

## MCMC를 이용한 비동질적 포아송과정에서의 일반화 순서통계량 모형의 연구

최 기 현<sup>1)</sup>, 김 희 철<sup>2)</sup>

### 요 약

컴퓨터의 발전에 따른 MCMC를 비동질적 포아송 과정에 이용하였다. 베이저안 추론에서 조건부 분포를 가지고 사후분포를 결정하는데 있어서의 계산 문제를 고려하였다. 특히 분포가 이중지수, 고펜르츠, 랄리, 감마, 그리고 겹벨인 일반 순서통계량 모형에 대하여 깁스 샘플링과 메트로폴리스 알고리즘을 활용한 베이저안 계산과 모형선택을 제시하였다.

### 1. 서론

소프트웨어 신뢰도에서 발견된 고장의 수를 모형화 하는 것은 비동질적 포아송 과정(nonhomogenous Poisson process, *NHPP*)으로 널리 사용하여 왔다.  $M(t)$ 를  $(0, t]$ 사이에서의 *NHPP*의 발견된 고장의 수라 하자. 또한  $m(t) = EM(t)$ ,  $t$ 에 대한 비감소 함수(nondecreasing function)라고 가정하자.  $m(t)$ 가 미분가능하면, 강도함수(intensity function)는  $\lambda(t) = m'(t)$ 이고 *NHPP*으로 표현할 수 있다. 강도함수를 고장 발생률(rate of occurrence of failure, *ROCOF*)이라고 한다. 예를 들어 동질적 포아송 과정(homogeneous Poisson process), Musa-Okumoto(1984) 과정, Weibull 과정 그리고 Cox-Lewis 과정은 각각 *NHPP*의 강도함수가 상수, 시간에 관한 부분함수, 멱함수, 대수선형함수이다.

시스템에서 고장 시간은 2가지로 나누어 생각할 수 있다. 하나는 일반 순서통계량 모형(general order statistics model, *GOS*)이고 다른 하나는 기록값 통계량 모형(record value statistics model, *RVS*)이다.

*GOS* 모형은 소프트웨어 테스트 시점에서 미지의  $N$ 개의 결함 수가 있다고 가정하자. 관찰된 고장 시점은  $N$ 으로부터 추출된 처음  $n$ 개의 순서통계량은  $R^+$ 에서 만들어진(supported) 밀도함수  $f$ 을 가진 독립이고 동일한 분포(*i.i.d.*)의 관찰값으로 모형화 한다.  $f$ 가 지수분포라면 Jelinski-Moranda 모형이 된다. 그러면 이  $f$ 을 다양하게 할 수 있다. 예를 들면 랄리, 파레토, 와이블, 지수 그리고 극값(truncated extreme value)분포를 따른다고 생각할 수 있다. 이러한 모형을 랄리, 파레토, 와이블, 지수 그리고 극값 순서통계량 모형이라고 한다.

1) (132-714) 서울시 도봉구 쌍문동 419, 덕성여자대학교 통계학과 부교수  
2) (100-715) 서울시 중구 필동 3가 26, 동국대학교 통계학과

GOS 모형은 각 수리 시점에서는 새로운 결함이 없다는 한계를 가지고 있다. 그러나 수리하는 동안에도 새로운 결함이 발생한다는 가정을 반영하기 위한 모형이 RVS 모형이고, 고장 시점은  $f$ 로부터 추출된 독립이고 동일한 분포의 관찰값에 대한 기록 멈춤 통계량(record breaking statistic)이 된다. RVS 모형에 대한 설명은 Choi & Kim(1997)을 참조하기 바란다.

GOS 모형에서  $N$ 은 평균  $\theta$ 을 가진 포아송 분포를 한다는 가정이 필요하다. 즉  $M(t)$ 는  $m(t) = \theta F(t)$ 를 가진 NHPP가 된다. 단,  $F$ 는 밀도함수  $f$ 를 따르는 누적분포함수이다. 이때  $\lim_{t \rightarrow \infty} m(t)$ 가 유한할 때의 과정을 NHPP(1)이라 하자. 특히 Jelinski-Moranda 모형에서  $F(t) = (1 - e^{-\beta t})$ 일 때  $M(t)$ 는 평균함수  $m(t) = \theta (1 - e^{-\beta t})$ 를 가진 NHPP가 된다. 이것을 Goel-Okumoto 과정(1979)이라고 한다. Jelinski-Moranda 모형과 Goel-Okumoto 과정의 관계는 Langberg & Singpurwalla(1985)에 의해서 재조명되었다. 따라서 본 논문에서는 Goel-Okumoto 과정과 연관하여 다루지 않은 NHPP 중에서  $F$ 가 이증지수, 고펜페르츠(Gompertz), 랄리, 감마, 그리고 검벨(Gumbel)일 때 이증지수 NHPP(1), 고펜페르츠 NHPP(1), 랄리 NHPP(1), 감마 NHPP(1), 검벨 NHPP(1)을 제시하였다.

NHPP에 대한 베이지 추정치들을 계산하기 위하여, 사후분포의 형태를 구하는데 깃스 알고리즘이 제안되어진다. 깃스 알고리즘은 Markov Chain Monte Carlo(MCMC) 알고리즘인데, 이것은 바람직한 사후분포로서 정상성 분포를 가지는 마코프 체인에 따라 표본들이 변화하는 알고리즘이다. 이 마코프 체인의 추이 측정은 주로 보통 조건부 밀도함수의 곱이다. NHPP(1) 모형에 대하여 하나는 두 개의 잠재변수들을 이용하는 데이터 증대(data augmentation) 기법을 도입하는 것이다. 즉, 한 변수는 소프트웨어에 남아있는 수많은 오류들이며, 다른 변수는 포아송 확률들 중에서 척도모수의 변화에 따라서 어떤 요소가 도움을 주는지를 나타내는 것이다. 또 하나는 깃스 알고리즘에서 메트로폴리스 방법인데 이것의 척도모수는 메트로폴리스 알고리즘에 의해 발생되어진다.

2절에서는 NHPP(1)에 대한 우도함수와 일반 순서통계량에 대해 논하고, 일반 순서통계량 모형들에 대한 사후분포를 구하는 깃스 샘플링이 3절에 제시된다. 4절에서는 베이저안 추론과 모형 선택에 대해 논한다. 마지막으로 수치 예를 5절에서 보여준다.

## 2. NHPP(1)에 대한 우도함수와 일반 순서통계량

$m(t)$ 는 미지의 모수  $\theta$ ,  $\beta$ 에 의하여 결정된다고 가정하고,  $\theta$ 는 음이 아닌(non-negative) 실수이고  $\beta$ 는 미지의 모수벡터도 될 수 있다. 간단히  $m(t) = m(t|\theta, \beta)$ 라고 하자.

시간  $t$ 까지 테스트하기 위한 시간절단모형이 주어지면,  $n$ 번의 고장 시점의 순서화된 자료를  $x_1, x_2, \dots, x_n$ 이라고 하자. 그러므로 데이터 집합  $D_t$ 는 다음과 같이 구성된다.  $\{n, x_1, x_2, \dots, x_n; t\}$ ,  $n$ 번째까지 고장 시점이 관찰된 고장절단모형일 경우 데이터 집합  $D_{x_n}$ 은  $\{x_1, x_2, \dots, x_n\}$ 으로 구성된다. 여기서 우리는 NHPP 모형에 대한 주어진 자료로부터 우도함수를 얻는 방법을 보인다. 주어진 시간절단모형에서 구간  $(0, t]$ 에서 관측된 고장의 수는 확률변수로서  $n$ 이다. 그러면 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 순서 통계량과 같다(Parzen 1962, p.143). 따라서 주어진 자료에 대한 시간절단모형에서의 우도 함수는 다음과 같다.

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

고장절단모형은 (2.1)식의  $t$ 을  $x_n$ 으로 대치하면 유사한 형태가 된다.

$m(t)$ 의 극한반응(limiting behavior)에 따라 다음과 같이 NHPP 과정을 2가지 형태로 구분할 수 있다.

- (1)  $\lim_{t \rightarrow \infty} m(t) < \infty$  일 때 NHPP(1).
- (2)  $t \rightarrow \infty$  이고  $m(t) \rightarrow \infty$  일 때 NHPP(2).

본 연구에서는 NHPP(1)의 경우만 논하고자 한다.

NHPP(1)에서  $\lim_{t \rightarrow \infty} m(t) = \theta < \infty$  이고,  $N$ 은 테스트링 시점에서의 결함의 수라고 가정하자.

NHPP(1)에서 시간절단 모형에 대한 데이터  $D_t$ 가 주어졌을 때  $\theta, \beta, N$ 에 대한 우도함수는 다음과 같다.

$$L_{NHPP(1)}(\theta, \beta, N | D_t) = \frac{N!}{(N-n)!} \left[ \prod_{i=1}^n \frac{\lambda(x_i)}{\theta} \right] \left( 1 - \frac{m(t)}{\theta} \right)^{N-n} \quad (2.2)$$

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

$$\sum_{N=n}^{\infty} L_{NHPP(1)}(\theta, \beta, N | D_t) \times p(N | \theta) = L_{NHPP}(\theta, \beta | D_t) \quad (2.3)$$

그러므로 (2.1)식의 우도함수는  $N$ 이 포아송 분포를 따른다는 가정을 이용하면 주변 우도함수가 된다. 고장 절단 모형에서는  $t$ 을  $x_n$ 으로 대치하면 (2.2)식과 유사한 형태가 된다.

### 3. NHPP(1)과 일반 순서통계량 모형에 대한 깃스 샘플링

NHPP(1)에 대해서는 미지의 남아 있는 결함의 수를 사용한 데이터 증대를 이용하여 깃스 알고리즘을 적용하는데 필요한 조건부 분포를 구체화 할 수 있다.

(2.1)식의 우도함수에 있는  $\exp[-\theta F(t | \beta)]$ 의 형태 때문에 깃스 샘플링에 있어서 필요로 하는 주어진  $\theta$ 와  $D_t$ 에  $\beta$ 대한 조건부 밀도함수 형태를 구체화하기 어렵다. 그러므로 모수  $\theta[1 - F(t | \beta)]$ 을 가지는 포아송 분포를 따르는 잠재변수  $N' = N - n$ 을 이용한다. 그러면 사후분포  $p(\theta, \beta | D_t)$ 는 결합밀도  $p(\theta, \beta, N' | D_t)$ 을 주변밀도화함으로서 얻을 수 있다. 이 밀도는 다음과 같은 조건부 밀도의 반복을 이용하여 깃스 샘플링을 통하여 근사밀도함수를 구할 수 있다. 즉  $p(N' | \theta, \beta, D_t)$ ,  $p(\theta | N', \beta, D_t)$ ,  $p(\beta | N', \theta, D_t)$ , 단, 이 밀도들은 구체화가 가능한 밀도라야 한다. 다음의 곱을 고려해 보자.

$$\begin{aligned}
L_{NHPP(1)} \times p(N' | \theta) &= [\prod_{i=1}^n \theta f(x_i | \beta)] e^{-\theta F(t|\beta)} e^{-\theta(1-F(t|\beta))} [\theta(1-F(t|\beta))]^{N'} / N'! \\
&= \frac{N!}{(N-n)!} [\prod_{i=1}^n f(x_i | \beta)] [1-F(x_n | \beta)]^{N'} \times e^{-\theta} \theta^N / N!. \quad (3.1)
\end{aligned}$$

이 식은 GOS 모형의 우도함수와  $\theta$ 에 대한 포아송 사전분포의 곱으로 이루어진다. 여기서  $X \perp Y$ 는  $X$ 와  $Y$ 가 독립을 의미하고  $N' = N - n$ 을 의미하고  $\Gamma(a, b)$ 는 평균이  $a/b$ 인 감마분포를 나타낸다. 그리고  $P(\theta)$ 는 평균이  $\theta$ 인 포아송 분포를 의미하고, 벡터  $\beta$ 은  $\beta$  혹은  $(\alpha, \beta)$ 을 나타낸다.

### 3.1 이중지수 일반 순서통계량 모형

(2.1)식의 우도함수를 이용하여 다음과 같이 나타낼 수 있다.

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

사전분포 :  $\theta \sim \Gamma(a, b)$ ;  $\beta \sim \Gamma(c, d)$ ;  $\theta \perp \beta$ .

깁스 샘플링 알고리즘 :

$$N' | \theta, \beta, D_t \sim P \left( \frac{\theta}{2} e^{-\beta t} \right); \quad (3.2)$$

$$\theta | N', \beta, D_t \sim \Gamma(a + n + N', b + 1); \quad (3.3)$$

$$\beta | N', \theta, D_t \sim \Gamma \left( c + n, d + tN' + \sum_{i=1}^n x_i \right); \quad (3.4)$$

### 3.2 감마 일반 순서통계량 모형

형상모수가 2인 경우를 고려하여 보면 우도함수는 다음과 같이 표현되고 알고리즘은 이중지수 모형과 유사한 방법으로 (3.5)식을 유도할 수 있다.

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

사전분포 :  $\theta \sim \Gamma(a, b)$ ;  $\beta \sim \pi_1(\beta)$

단,  $\pi_1$ 은  $\beta > 0$ 을 만족하는 임의의 사전분포이고  $\theta \perp \beta$ 을 만족한다.

깁스 샘플링 알고리즘 :

$$\begin{aligned}
N' | \theta, \beta, D_t &\sim P \{ \theta e^{-\beta t} (1 + \beta t) \}; \\
\theta | N', \beta, D_t &\sim \Gamma(a + n + N', b + 1); \quad (3.5)
\end{aligned}$$

$$\beta | N', \theta, D_t \propto \beta^{2n} e^{-\beta \sum x_i} [e^{-\beta t} (1 + \beta t)]^{N'} \times \pi_1(\beta);$$

단,  $\beta$ 는 메트로폴리스 알고리즘에 의하여 발생되어진다.

### 3.3 랠리 일반 순서통계량 모형

랄리 분포는 형상모수가 2인 와이블 분포의 특수한 경우이므로 우도함수는 다음과 같이 표현되고 유사한 방법으로 (3.6)식을 유도할 수 있다.

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

사전 분포 :  $\theta \sim \Gamma(a, b)$ ;  $\beta \sim \Gamma(c, d)$ ;  $\theta \perp \beta$ .

깁스 샘플링 알고리즘 :

$$\begin{aligned} N' | \theta, \beta, D_t &\sim P(\theta e^{-\beta t^2}); \\ \theta | N', \beta, D_t &\sim \Gamma(a + n + N', b + 1); \\ \beta | N', \theta, D_t &\sim \Gamma\left(n + c, N' t^2 + d + \sum_{i=1}^n x_i^2\right); \end{aligned} \quad (3.6)$$

### 3.4. 콤포르츠 일반 순서통계량 모형

우도함수는 다음과 같이 표현된다. 유사한 방법으로 (3.7)식을 유도할 수 있다

$$\begin{aligned} L_{NHPP(1)}(\theta, \underline{\beta} | D_t) &= \left[ \prod_{i=1}^n \theta \cdot \exp\left\{(\beta + \alpha x_i) - \frac{1}{\alpha}(e^{\beta + \alpha x_i} - e^\beta)\right\} \right] \\ &\quad \cdot \exp\left[-\theta \left(1 - \exp\left\{\frac{-e^{\beta + \alpha t} + e^\beta}{\alpha}\right\}\right)\right] \end{aligned}$$

사전분포 :  $\theta \sim \Gamma(a, b)$ ;  $\beta \sim \pi_2(\beta)$ ;  $\alpha \sim \pi_3(\alpha)$ ;

단,  $\pi_2(\beta), \pi_3(\alpha)$ 는  $\alpha > 0, \beta > 0$ 을 만족하는 임의의 사전 분포이고

$\theta \perp \beta \perp \alpha$ 을 만족한다.

깁스 샘플링 알고리즘 :

$$\begin{aligned} N' | \theta, \beta, \alpha, D_t &\sim P\left\{\theta \exp\left[\frac{-e^{\beta + \alpha t} + e^\beta}{\alpha}\right]\right\}; \\ \theta | N', \beta, \alpha, D_t &\sim \Gamma(a + n + N', b + 1); \\ \beta | N', \theta, \alpha, D_t &\propto \\ &\quad \exp\left\{n\beta + n\frac{1}{\alpha}e^\beta - \frac{1}{\alpha}\sum_{i=1}^n e^{\beta + \alpha x_i} - \frac{N'}{\alpha}(e^{\beta + \alpha t} - e^\beta)\right\} \pi_2(\beta); \\ \alpha | N', \theta, \beta, D_t &\propto \\ &\quad \exp\left\{\alpha \sum x_i + n\frac{1}{\alpha}e^\beta - \frac{1}{\alpha}\sum_{i=1}^n e^{\beta + \alpha x_i} - \frac{N'}{\alpha}(e^{\beta + \alpha t} - e^\beta)\right\} \pi_3(\alpha); \end{aligned} \quad (3.7)$$

단,  $\alpha, \beta$ 는 메트로폴리스 알고리즘에 의하여 발생되어진다.

### 3.5. 겐벨 일반화 순서통계량 모형

우도함수는 다음과 같이 표현되고 유사한 방법으로 (3.8)식을 유도할 수 있다.

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

사전분포 :  $\theta \sim \Gamma(a, b)$ ;  $\beta \sim \pi_4(\beta)$ ;

단,  $\pi_4$ 는  $\beta > 0$ 을 만족하는 임의의 사전 분포이고  $\theta \perp \beta$ .

깁스 샘플링 알고리즘 :

$$N' | \theta, \beta, D_t \sim P\{\theta \exp(e^{-t^\beta})\};$$

$$\begin{aligned} \theta | N', \beta, D_t &\sim \Gamma(a + n + N', b + 1); \\ \beta | N', \theta, D_t &\propto \beta^n \cdot \exp\left\{\beta \sum_{i=1}^n x_i - \sum_{i=1}^n e^{-\beta x_i} - N' \beta t\right\} \times \pi_4(\beta); \end{aligned} \quad (3.8)$$

단,  $\beta$ 는 Metropolis 알고리즘에 의하여 발생되어진다.

#### 4. NHPP(1)에 대한 베이지안 추론과 모형 선택

미지의 모수를 추론하기 위해서는 조건부 예측순서량(conditional predictive ordinates)과 미래 신뢰성 함수를 이용하여 고장 시점 사이의 평균 시간을 추정해야 한다.  $N, \theta$  그리고  $\beta$  같은 모수를 추정하기 위해서는 깃스 추출법의 경험적 측도(empirical measure)을 사용할 수 있다. 예측은 고장 절단 상황에서 고려하는 것이 보다 쉽다. 즉  $n$ 번째 고장까지 조사한다. 모든 사후 분포는 앞절에서와 같이  $t$ 을  $x_n$ 으로 대체함으로써 유도할 수 있다. 완전 데이터  $D_{x_n}$  대신에 베이지안 추론을 각  $D_{x_i} = (x_1, x_2, \dots, x_i)$ ,  $i = 1, \dots, n$ 에 적용할 수 있는 축차적 구성을 고려하자.

$x_i$ 로부터 거리  $x$ 만큼 떨어진 시점에서의 미래 생존함수의 추론은 다음과 같은 식을 통해 얻을 수 있다(이 식의 유도 과정은 Cinlar 1975, p.97 참조).

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

미래시점  $x_{i+1}$ 에 대한 PCPO(Prequential Conditional Predictive Ordinate)는  $(x_1, \dots, x_i)$ 가 주어진 미래의 관찰 시점  $x_{i+1}$ 에서 나타나는 조건부 밀도  $c_{i+1} = p(x_{i+1} | D_{x_i})$ 으로 정의할 수 있다. 이 PCPO는 과거의 데이터가 주어진 상태에서  $X_{i+1}$ 의 값을 예측하기 때문에 모형 선택을 하는데도 적절한 도구가 된다.

NHPP를 변환하여 동질적인 포아송 과정으로 되기 위해서는  $m(x_1) < m(x_2) < \dots$  들은 비율 1를 가진 동질적 포아송 과정에서의 도착시간들이 되어야 한다. 동질적 포아송과정에서의 도착시간 간격들은 비율 1이고 독립인 지수확률변량이기 때문에 사건  $[m(X_n) \geq m(X_{n-1} + t)]$ 와 사건  $[X_n - X_{n-1} > t]$ 은 동등하다. 그러므로 다음과 같은 식이 성립한다.

$$\begin{aligned} p(X_n - X_{n-1} > t | X_1, \dots, X_{n-1}) &= p(m(X_n) \geq m(t + X_{n-1}) | X_1, \dots, X_{n-1}) \\ &= \exp[-m(X_{n-1} + t) + m(X_{n-1})] = p(X_n - X_{n-1} > t | X_{n-1}) \end{aligned}$$

이 사실은 수열  $\{X_i\}_{i \geq 1}$ 이 주어지면 하나의 마코브 연쇄가 된다.  $c_{i+1} = p(x_{i+1} | x_i)$ 으로 표기하고 PCPO는 다음과 같은 식을 통해 계산된다.

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

(4.1)식과 (4.2)식은 깃스 추출법을 이용하거나 Rao-Blackwell 정리를 사용하여 평가할 수 있다. 이것을 Gelman & Rubin (1992)이 제시한 MC방법을 적용하고자 한다.  $\alpha^{(r,s)}$ 은 MCMC의 반복 알고리즘을 사용하여  $r$ 번 반복후  $s$ 번 적용을 통해  $\alpha$ 를 발생시킨 깃스 추출법을 의미한다. 유사한 방법으로  $\beta^{(r,s)}$ 을 발생 시킨다.  $S$ 가 충분히 큰 짝수를 대입한 깃스 추출법을 사용하면 (4.2)의 추정치를 유도할 수 있다. 예를들어 이중지수 분포를 고려하면 다음과 같다.

$$\begin{aligned} \hat{p}(x_{i+1}|x_i) &= \frac{2}{RS} \sum_{r=1}^R \sum_{s=\frac{S}{2}+1}^S 1/2 \theta^{(r,s)} \beta^{(r,s)} e^{-\beta^{(r,s)} x_{i+1}} \\ &\cdot \exp\left(\theta^{(r,s)} e^{-\beta^{(r,s)} x_i} - \theta^{(r,s)} e^{-\beta^{(r,s)} x_{i+1}}\right) \end{aligned} \quad (4.3)$$

NHPP(1)은 여러 모형들이 존재한다. 최적 모형의 선택은 베이지안 PCPO를 사용하여 해결할 수 있다.  $x_{i+1}$ 에 대한 PCPO는  $c_{i+1} = p(x_{i+1}|x_1, \dots, x_i), i \geq 2$ 에 의해 정의된다. 데이터에 대한 사전분포를 사용하여 각 모형에서  $i$ 에 대한  $c_i$ 를 플롯시킨다. 단  $i=2, 3, \dots, n$ . 평균적으로 가장 큰  $c$ 를 가진 모형이 더 좋은 모형이 된다. 그러나 플롯에서  $c$ 의 값이 아주 작을때는 이상치(outlier)가 존재할 수도 있다. 베이지안 예측 우도 기준은 예측 밀도

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

를 최대화 시키는 모형이 선택되어 진다. 단  $l$ 은 모형들에 대한 지수를 의미한다. 이러한 계산은 (4.3)로부터 직접 이루어진다. PCPO는 과거의 고장시점  $x_1, x_2, \dots, x_i$ 의 집합에 조건화를 시킴으로서  $x_{i+1}$ 의 CPO(conditional predictive ordinates)를 정의한다. 각 고장 시점에서 그다음 고장 시점을 예측하는 것이 주 관심사이다. CPO는 현재의 데이터의 견지에서 선호하는 모형의 접근을 제공한다.

## 5. 수치적인 예

<표1>에 있는 자료는 Geol & Okumoto(1979) 등이 이 자료를 이용하여 결과를 언급하고 있다. 본 연구도 이 자료를 이용하여 결과를 나타내었다.

적당한 사전 분포를 사용하여 2000번 적용을 하였고 70번 반복을 하였다. 사전분포는  $\alpha \sim u(0, 1), \beta \sim \Gamma(1, 0.0001), \theta \sim \Gamma(1, 0.0001)$ 을 선택 이용하였고 메트로폴리스 기법을 적용하기 위하여 사후분포  $\pi(\beta) = 1, \beta > 0$ 을 가정하였다.

FORTTRAN IMSL<sup>3)</sup>를 사용하여 그 결과를 <표3>에 나타내었고 수렴성을 확인하기 위해서 500, 2000번 적용에 50, 70번의 결과를 나타내었다. 그리고 분산분석표를 이용하는 Gelman &

3) MS-FORTTRAN IMSL(INTERNATIONAL MATHEMATICAL AND STATISTICAL LIBRARY(PC 버전 4.0))

Rubin(1992)이 제시한 MC방법을 이용하여 모형선택을 하였다. 모형선택에 있어서  $S=70$ ,  $R=2000$ ,  $i=22$ , 24는 아주 작기 때문에 이상치로 간주하여 계산되었다. <표2>는 밀도와 평균함수를 요약하였고 <그림1>은 생존 함수를 이용하여 미래의 신뢰도를 예측한 결과 시간이 지날수록 모형들이 감소추세를 보이고 있음을 나타내었고 <표4>는 PCPO를 이용하여 모형 선택을 할 수 있다. 즉 (4.4)를 최대로 하는 모형이 효율적인 모형이 된다. <그림2>는 PCPO의 그림에서  $c_2$ 에서  $c_6$  사이는 랄리모형과 이중지수모형이 PCPO의 값이 유사하나 그 이후는 랄리 모형이 커지고 있다 그러나 대체적으로 감마모형은 다른 두모형보다 작은 PCPO의 값을 보여주고 있다. 그 결과로 파악되는 것은 대체적으로 랄리모형은 이중지수모형, 감마모형에 비해 거의 전구간에서 고장이 발생하는 시간이 길다고 볼 수 있으며 또 전체적으로 볼 때 (4.4)를 최대로 하기 때문에 효율적 모형이라고 결론을 내릴 수 있다.

<표1> 소프트웨어의 고장시간(단위 분)

고장번호	고장간격시간	고장번호	고장간격시간
1	9	14	9
2	12	15	4
3	11	16	1
4	4	17	3
5	7	18	3
6	2	19	6
7	5	20	1
8	8	21	11
9	5	22	33
10	7	23	7
11	1	24	91
12	6	25	2
13	1	26	1

<표2> 밀도 함수와 평균 함수

	밀도함수	GOS/NHPP(1)
이중지수	$1/2\beta \exp(- t \beta)$	$\theta[(1 - \frac{1}{2}e^{-\beta t})]$
감마-2	$\beta^2 t \exp^{-\beta t}$	$\theta[1 - e^{-\beta t}(1 + \beta t)]$
랄리	$2\beta t \exp(-\beta t^2)$	$\theta[1 - \exp(-\beta t^2)]$
검벨	$\beta e^{\beta t} \cdot \exp(-e^{\beta t})$	$\theta[(1 - \exp(e^{-t\beta}))]$
곰페르츠	$\exp\{(\beta + \alpha t) - \frac{1}{\alpha} \exp(e^{\beta + \alpha t} - e^\beta)\}$	$\theta[1 - \exp(-\frac{1}{\alpha}(e^{\beta + \alpha t} + e^\beta))]$



<표3> 깃스 표본의 추정치 ( $NHPP(1)$ ), R번 반복 및 S번 적용((5.1) 참조)

1. 이중지수 순서통계량 모형

S	R	$\hat{E}(\theta)$	$\hat{E}(\beta)$
500	50	27.33334	0.008824
500	70	27.33434	0.008922
2000	50	27.35833	0.008904
2000	70	27.39831	0.008701

2. 감마 순서통계량 모형

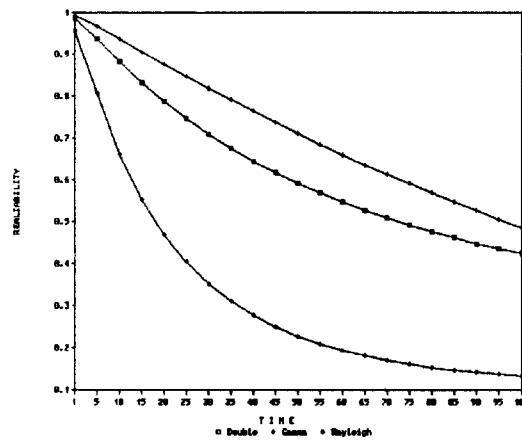
S	R	$\hat{E}(\theta)$	$\hat{E}(\beta)$
500	50	32.48748	0.001023
500	70	32.18952	0.001106
2000	50	32.52896	0.000989
2000	70	32.50198	0.001006

3. 랄리 순서통계량 모형

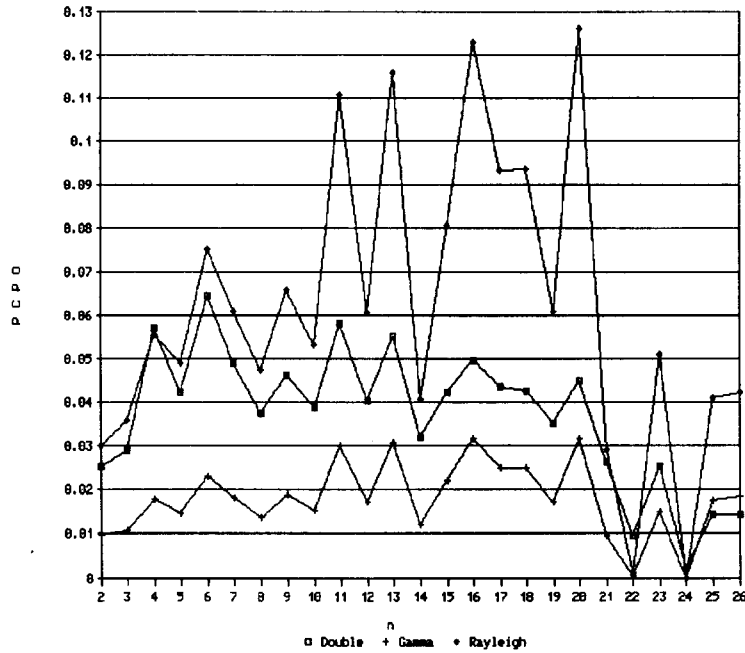
S	R	$\hat{E}(\theta)$	$\hat{E}(\beta)$
500	50	27.37362	0.000491
500	70	27.16149	0.000428
2000	50	26.84469	0.000437
2000	70	27.01269	0.000429

<표4> 모형 선택

	$GOS / NHPP(1)$		
	이중지수 모형	감마 모형	랄리 모형
$\ln C(I)$	-69.34	-70.05	-68.26



<그림1>  $NHPP(1) \hat{P}(x | D_i)$



<그림2> PCPO(NHPP(1))

### 참 고 문 헌

- [1] Casella, G. and George, E. I. (1992), Explaining the Gibbs Sampler, *The American Statistician*, 46, 167-174.
- [2] Choi, K. and Kim H. (1997), Bayesian Approach for Record Value Statistics Using Nonhomogeneous Poisson Process. *The Korean Communication in Statistics Vol 4, No.1 pp. 259-269*
- [3] Cinlar, E. (1975), *Introduction to Stochastic Process*, New Jersey:Prentice-Hall.
- [4] Cox, D. R. and Lewis, P. A., (1966), *Statistical Analysis of Series of Events*, London: Methuen.
- [5] Geisser, S. and Eddy, W. (1979), A Predictive Approach to Model Selection, *Statistical Sciences*, 74, 153-160.
- [6] Gelfand, A. E. and Smith, A. F. M. (1990), Sampling-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, 85, 398-409.
- [7] Gelman, A. E. and Rubin D. (1992), Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, 7, 457-472.

- [8] Goel, A. L. and Okumoto, K. (1979), Time Dependent Error Detection Rate Model for Software Reliability and Other Performance Measures, *IEEE Transactions on Reliability*, 28, 206-211.
- [9] Hastings, W. K., (1970), Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, 87, 97-109
- [10] Jelinski, Z. and Moranda, P. B. (1972), *Software Reliability Research*, in Statistical Computer Performance Evaluation, ed. W. Freiberger, New York: Academic Press, pp. 465-497.
- [11] Kuo, L. and Yang, T. (1993), Bayesian computation of Software Reliability, *Technical Report No. 83-18*, University of Connecticut, Statistics Dept.
- [12] Langberg, N. and Singpurwalla, N. D. (1985), A Unification of Some Software Reliability Models, *SIAM Journal on Scientific and Statistical Computing*, 6, 781-790.
- [13] Lewis, P. A. W. and Shedler, G. S. (1979), Simulation of Nonhomogeneous Poisson Processes by Thinning, *Naval Research Logistic Quarterly* 26, 403-413.
- [14] Musa, J. D., Iannino, A. and Okumoto, K. (1987), *Software Reliability: Measurement, Prediction, Application*, New York: McGraw Hill.
- [15] Musa, J. D. and Okumoto, K. (1984), A Logarithmic Poisson Execution Time Model for Software Reliability Measurement, in *Proceedings Seventh International Conference on Software Engineering Orlando*, pp. 230-238
- [16] Parzen, E. (1962), *Stochastic Process*, San Francisco: Holden-Day.
- [17] Raftery, A. E. (1987), Inference and Prediction for a General Order Statistic Model with Unknown Population Size, *Journal of the American Statistical Association*, 82, 1163-1168.
- [18] Smith, A. F. M. and G. O. Roberts (1993), Bayesian computation via the Gibbs Sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Ser. B*, 55, 3-24.
- [19] Tanner, M. and Wong, W. (1987), The Calculation of Posterior Distributions by Data Augmentation(with discussion), *Journal of the American Statistical Association*, 81, 82-86.