erlang과정(□), (c) 중첩과정(●)

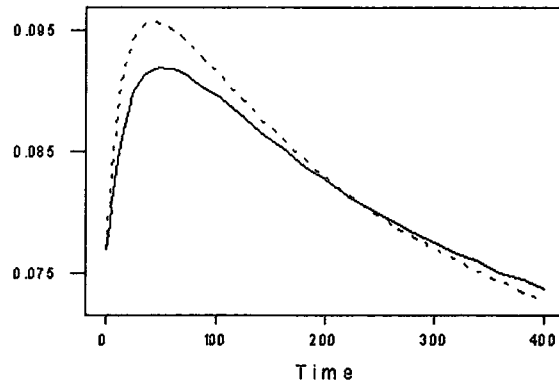


그림 5.3: 시뮬레이트한 강도함수 $\lambda(t) = 15/(t + 200) + (0.05^2 \cdot t)/(0.05 \cdot t + 1)$ 와 베이스추정치를 이용한 중첩모형의 강도함수(--)와의 비교

참고문헌

- [1] Blumenthal, S. and Marcus, R.(1975). Estimating Population Size with Exponential Failure, *Journal of the American Statistical Association*, vol. 70, 913-922.
- [2] Berger, J. O. and Sun, D.(1993). Bayesian Analysis For The Poly-Weibull Distribution, *Journal of the American Statistical Association*, vol. 88, 1412-1418.
- [3] 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.
- [4] Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler, *The American Statistician*, vol.46, 167-174.
- [5] Cinlar, E.(1975). *Introduction To Stochastic Process*, New Jersey : Prentice-Hall.
- [6] Chib, S and Greenberg, E.(1995). Understanding the Metropolis-Hastings Algorithm, *The American Statistician*, vol.49, 327-335.

- [7] Dawid, A. P.(1984). Statistical Theory : The Prequential Approach, *Journal of the Royal Statistical Society, Ser. A*, vol.147, 278-292.
- [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., Dey, D. K. and Chang, H.(1992). *Model Determination Using Predictive Distributions with Implementation via Sampling-Based Methods (with discussion)*, in Bayesian Statistics 4, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Cambridge, MA ; Oxford University Press, pp.147-169.
- [10] 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.
- [11] Gelman, A. E. and Rubin D.(1992). Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, vol.7, 457-472.
- [12] Kuo, L. and Yang, T. Y.(1995). Bayesian Computation of Software Reliability, *Journal of Computational and Graphical Statistics*, 65-82. 781-790.
- [13] Kuo, L. and Yang, T. (1996). Bayesian computation of Software Reliability, *Journal of the American Statistical Association*, Vol.91, pp.763-773.
- [14] Lawless, J. F.(1982). *Statistical Models and Methods for lifetime Data*, New York: John Wiley and Sons.
- [15] Musa, J. D. , Iannino, A., and Okumoto, K.(1987). *Software Reliability: Measurement, Prediction, Application*, New York: McGraw Hill.
- [16] 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.
- [17] *USER'S MANUAL STAT/LIBRARY FORTRAN Subroutines for statistical analysis*, IMSL, Volume 3, 1050-1054.
- [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.

[1998년 2월 접수, 1998년 7월 최종수정]

Bayesian Computation for Superposition of MUSA-OKUMOTO and ERLANG(2) processes

Kiheon Choi¹⁾ Heecheul Kim²⁾

ABSTRACT

A Markov Chain Monte Carlo method with data augmentation is developed to compute the features of the posterior distribution. For each observed failure epoch, we introduced latent variables that indicates with component of the Superposition model. This data augmentation approach facilitates specification of the transitional measure in the Markov Chain. Metropolis algorithms along with Gibbs steps are proposed to preform the Bayesian inference of such models. For model determination, we explored the Prequential Conditional Predictive Ordinate(PCPO) criterion that selects the best model with the largest posterior likelihood among models using all possible subsets of the component intensity functions. To relax the monotonic intensity function assumptions, we consider in this paper Superposition of Musa-Okumoto and Erlang(2) models. A numerical example with simulated data set is given.

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

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