creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

[UCI]I804:24011-200000241944

2011년 8월
석사학위논문

베이지안 소프트웨어 신뢰성 성장 모형에 관한 연구

조선대학교 대학원

전산통계학과

정 덕 환

베이지안 소프트웨어 신뢰성 성장 모형에 관한 연구

조선대학교 대학원

전산통계학과

정 덕 환

베이지안 소프트웨어 신뢰성 성장 모형에 관한 연구

A Study of Bayesian Software Reliability Growth Model

2011년 8월 25일

조선대학교 대학원

전산통계학과

정 덕 환

베이지안 소프트웨어 신뢰성 성장 모형에 관한 연구

지도교수 장 인 홍

이 논문을 이학 석사학위신청 논문으로 제출함

2011년 4월

조선대학교 대학원

전산통계학과

정 덕 환

정덕환의 석사학위논문을 인준함

위 원 장 조선대학교 교수 이 웅 기 (인)

위 원 조선대학교 교수 권 용 만 (인)

위 원 조선대학교 교수 장 인 홍 (인)

2011 년 6 월

조선대학교 대학원

목 차

ABSTRACT ····· v

1. 서 론 ····· 1
2. 신뢰성과 신뢰성 수명분포 ····· 3
   2.1. 신뢰성의 개념 ····· 3
   2.2. 지수분포 ····· 4
   2.3. 와이블분포 ····· 5
   2.4. 감마분포 ····· 6
   2.5. 정규분포 ····· 7
   2.6. 대수정규분포 ····· 9
3. 소프트웨어 신뢰성장모형 ····· 9
   3.1. 소프트웨어 신뢰도 ····· 9
   3.2. 비동질성 포아송 과정 ····· 10
      3.2.1 Jelinski-Moranda 모형 ····· 11
      3.2.2 Musa-Okumoto 모형 ····· 12
      3.2.3 Goel-Okumoto 모형 ····· 12
      3.2.4 Yamada-Ohba-Osaki 모형 ····· 13
      3.2.5 Rayleigh 모형 ····· 14
4. NHPP을 이용한 베이지안 추론 ····· 14
   4.1. 베이지안 추론 ····· 14
   4.2. 깁스 샘플링 알고리즘 ····· 15
   4.3. 메트로폴리스-해스팅스 ····· 16
   4.4. 공액 사전 분포 ····· 18
   4.5. 균등분포를 이용한 무정보적 사전 분포 ····· 19

4.6. NHPP에 대한 추론 및 모형선택 ····· 19
4.6.1 고전적 추론 ····· 19
4.6.2 공액 사전분포를 적용한 베이지안 추론 ····· 21
4.6.3 균등분포를 적용한 무정보적 베이지안 추론 ····· 23
5. 수치적 예제 및 결론 ····· 26
5.1 수치적 예제 ····· 26
5.2. 결론 및 제언 ····· 32
참고문헌 ····· 33

표 목 차

<표 1> 지수분포의 신뢰성 척도 ····· 4
<표 2> 와이블분포의 신뢰성 척도 ····· 5
<표 3> 정규분포의 신뢰성 척도 ····· 8
<표 4> 대수정규분포의 신뢰성 척도 ····· 9
<표 5> 각 표본의 분포에 대해 알려진 공액 사전 분포 ····· 18
<표 6> 소프트웨어 모의 고장시간 ····· 26
<표 7> θ에 대한 모수의 추정치 ····· 27
<표 8> β에 대한 모수의 추정치 ····· 27
<표 9> 상대오차에 의한 모형 비교 ····· 31

그 림 목 차

<그림 1> 지수분포의 그래프 ····· 4
<그림 2> 감마분포의 그래프 ····· 7
<그림 3> 표준정규분포의 그래프 ····· 8
<그림 4> GO모델에서 사전분포 GG 적용 시의 수렴성 ····· 28
<그림 5> GO모델에서 사전분포 UG 적용 시의 수렴성 ····· 28
<그림 6> GO모델에서 사전분포 GU 적용 시의 수렴성 ····· 29
<그림 7> Rayleigh모델에서 사전분포 GG 적용 시의 수렴성 ····· 29
<그림 8> Rayleigh모델에서 사전분포 UG 적용 시의 수렴성 ····· 30
<그림 9> Rayleigh모델에서 사전분포 GU 적용 시의 수렴성 ····· 30
<그림 10> 두 모형에 대한 신뢰도 그래프 ····· 31

Abstract

A study of Bayesian Software Reliability Growth Model

Jung Deok-hwan
Advisor : Prof. Chang, In-Hong, Ph.D.
Department of Computer Science and Statistics,
Graduate School of Chosun University

This paper presents a stochastic model for the software failure phenomenon based on a Nonhomogeneous Poisson Process(NHPP) and performs Bayesian inference using Conjugate prior and noninformative prior by Laplace. The failure process is analyzed to develop a suitable mean value function for the NHPP; expressions are given for several performance measure. The parametric inference of the model using Goel-Okumoto model, and Rayleigh model is discussed. Sum of relative error was used for efficient bayesian selection. The numerical results of this models are applied to real software failure data. We could avoid multiple integration using Gibbs sampling, which is a kind of Markov Chain Monte Carlo method to compute the posterior distribution, and numerical estimates were used for the NTDS data.

1. 서 론

신뢰성(Reliability) 이론은 제품을 사용함에 있어서 의도하는 기간 동안에 사용자가 요구하는 기능을 만족스럽게 수행할 수 있는 확률을 예측하고 높이기 위한 도구로 사용된다. 따라서 의도하는 기간 동안 고장 없이 제품을 사용할 수 있는 확률을 정확하게 구하는 것이 신뢰성의 핵심 목적이다.

과거에는 하드웨어의 신뢰성과 효율성을 유추하기위해 신뢰성 이론을 적용시켰으나, 소프트웨어가 발달하고 삶에 밀접하게 이용됨에 따라, 소프트웨어의 고장률에 대해 관심을 갖기 시작하였다.

소프트웨어의 신뢰성을 예측 하는데 가장 대표적인 방법이라고 할 수 있는 소프트웨어 신뢰성 성장모형(Software Reliability Growth Model)은 소프트웨어의 고장에 대한 원인을 찾아 개선을 하거나 새로운 디자인, 새로운 기술을 도입하게 되므로 시간이 지남에 따라 신뢰도가 증가하는 것을 기대할 수 있는 모형이다. 이모형은 소프트웨어 테스트 단계에서 소프트웨어 고장수와 고장 간격을 근거로 소프트웨어 고장현상을 모형화하였으며 고장발생간격시간, 소프트웨어 신뢰도 및 고장률 등의 신뢰성 평가측도들이 추정되어 미래에 어떤 방식으로 고장이 나타날 것인가를 예측 할 수 있다.

Jelingski와 Moranda(1972)는 지수함수의 성질은 이용한 소프트웨어 성장모형을 제안하였다. 고장의 과정을 비동질적 포아송 과정(Nonhomogenous Poisson)의 성질을 이용하여 모델을 제안한 것으로 Goel-Okumoto(1979)와 Musa, Iannino와 Okumoto(1990)가 가장 대표적인 연구를 진행하였으며, 이 연구를 기반으로 Yamada와 Ohba 그리고 Osaki(1983), Kuo와 Lee, 그리고 Yang(1997)에 의하여 S-shaped NHPP 모델에 대한 연구가 진행되었다.

소프트웨어 고장시간은 시간에 따른 분포로 수명분포가 된다. 따라서 비음(Nonnegative)의 값을 가지기 때문에 이 분야에서는 주로 지수분포, 와이블분포, 감마분포 등 일반화 감마분포(혹은 어랑분포)가 많이 사용되어 왔다. 그러나 이러한 감마분포 외에도 통계자료 해석시 자료를 대수변환 이후에 정규분포를 적용하는 대수정규분포(Lognomal distribution), 이 분포와 유사한 대수로지스틱분포(Loglogistic distribution)도 이 분야에 사용이 가능하다. 또, 자유도(Degree of freedom ; df)에 의존하는 카이제곱분포 (χ^2 distribution)도 적용이 가능하다. 이러한 카이제곱 분포는 변화가 일어날 때까지의 대기시간을 나타내는데 많이 사용되

는 감마분포(Gamma distribution)의 특수한 경우로도 접근 할 수 있기 때문에 소프트웨어 신뢰성 수명분포로 설명할 수 있다.

신뢰도 모수의 값을 계산하는데 있어서 일반적인 방법으로 최우추정법(Maximum Likelihood estimation, MLE)을 사용하였다. 그러나 관측된 데이터에 관한 모형만을 사용하여 모형이 가지는 미지의 모수를 추론하므로, 모수의 불확실성을 고려하지 않는다. 이 단점을 보완하기 위한 모수와 관측된 데이터 모두에 대한 확률모형이 베이지안 추론으로 알려져 있다. 베이지안 추론의 주된 목적은 데이터로부터 얻은 모수에 관한 정보뿐만 아니라 모수에 관한 과거의 경험이나 사전 지식 같은 주관적 견해를 수량화한 모수의 특성을 결합시켜 보다 정확한 추론을 하고자 하는데 있다. 베이지안 이론에 의한 베이지안 추정량이 갖는 좋은 성질들 때문에 이 방법의 이용으로 많은 연구결과를 낳고 있는 추세이지만 베이지안 추정법에서 사전 확률분포인 수명분포가 복잡하면 적분이 불가능하므로 사후정보의 추출이 불가능해진다.

이러한 분야에서 베이즈와 Yang(1996), Choi(1999), Kim과 Shin(2006)등은 깁스 알고리즘(Gibbs algorithm)과 잠재변수(latent variable)등을 이용하는 베이지안 분석방법을 연구하였다.

베이지안 연구에서 지금까지의 연구로 Achcardhk Dey 그리고 Niverthi(1998), Kim과 Shin(2006)은 사전분포로 감마분포를 가정하여 모수에 대한 추정과 모형 선택에 대한 효율적인 방법을 제안하였다. 그러나 사전정보가 없는 경우에 대하여 연구가 이루어져 있지 않으므로 무정보적 사전분포의 베이지안 접근방법이 요구된다.

본 연구에서는 Goel-Okumoto모형과 Rayleith 모형에 대하여 공액사전분포와 무정보적 사전분포 중 라플라스에 의해서 처음 고안된 사전분포를 이용하여 모형의 효율성에 대해 비교하고자 한다.

논문의 구성은 2장에서 신뢰성과 신뢰성에 사용되는 수명분포에 대해 소개하고, 3장에서 소프트웨어 신뢰성장모형의 개념과 개발된 모형에 대해 소개한다. 4장에서는 베이지안 추론의 개념과 사용되는 시뮬레이션 방법, 각각의 사전분포를 고려한 NHPP모형에 대해 설명하였으며, 마지막으로 5장에서는 실제 데이터를 이용하여 수치적 예제 실행 및 결론을 제시 하였다.

2. 신뢰성과 신뢰성 수명분포

2.1 신뢰성의 개념

근래에 시스템이 복잡하고, 다양해짐에 따라 모든 시스템에서 신뢰성의 중요성이 대두되고 있다. 여기서 신뢰성이란 제품이 주어진 사용 조건 아래서 의도하는 기간 동안 정해지는 기능을 성공적으로 수행하는 능력 또는 성질이다.

이 신뢰성을 가장 기본적으로 나타내는 척도로 제품의 수명이 t시간 이상이 될 가능성을 나타내는 신뢰성 $R(t)$를 생각한다. $f(t)$가 제품의 분포인 경우에 아이템 분포의 누적분포 함수인 $F(t)$는 제품이 t시간 동안에 고장이 발생할 확률을 나타내므로

$$R(t) = \int_t^{\infty} f(t)\,dt = 1 - \int_0^t f(t)\,dt = 1 - F(t)$$

와 같은 관계가 나타나게 된다.

n개의 데이터가 서로 독립적인 경우 해당 제품의 신뢰성을 재는 척도로 MTTF (Mean Time To Failure, 평균고장시간)라고 한다. 수명 분포가 $f(t)$인 경우 모집단의 평균 MTTF는

$$MTTF = \int_0^{\infty} t f(t)\,dt$$

이며, 이 평균은 분포의 무게중심을 나타낸다.

한편, 냉장고, TV등 수리가 가능한 제품이 서로 이웃하는 고장 간 시간동작의 평균 수명은 MTBF(Mean Time Between Failures, 평균 고장 간격)이라 한다.

내구성을 측정하는 또 다른 척도로 임의의 시점 t에서 잔존하는 것 중 몇 개가 순간적으로 고장 날 것인지를 나타내는 (순간)고장률(Failure Rate, Hazard Rate)은

$$\lambda(t) = \frac{f(t)}{R(t)}$$

과 같이 나타난다.

2.2 지수분포

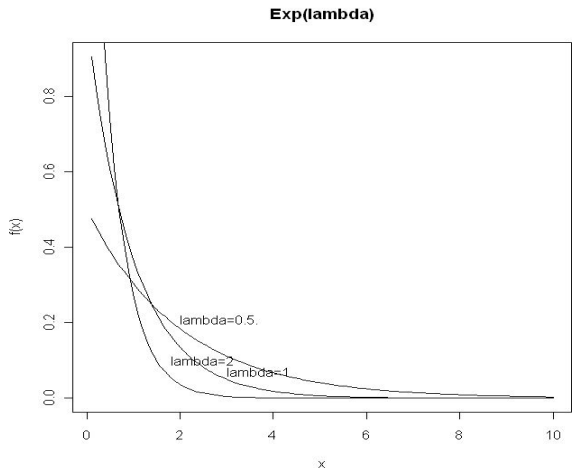
지수분포(Exponential distribution)는 제품의 신뢰도 예측, 신뢰성 시험의 설계 등 신뢰성공학 이론에서 중요한 역할을 한다. 지수분포의 신뢰성 이론에서의 중요성은 무엇보다도 노화 영향(Aging effect)을 보이지 않는다는 것이다. 따라서 제품의 수명이 지수분포에 의하여 설명되어진다는 가정 하에서는 부품의 노화가 마모를 측정하는 관심 모수들의 값에 영향을 미치지 못한다는 것이다. 특히 지수분포는 무기억성(Memoryless property)을 가지고 있는 유일한 연속 확률분포로 잘 알려져 있으며 이러한 성질은 지수분포를 따르는 제품이 앞으로 일정 시간 이상을 가동하는 확률이 현재의 부품의 나이(Age)에 무관하게 일정하다는 것을 의미한다.

지수분포의 확률 밀도 함수 및 관련 모수들의 값은 다음과 같다.

<표 1> 지수분포의 신뢰성 척도

신뢰성 척도	지수분포
분포함수	$f(t)=\lambda e^{-\lambda t}$
불신뢰도 함수	$F(t)=1-e^{-\lambda t}$
신뢰도 함수	$R(t)=e^{-\lambda t}$
고장률함수	$\lambda(t)=\lambda$

R을 이용하여 자유도가 0.5, 1, 2 인 지수분포를 그래프로 나타내면 <그림 1>과 같이 나타난다.



<그림 1> 지수분포의 그래프

Drenick(1960)에 의해 보전을 수반하는 여러 개의 다른 형태의 부품들이 조립되어 만들어진 복잡한 기기나 시스템의 고장 간격은 비교적 넓은 조건하에서 근사적으로 지수 분포를 따른다고 하는 중요한 성질이 알려졌다.

(1) 시스템은 반드시 동종이 아닌 구성 요소로 이루어지며 고장 시간은 상호간에 확률적으로 독립이다.
(2) 어떤 구성 요소가 고장이 나도 전체 시스템의 고장을 일으킨다.
(3) 각 구성 요소는 고장이 나면 즉시 교체할 수 있다.
(4) 고장의 발견 및 고체에 소요되는 시간은 무시할 수 있다.

이때에 시스템을 충실히 오랜 기간 동작시키고 또한 구성 요소수가 충분히 크다면 시스템의 고장 시간의 간격은 점근적으로 지수 분포를 따른다.

2.3 와이블분포

신뢰성 모델로써 가장 많이 사용되는 것이 와이블 분포이다. 스웨덴의 과학자 와이블(W. Weibull)이 재료의 파괴강도를 설명하기 위하여 처음으로 제안하였다. 와이블 분포는 지수 분포를 일반화한 분포로서 고장률 함수가 상수, 증가함수 또는 감소함수인 수명 분포들을 모형화 할 때 적당한 분포이다. 3모수를 갖는 와이블 분포의 신뢰도 함수 등은 다음과 같이 표현된다.

<표 2> 와이블분포의 신뢰성 척도

신뢰성 척도	와이블분포
분포함수	$f(t)=\left(\frac{\beta}{\eta}\right)\left(\frac{t-\gamma}{\eta}\right)^{\beta-1}\exp\left\{\left(-\frac{t-\gamma}{\eta}\right)^{\beta}\right\}$
불신뢰도 함수	$F(t)=1-\exp\left\{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right\}$
신뢰도 함수	$R(t)=\exp\left\{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right\}$
고장률함수	$\lambda(t)=\left(\frac{\beta}{\eta}\right)\left(\frac{t-\gamma}{\eta}\right)^{\beta-1}$

γ는 위치모수로 특별하게 주어지지 않은 경우 $\gamma=0$으로 가정하고, β는 형상모수로 분포의 모양을 결정하며, η는 척도모수로 가로 축의 척도를 규정한다. 여기서 형상 모수 β에 따라 분포의 모양이 결정된다. $\beta>1$이면 $\lambda(t)$가 단조증가하며

$\beta=1$이면 지수 분포가 되어 $\lambda(t)$가 일정하고, $\beta<1$이면 $\lambda(t)$가 단조 감소한다. 특히, $\beta=2$인 경우에는 Rayleigh 분포로 알려져 있으며 이 경우의 고장률 함수는 기울기가 $2/\eta^2$인 선형증가 함수이다. 또한 $m=3.44$인 경우의 와이블 분포는 거의 정규 분포와 가까우며 이 경우 이 분포의 평균값은 중앙값과 같다.

2.4 감마분포

감마분포는 2개의 모수에 의하여 특정 지어지는 수명분포로 지수분포와 와이블 분포와 함께 신뢰성 분석에서 가장 많이 응용되는 연속확률분포이다. 연속확률변수 T가 감마분포일 경우 확률밀도함수는

$$f(t)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}t^{\alpha-1}e^{-\frac{t}{\beta}} \qquad t\geq 0$$

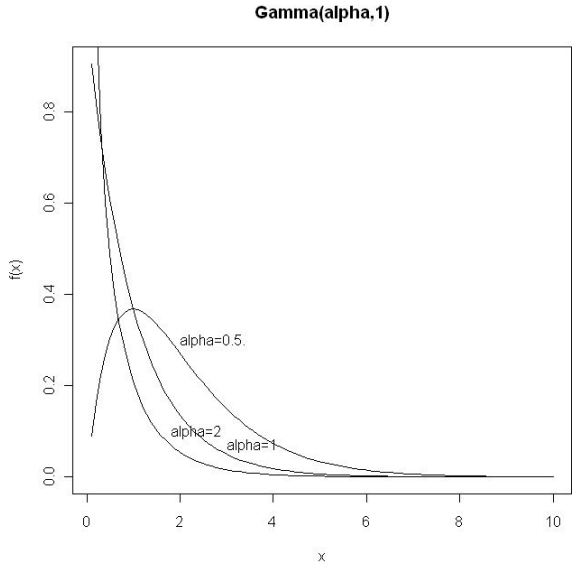
과 같이 나타난다. 여기서 감마함수 $\Gamma(\alpha)=\int_0^{\infty}t^{\alpha-1}e^{-t}dt$이다. 특히 감마분포의 밀도 함수에서 $\beta=1$인 경우를 표준 감마분포로 알려져 있다.

감마분포의 응용분포로 $1/t$가 감마 분포를 따르면 t는 역감마 분포를 따르며, 이 경우의 밀도 함수는

$$f(x)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}t^{-(\alpha+1)}e^{-\beta t} \qquad t\geq 0$$

과 같다.

R을 이용하여 $\beta=1$일 때 α가 각각 0.5, 1, 2인 감마분포를 그림으로 나타내면 <그림 2>와 같이 나타난다.



<그림 2> 감마분포의 그래프

2.5 정규분포

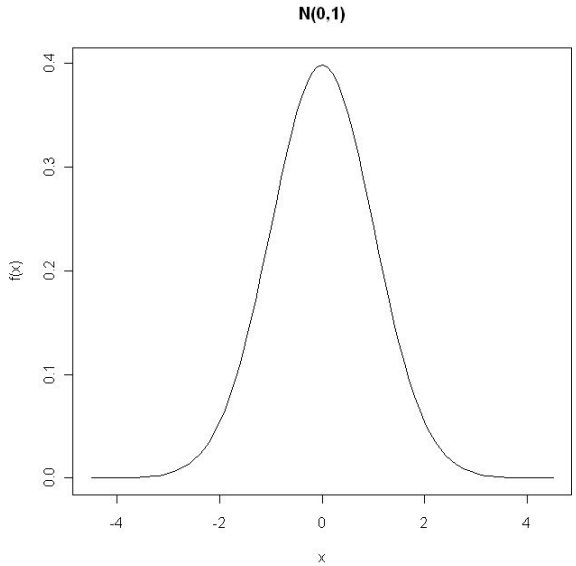
정규 분포는 연속확률변수를 설명하는 분포로서 분포이론에서 가장 핵심적인 역할을 한다. 특히 대표본의 경우에는 모집단의 분포에 관계없이 표본의 평균이 근사적으로 정규 분포를 따른다는 중심극한정리 덕분에 광범위하게 응용되고 있다. 신뢰성 분야에서는 어떤 원인 때문에 특정한 시간에 집중적으로 고장이 발생하는 경우에 정규 분포를 많이 사용한다. 특히 마모, 피로, 부식 등의 원인으로 고장이 평균 근방에서 집중적으로 발생하는 경우에 적합하며, 강도나 스트레스에 대한 데이터를 적합시키는 데에도 많이 이용된다. 2모수를 갖는 정규 분포의 신뢰도 함수 등은 다음과 같이 나타난다.

<표 3> 정규분포의 신뢰성 척도

신뢰성 척도	정규분포
분포함수	$f(t)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}$
불신뢰도 함수	$F(t)=\Phi\left(\frac{t-\mu}{\sigma}\right)$
신뢰도 함수	$R(t)=1-\Phi\left(\frac{t-\mu}{\sigma}\right)$
고장률함수	$\lambda(t)=\frac{\left(\frac{1}{\sigma}\right)\Phi\left(\frac{t-\mu}{\sigma}\right)}{1-\Phi\left(\frac{t-\mu}{\sigma}\right)}$

특히, $\mu=0$, $\sigma=1$을 따르는 정규분포는 표준 정규분포로 알려져 있다. R을 이용한 표준정규분포의 그래프는 <그림 3>과 같다.

<그림 3> 표준정규분포의 그래프



2.6 대수정규분포

대수정규분포(Log-normal distribution, 로그정규분포)는 고장률함수가 어느 시점에서 단조성에 변동이 생긴다는 것이 발견되어 수명분포의 검정 등에 많이 응용되고 있다. 2모수를 갖는 신뢰도 함수 등은 다음과 같이 나타난다.

<표 4> 대수정규분포의 신뢰성 척도

신뢰성 척도	대수정규분포
분포함수	$f(t)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(\ln t-\mu)^2}{2\sigma^2}}=\frac{1}{\sigma t}\Phi\left(\frac{\ln t-\mu}{\sigma}\right)$
불신뢰도 함수	$F(t)=\Phi\left(\frac{\ln t-\mu}{\sigma}\right)$
신뢰도 함수	$R(t)=1-\Phi\left(\frac{\ln t-\mu}{\sigma}\right)$
고장률함수	$\lambda(t)=\frac{\Phi\left(\frac{\ln t-\mu}{\sigma}\right)}{\sigma t\left\{1-\Phi\left(\frac{\ln t-\mu}{\sigma}\right)\right\}}$

특히 $\lambda(t)$가 처음 일정 기간 동안 증가하다가 나중에는 감소하는 UBR(Upside down Bathtub failure Rate)을 보여주는데 이러한 경향 변동의 특성 때문에 대수정규 분포는 수리 시간을 나타내는 분포 등에 적당한 것으로 알려져 있다.

3. 소프트웨어 신뢰 성장 모형

3.1 소프트웨어 신뢰도

소프트웨어는 하드웨어와 같이 제품이 마모, 과부하 등의 제품 손상이 신뢰성에 영향을 주는 것이 아니라, 제품의 합리적인 개발 여부에 따라서 신뢰도가 영향을 받게 된다. 그리고 소프트웨어의 신뢰성에 대하여 동일 소프트웨어의 중복 사용은 소스가 같기 때문에 다른 두 개의 복사본은 같은 상황에서 같은 고장을 유발하므로 하드웨어와 같이 반복적인 실험으로 고장의 발생을 낮추는 방법으로는 신뢰

도를 높일 수 없다. 따라서 독립적으로 개발된 같은 목적의 소프트웨어를 여러 가지 생산하여 신뢰성을 향상 시킬 수 있다. 독립적인 소프트웨어를 병렬 중복 사용해서 개별적인 결과 값들을 얻고 가장 많이 일치하는 결과 값을 사용하는 방식으로 하는 투표방식(Voting method)을 이용해 신뢰성을 향상 시킬 수 있다.

소프트웨어의 신뢰성 측정은 개발단계에서 고장을 유발하는 결함을 찾아내는 것이 시간, 경비의 모든 면에서 완성된 제품의 결함을 찾아내는 것보다 절약되기 때문에 개발의 과정에 치중되어 신뢰성을 측정하게 된다.

3.2 비동질성 포아송 과정

시간 t까지 발견된 총 고장 수를 나타내는 계수과정 $\{N(t), t \geq 0\}$를 가정할 때, 이 확률과정에 대하여 단위 시간당에 발견되는 에러수를 나타내는 강도함수 $\lambda(t)$를 가지는 비동질적 포아송 과정(Nonhomogeneous Poisson Process : NHPP)을 가정한다. 따라서 테스트 공정 혹은 운용단계에 있어서 발견되는 고장 수에 대하여 다음의 가정을 한다.

(1) $N(0) = 0$
(2) $\{N(t), t \geq 0\}$은 독립증분을 갖는다. 즉, 서로 다른 테스트 시간구간에서 발견되는 총 에러 수는 통계적으로 독립이다.
(3) $\Pr\{N(t+h) - N(t) = 1\} = \lambda(t)h + o(h)$. 즉, 임의의 극소 테스트 시간 구간 Δt에서 한 개의 고장이 발견될 확률은 강도함수 $\lambda(t)$에 비례한다.
(4) $\Pr\{N(t+h) - N(t) \geq 2\} = o(h)$. 즉 임의의 극소 테스트 시간구간 Δt에서 두 개 이상의 고장이 발견될 확률은 무시해도 좋을 만큼 적다

만약 $m(t) = \int_0^t \lambda(h)dh$이면

$$\Pr\{N(t+h) - N(t) = n\} = e^{-\{m(t+h) - m(t)\}} \frac{\{m(t+h) - m(t)\}^n}{n!}$$

이므로 $N(t+h) - N(t)$는 평균 $m(t+h) - m(t)$인 포아송분포를 따른다. 따라서 $N(t)$는 평균 $m(t)$인 포아송분포를 따른다.

3.2.1 Jelinski-Moranda 모형

Jelinski-Moranda(JM) 모형은 소프트웨어 신뢰 성장 모형에서 가장 고전적이고 기본적인 모형으로 초기 고장 수에 의존하는 모형이며, 소프트웨어 고장 데이터를 기술하는 가장 간단한 모형으로 지수함수의 성질을 이용한 소프트웨어 신뢰 성장 모형에 대한 연구를 시작하면서 연구 초기에는 소프트웨어의 고장률이 남아있는 고장 수에 비례하는 고정된 상수로 가정한 것을 바탕으로 모델연구를 시작했다.

N_0을 테스트 단계에서 초기의 소프트웨어 오류의 총 개수라고 하면, ϕ는 실패강도로서 오류마다 오류 발견률을 나타내는 상수라고 하면 Jelinski-Moranda 모형의 가정은 다음과 같다.

(가정 1) N_0는 소프트웨어에 의존하는 초기 오류의 총 개수이고 미지의 고정 된 상수이다.

(가정 2) 고장의 원인은 한 개의 오류에 의해 발생하고, 추가적인 오류는 발생하지 않으며 디버깅 절차는 완벽하다.

(가정 3) 고장이 발생되면 즉시 수정되고 고장 발생 간격 시간 $x_i = t_i - t_{i-1}$는 각각 독립적으로 평균이 $1/\phi(N_0 - i + 1)$인 지수분포를 따른다.

(가정 4) 남아 있는 소프트웨어 실패율은 항상 같은 상수이다.

위의 네 가지 가정은 오늘날 소프트웨어 환경변화를 따라가기 위하여 다양한 수정·보완 연구가 진행되어지고 있다. 첫째, 고장발생 사이의 시간이 종속적인 경우에 대한 연구가 진행되고 있고, 둘째 소프트웨어에 남아있는 결함은 고장률이 같다는 가정에 대하여 오늘날 연구는 초기에 발견되는 고장의 크기와 시험 중간 단계 중에서 발생되어지는 고장의 크기 및 마지막 단계에서 발견된 고장의 크기가 다르다는 관점에서 모델에 대한 연구가 진행되고 있다.

Jelinski-Moranda 모형에서 모든 x_i에 대하여 고장 확률 밀도 함수는 다음과 같다.

$$L(N_{0,}\phi|D_t) = \prod_{i=1}^{n}(N_0 - i + 1)\phi \exp\left\{-\phi\sum_{i=1}^{n}(N_0 - i + 1)x_i\right\}$$

3.2.2 Musa-Okumoto 모형

Musa-Okumoto(MO)모형은 JM 모형의 변형된 모형으로 대수 포아송 실행 시간 모형(Logarithmic Poisson Execution Time Model)이라고도 불린다. $\tilde{\beta}=\{\beta_1,\beta_2,\cdots\beta_n\}$를 갖는다 하고, $\lambda(t)=m'(t)$이며, $0<x_i<t(i=1,2,\cdots n)$인 Musa-Okumoto 모형의 가정은 다음과 같다.

(가정 1) 실행시간의 간격과 고장과정은 NHPP이다.
(가정 2) 제거된 오류의 수의 함수에 따라 고장 강도가 지수적으로 감소한다.
(가정 3) 고장 간격 시간 $t_i=x_i-x_{i-1}$은 각각 독립적인 모수이다.

Musa-Okumoto 모형에서 모든 x_i에 대하여 강도함수와 평균함수는

$$m_{MO}(t)=\alpha\ln\left(1+\frac{t}{\beta}\right),$$

$$\lambda_{MO}(t)=\frac{\alpha}{x_i+\beta}$$

여기서

$$L(\beta|D_t)=\left(\prod_{i=1}^{n}\frac{\alpha}{x_i+\beta}\right)\exp\left[-\alpha\ln\left(1+\frac{t}{\beta}\right)\right]$$

3.2.3 Goel-Okumoto 모형

Jelinski-Moranda 모형은 모수 추정에 초기치의 값에 영향을 많이 받는 점과 모든 고장이 같은 크기를 가진다는 가정을 포함한다. 이러한 문제점을 보완하기 위하여 Goel-Okumoto 모형이 제안 되었는데, 이 모형은 가장 기본적인 NHPP 모형으로서 단위 시간당 발견된 결함 수는 그 시간에 잔존하는 결함 수에 비례한다는 기본적인 개념을 가지고 있다. Goel-Okumoto(GO) 모델에 대한 가정은 다음과 같다.

(1) t시간까지 발견된 고장의 누적 수는 포아송분포를 따른다.
(2) 모든 고장은 독립이며 발견될 가능성은 같다.
(3) 모든 발견된 고장은 즉시 제거되어지며 새로운 고장을 만들어 내지 않는

다.

NHPP의 성질을 만족하는 GO 모델의 평균값 함수 $m(t)$와 강도함수 $\lambda(t)$는 각각 다음과 같다.

$$m_{GO}(t) = \theta\left(1 - e^{-\beta t}\right),$$
$$\lambda_{GO}(t) = \theta\beta e^{-\beta t}$$

여기서 θ는 소프트웨어 시험과정에서 발견되어질 수 있는 전체 결함의 수이다. t를 최종 고장시점 x_n으로 대체하고 강도함수 $\lambda(t)$를 이용하여 우도함수를 구하면 다음과 같다.

$$L_{GO}(\theta,\beta|D_{x_n}) = \left(\prod_{i=1}^{n}\theta\beta e^{-\beta x_i}\right)\exp\left[-\theta(1-e^{-\beta t})\right]$$

3.2.4 Yamada-Ohba-Osaki 모형

Yamada-Ohba-Osaki 모형은 지연 S자형(delayed S-Shaped) SRGM이라고도 하는데, 고장 발생 시간으로부터 그 고장의 원인인 결함을 제거하는 시간까지의 시간적인 지연을 고려하고 있으며, 테스트에 의해 결함을 발견하기 위해서는 우선 소프트웨어 고장의 현상을 확인하고 그 다음에 원인을 규명해야 한다는 개념을 가지고 있다. Yamada -Ohba-Osaki 모형의 평균값 함수 $m(t)$와 강도함수 $\lambda(t)$는 각각 다음과 같다.

$$m_{YOO}(t) = \theta\{1-(1+\beta t)\exp(-\beta t)\}, \qquad (\theta > 0,\ \beta > 0)$$
$$\lambda(t) = \theta\beta^2 t\exp(-\beta t)$$

위의 두 함수를 이용하여 우도 함수를 구하면 다음과 같다.

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

3.2.5 Rayleigh 모형

Goel-Okumoto 모형과 Yamada-Ohba-Osaki 모형의 평균값 함수와 강도함수를 일반화하게 되면 다음과 같이 표현된다.

$$m(t) = \theta\{1 - \exp(-\beta t^2)\},$$
$$\lambda(t) = 2\theta\beta t \exp(-\beta t^2)$$

위의 일반화 된 함수를 이용하여 Kan과 Kim, Lee(2001)는 Rayleigh 함수를 적용한 NHPP모형을 제시하였으며, 우도함수는 다음과 같다.

$$L_R(\theta,\beta)|D_{x_n}) = \left(\prod_{i=1}^{n} 2\theta\beta x_i e^{-\beta x_i^2}\right)\exp\left[-\theta\left(1 - e^{-\beta x_n^2}\right)\right]$$

4. NHPP을 이용한 베이지안 추론

4.1 베이지안 추론

베이지안 추론은 관측된 데이터와 모수 모두에 대한 확률 모형은 고려하여 모수에 대한 추론을 하는 방법이다. 베이지안 추론을 이용한 분석 과정은 다음과 같다.

(1) 완전확률모형(full probability model) 수립 : 데이터와 모수에 대한 결합 확률분포를 수립한다. 이 때 완전확률모형은 과학적 문제에 대한 이해와 데이터 수집과정에 관련된 모든 정보를 고려하여 수립한다.

(2) 사후분포(posterior distribution)의 계산 : 과학적 문제에 대한 정보를 나타낸 모수에 대한 분포와 관측된 데이터의 확률분포가 결합하여 사후분포를 생성한다.

(3) 모형의 진단 : 모형의 데이터에 대한 적합정도, 결론의 타당성, 가정에 대한 결과의 민감성을 평가한다. 필요한 경우 모형을 바꾸거나 확장한다.

완전 확률 모형을 수립하기 위한 모수 θ와 관측된 데이터 y에 대한 확률모형은 θ를 인덱스로 생각하여 $p(y|\theta)$로 나타내면 $L(\theta) \propto p(y|\theta)$으로 나타난다. 이 때 θ에 대한 불확실성을 수량화 하는 사전분포(prior distribution)를 결정하고 $\pi(\theta)$로 나타낸다. 사전분포는 데이터를 관측하기 이전에 결정되어지고 과거의 경험이나 데이터의 충분한 이해를 통해 정해진다.

베이즈 정리에 의해 y가 주어졌을 때 θ의 조건부 분포인 사후분포 $p(\theta|y)$를 구하면

$$p(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{p(y)}$$

가 되고 정규화 상수 $p(y)$를 생략하면 다음과 같이 표현된다.

$$p(\theta|y) \propto p(y|\theta)\pi(\theta)$$

4.2 깁스 샘플링 알고리즘(Gibbs Sampling Algorithm)

깁스 샘플링 알고리즘은 Geman에 의해 연구가 되어 Hastings에 의해 발전되었다. 깁스 샘플링 알고리즘은 다차원 이상의 적분을 계산하기 위한 것으로 신경망, 전문가 시스템과 같은 척도가 크고 복잡한 모형에 적용되어 사용되어 왔다.

깁스 샘플링 알고리즘은 다차원의 결합확률분포가 복잡하여 직접 랜덤표본을 생성하기 어려운 경우 각 변수의 조건부확률분포로부터 랜덤표본을 반복적으로 생성하면서 적절한 조건 하에서 이들의 극한분포가 결합확률밀도함수가 된다는 사실에 근거하여 난수를 생성하는 알고리즘이다. 이때 개개의 조건확률분포로부터의 난수 발생이 쉬워야 한다.

깁스 샘플링 알고리즘에서 각 변수의 조건부 확률분포로부터 반복적으로 생성된 난수들이 결합확률밀도함수를 극한분포로 갖는 벡터 마코브 연쇄를 구성한다. 뿐만 아니라 각 변수의 조건부확률분포로부터 생성된 랜덤표본의 극한분포는 각 변수의 주변확률분포가 된다.

모수 θ에 대하여 $\Theta = (\theta_1, \theta_2 \cdots, \theta_p)^T$의 결합사후분포 $p(\Theta|y)$로부터의 베이지안 추론을 위한 깁스 샘플링 알고리즘은 다음과 같다.

(1) 적절한 초기값 $\Theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_p^{(0)})^T$를 생각한다.

(2) $\theta_1^{(1)} \sim p[\theta_1 | y, \theta_2^{(0)}, \cdots, \theta_p^{(0)}]$,

$\theta_2^{(1)} \sim p[\theta_2 | y, \theta_1^{(1)}, \theta_3^{(0)}, \cdots, \theta_p^{(0)}]$,

$\theta_3^{(1)} \sim p[\theta_3 | y, \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \cdots, \theta_p^{(0)}]$,

$\vdots$

$\theta_p^{(1)} \sim p[\theta_p | y, \theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \cdots, \theta_{p-1}^{(1)}]$

(3) (1), (2)를 k번 반복한다.

위의 깁스 샘플링 과정을 통하여 나타나는 모수 θ에 대한 샘플링의 추정 결과는 다음과 같다.

$$\theta_1^{(k)} \sim p[\theta_1 | y, \theta_2^{(k-1)}, \cdots, \theta_p^{(k-1)}],$$
$$\theta_2^{(k)} \sim p[\theta_2 | y, \theta_1^{(k)}, \theta_3^{(k-1)}, \cdots, \theta_p^{(k-1)]},$$
$$\theta_3^{(k)} \sim p[\theta_3 | y, \theta_1^{(k)}, \theta_2^{(k)}, \theta_4^{(k-1)}, \cdots, \theta_p^{(k-1)}],$$
$$\vdots$$
$$\theta_p^{(k)} \sim p[\theta_p | y, \theta_1^{(k)}, \theta_2^{(k)}, \theta_3^{(k)}, \cdots, \theta_{p-1}^{(k)}]$$

따라서 깁스 표본 $\Theta^G = (\theta_1^{(k)}, \cdots, \theta_p^{(k)})$를 얻을 수 있으며, 이러한 과정을 전체 m번 반복하여 깁스 표본 $\Theta_1^G, \cdots, \Theta_m^G$를 얻을 수 있다. 이 깁스 표본을 사용하여 사후분포를 추정하며 사후평균과 사후분산을 계산할 수 있다.

4.3 메트로폴리스-해스팅스 알고리즘 (Metropolis-Hastings Algorithm)

메트로폴리스-해스팅스 알고리즘은 볼츠만분포의 특별한 경우를 위해 이것을 발표한 Metropolis와 이것을 일반화한 Hastings의 이름을 따서 명명되었다. 깁스 샘플링 알고리즘은 메트로폴리스-해스팅스 알고리즘은 특별한 경우이며, 일반적인

적용에는 제약이 있지만 보통 더욱 빠르고 사용하가 쉽다.

메트로폴리스-해스팅스 알고리즘은 난수를 발생하고자 하는 목표 확률분포가 $f(\theta) \propto \pi(\theta)$ 혹은 $f(\theta) = c \cdot \pi(\theta)$와 같이 주어지고, 이 때 정규화 상수를 모르는 경우 $f(\theta)$로부터 난수를 발생시키는 알고리즘이다.

메트로폴리스-해스팅스 알고리즘을 사용하기 위해서 다음과 같은 과정이 필요하다.

(1) 초기값 θ^0를 선택한다.

(2) $k=1$로 놓는다.

(3) 적당한 조건부 확률밀도함수 $q(\theta^*|\theta^{k-1})$로부터 후보난수 θ^*를 생성한다.

(4) 균일분포 $U(0,1)$로부터 난수 u를 생성하여 $u \le a$이면 $\theta^k = \theta^*$으로 갱신하고 그렇지 않을 경우에는 $\theta^k = \theta^{k-1}$으로 한다. 이 때 α는 다음과 같다.

$$\alpha = \min\left\{1,\ \frac{\pi(\theta^*|y)}{\pi(\theta^{k-1}|y)} \frac{q(\theta^{k-1}|\theta)}{q(\theta^*|\theta^{k-1})}\right\}$$

(5) $k=k+1$을 적용하여 (3)으로 돌아간다.

다음의 결과로부터 얻은 $\theta^{(1)}, \theta^{(2)}, \cdots$는 $f(\theta)$를 극한분포로 가지는 마코브 연쇄이며, 이 표본을 통하여 사후분포의 사후평균과 사후분산을 계산할 수 있다.

메트로폴리스-해스팅스 알고리즘에서 후보생성밀도함수 $q(\theta^*|\theta^{k-1})$를 어떻게 정하는가에 따라 난수를 생성하는 과정이 많은 영향을 받게 된다. 이 후보 생성 밀도 함수 $q(\theta^*|\theta^{k-1})$에 대해 주로 사용 되는 것은 확률보행연쇄와 $\pi(\theta^*)$의 알려지는 형태를 사용하는 방법이 있다.

첫 번째, 확률보행연쇄를 이용한 방법은 $q(\theta^*|\theta^{k-1}) = q_1(|\theta^* - \theta^{k-1}|)$로 주어진다. 여기서 q_1은 특정된 밀도함수 이다. 따라서 난수 $\theta^* = \theta^{k-1} + \theta^e$의 방법으로 생성된다. 여기서 θ^e는 증분 확률변수로서 $\theta^e \sim q_1(|\theta^e|)$를 따른다. 흔히 q_1의 대표적인 선택으로는 균일분포, 정규분포, 코쉬분포 등을 사용한다. 확률보행연쇄를 이용한 경우 이동확률 α는 다음과 같다.

$$\alpha = \min\left[\frac{\pi(\theta^*)}{\pi(\theta^{k-1})}\right]$$

두 번째, 알려진 형태의 $\pi(\theta^*)$를 사용하는 방법은 만약 $\pi(\theta^*)$가 $\pi(\theta^*) \propto \psi(\theta^*) \cdot h(\theta^*)$와 같이 표현될 때 $h(\theta^*)$가 난수를 쉽게 발생할 수 있는 밀도 함수이고 $\psi(\theta^{k-1})$가 균일유계이면 $q(\theta^*|\theta^{k-1}) = h(\theta^*)$를 선택한다. 알려진 형태의 $\pi(\theta^*)$를 사용하는 경우 이동확률 α는 다음과 같다.

$$\alpha = \min\left[\frac{\psi(\theta^*)}{\psi(\theta^{k-1})}\right]$$

4.4 공액 사전 분포

모수에 대한 사전분포를 선택할 때 그 분포의 형태를 모른다면 특정계열의 분포를 이용하면 편리하다. 이때 널리 알려져 있는 분포족(Distribution family)을 주로 이용한다. 사전분포와 사후분포가 같은 분포족에 속할 때 사전분포를 공액사전분포(Conjugate prior)라고 한다.

공액 사전분포를 이용하는 이유는 우도함수로 이항분포와 정규분포와 같은 표준모형을 사용하는 것과 같이 결과를 이해하기가 쉬우며, 수학적 계산을 매우 간편하게 해주기 때문이다. 흔히 사용되는 공액 사전 분포를 아래와 같이 요약할 수 있다.

<표 5> 각 표본의 분포에 대해 알려진 공액 사전 분포

표본 분포	공액 사전 분포
$B(n,\theta)$	$\theta \sim Beta(\alpha,\beta)$
$Poi(\theta)$	$\theta \sim Gamma(a,b)$
$N(\mu,\sigma^2)$, σ^2 : known	$\mu \sim N(\mu_0,\tau_0^2)$
$N(\mu,\sigma^2)$, μ^2 : known, $\tau = 1/\sigma^2$	$\tau \sim Gamma(a,b)$
$Gamma(\alpha,\lambda)$, α : know	$\lambda \sim Gamma(a,b)$

4.5 균등분포를 이용한 무정보적 사전분포

모수에 대한 사전정보가 거의 없는 경우 사전분포를 결정하기가 어려우며, 이런 경우 사후분포에 대한 최소한의 임무만 수행할 사전분포가 필요하다. 이러한 사전분포를 일반적으로 무정보적 사전분포(Noninfomation prior)라고 한다.

흔히 모수에 관한 사전정보에 확신이 없을 때 혹은 모수에 관한 사전정보가 미흡할 때 사전분포는 우도함수에 비해 상대적으로 편편하게 주어지며 이를 모호사전분포 혹은 편편한 사전분포, 퍼진 사전분포라 부른다. 이와 같은 경우는 우도함수가 사전분포를 지배하는 상황이다.

무정보적 사전분포의 대표적 예로서 다음의 두 경우를 들 수 있다.

1) 만약 $0 \le \theta \le 1$이면 $\pi(\theta)=1$, $0 \le \theta \le 1$이 θ에 대한 무정보적 사전분포이다.
2) 만약 $-\infty \le \theta \le \infty$이면 $\theta \sim N(\mu_0, \tau_0^2)$에서 $\tau_0^2 \to \infty$로 두면 무정보적 사전분포가 된다. 즉, 사전분포의 분산 τ_0^2을 충분히 크게 함으로 무정보적 사전분포를 얻을 수 있다.

무정보적 사전분포를 사용할 때 $\int_{-\infty}^{\infty} \pi(\theta) d\theta = +\infty$이면 부적절한 사후분포가 된다. 부적절한 사전분포는 부적절한 사후분포를 초래할 수 있으므로 주의해야 한다.

4.6. NHPP에 대한 추론 및 모형선택

본 논문에서는 Goel-Okumoto 모형에 대한 고전적 추론 방법과 베이지안 추론 방법을 이용하여 모수에 대한 추정치를 추정하였으며, 베이지안 추론 방법에서는 공액 사전 분포를 사용한 경우와 무정보적 사전분포를 고려한 경우의 모수 추론을 실시하였다.

4.6.1 고전적 추론

Goel-Okumoto 모형의 평균함수 $m(t)$와 강도함수 $\lambda(t)$는 다음과 같다.

$$m(t) = \theta\left(1 - e^{-\beta x_n}\right),$$
$$\lambda(t) = \theta\beta e^{-\beta x_n}$$

두 함수가 적용된 Goel-Okumoto 모형의 우도함수는 다음과 같다.

$$L(\theta, \beta | D_{x_n}) = \left(\prod_{i=1}^{n} \theta\beta e^{-\beta x_i}\right) \exp\left\{-\theta\left(1 - e^{-\beta x_n}\right)\right\}$$

이 우도함수의 두 모수 θ와 β의 최우추정치 $\hat{\theta}$과 $\hat{\beta}$를 구하기 위하여 수치해석 방법의 하나인 Newton-Raphson방법을 이용하였다.

먼저 두 모수에 대한 편미분함수를 구하기 위해 우도함수에 로그를 취하면

$$\ln L(\theta, \beta | D_{x_n}) = n\ln\theta + n\ln\beta - \beta\sum_{i=1}^{n} x_i - \theta\left(1 - e^{-\beta x_n}\right)$$

와 같다.

로그우도함수를 이용하여 각 모수에 대한 편미분함수를 구하면

$$F(\theta, \beta) = \begin{pmatrix} \dfrac{\partial \ln L(\theta, \beta | D_{x_n})}{\partial \theta} \\ \dfrac{\partial \ln L(\theta, \beta | D_{x_n})}{\partial \beta} \end{pmatrix} = \begin{pmatrix} \dfrac{n}{\theta} - 1 + e^{-\beta x_n} \\ \dfrac{n}{\beta} - \sum_{i}^{n} x_i - \theta x_n e^{-\beta x_n} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

와 같은 벡터를 얻을 수 있으며, $F(\theta, \beta)$에 대한 자코비안 행렬을 구하면

$$J(\theta, \beta) = \begin{pmatrix} -\dfrac{n}{\theta^2} & -x_n e^{-\beta x_n} \\ -x_n e^{-\beta x_n} - \dfrac{n}{\beta^2} + \theta x_n e^{-\beta x_n} & \end{pmatrix}$$

와 같이 나타난다. 여기서 최우추정치 $\hat{\theta}, \hat{\beta}$에 대해 $\triangle\Theta = (\triangle\theta,\ \triangle\beta) = (\hat{\theta},\ \hat{\beta}) - (\theta_0,\ \beta_0)$라고 하면, $\triangle\Theta \approx -J(\theta_0,\ \beta_0)^{-1} F(\theta_0,\ \beta_0)$에 의해 Θ에 대한 근사

값 Θ_1은

$$\Theta_1 = \Theta_0 + \triangle\Theta = \Theta_0 - J(\theta_0,\ \beta_0)^{-1} F(\theta_0,\ \beta_0)$$

이 된다. 이 방법을 반복하면 Θ의 근사적인 추정치를 구할 수 있다.

다음으로 Rayleigh 분포를 적용한 NHPP 모형에 대하여 편미분함수를 구하기 위한 로그 우도함수는

$$\ln L(\theta,\beta|D_{x_n}) = n\ln 2 + n\ln\theta + \sum_{i=1}^{n}\ln x_i + n\ln\beta - \beta\sum_{i=1}^{n}x_i^2 - \theta\left(1 - e^{-\beta x_n^2}\right)$$

이고, 로그우도함수를 이용하여

$$F(\theta,\beta) = \begin{pmatrix} \dfrac{\partial \ln L(\theta,\beta|D_{x_n})}{\partial\theta} \\ \dfrac{\partial \ln L(\theta,\beta|D_{x_n})}{\partial\beta} \end{pmatrix} = \begin{pmatrix} \dfrac{n}{\theta} - 1 + e^{-\beta x_n^2} \\ \dfrac{n}{\beta} - \sum_{i}^{n} x^{2_i} - \theta x_n^2 e^{-\beta x_n^2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

를 유도할 수 있다. 이 $F(\theta,\beta)$에 대해 자코비안 행렬은

$$J(\theta,\beta) = \begin{pmatrix} -\dfrac{n}{\theta^2} & -x_n^2 e^{-\beta x_n^2} \\ -x_n^2 e^{-\beta x_n^2} & -\dfrac{n}{\beta^2} + \theta x_n^4 e^{-\beta x_n^2} \end{pmatrix}$$

와 같이 나타난다. 이 모형 또한 앞의 Goel-Okumoto 모형과 같이 $\hat{\theta},\hat{\beta}$에 대하여 추정할 수 있다.

4.6.2 공액 사전분포를 적용한 베이지안 추론

먼저 Goel-Okumoto 모형에 대하여 공액사전분포를 적용하고자 한다. 두 모수

θ와 β에 대하여 데이터 집합 D_t가 주어졌을 때, θ와 β에 대한 조건부 밀도함수를 구체화하기가 어렵기 때문에 모수 $\theta[1-F(t|\beta)]$를 가진 포아송분포를 따르는 잠재변수 $N'=N-n$를 도입한다. $\pi(\theta) \sim Gamma(a,b)$, $a,b>0$, $\pi(\beta) \sim Gamma(c,d)$, $c,d>0$, $\theta \perp \beta$이라고 하면 사후 결합 밀도는 다음과 같다.

$$L(\theta,\beta,N'|D_{x_n}) \propto L(D_{x_n}|\theta,\beta) \cdot \pi(\theta) \cdot \pi(\beta) \cdot \pi(N')$$

$$\propto \frac{(\theta e^{-\beta x_n})^{N'}}{N'!} \cdot e^{-\theta e^{-\beta x_n}} \cdot \theta^{n+a-1} \cdot \beta^{n+c-1} \cdot e^{-b\theta} \cdot e^{-(d+\sum_{i=1}^{n} x_i)\beta} \exp[-\theta(1-e^{-\beta x_n})]$$

$$= \frac{1}{N'!}\theta^{N'+n+a-1} \cdot \beta^{c+n-1} \cdot e^{-\theta(b+1)} e^{-\beta(d+N'x_n+\sum_{i=1}^{n} x_i)}$$

이 식을 장애모수 개념을 이용하여 깁스 샘플링 알고리즘에 사용되는 각 모수에 대한 사후 조건부 밀도는 다음과 같다.

$$N'|\theta,\beta,D_{x_n} \sim P(\theta e^{-\beta x_n}),$$

$$\theta|N',\beta,D_{x_n} \sim Gamma(N'+a+n,\ b+1),$$

$$\beta|N',\theta,D_{x_n} \sim \Gamma(n+c,\ d+N'x_n+\sum_{i=1}^{n} x_i)$$

다음으로 Rayleigh 모형에 대해서 Goel-Okumoto와 같이 잠재변수 N'를 도입하고 , $\pi(\theta) \sim Gamma(a,b)$, $a,b>0$, $\pi(\beta) \sim Gamma(c,d)$, $c,d>0$, $\theta \perp \beta$이라고 하면 사후 결합 밀도는 다음과 같다.

$$L(\theta,\beta,N'|D_{x_n}) \propto L(D_{x_n}|\theta,\beta) \cdot \pi(\theta) \cdot \pi(\beta) \cdot \pi(N')$$

$$\propto \frac{(\theta e^{-\beta x_n^2})^{N'}}{N'!} \cdot e^{-\theta e^{-\beta x_n^2}} \cdot \theta^{n+a-1} \cdot \beta^{n+c-1} \cdot e^{-b\theta} \cdot e^{-(d+\sum_{i=1}^{n} x_i^2)\beta} \exp[-\theta(1-e^{-\beta x_n^2})]$$

$$\cdot \prod_{i=1}^{n} x_i \propto \frac{1}{N'!}\theta^{N'+n+a-1} \cdot \beta^{c+n-1} \cdot e^{-\theta(b+1)} e^{-\beta(d+N'x_n^2+\sum_{i=1}^{n} x_i^2)} \cdot \prod_{i=1}^{n} x_i$$

위의 식 또한 Goel-Okumoto 모형에서와 같이 장애모수 개념을 이용하여 사후 조건부 밀도 함수를 구하면 다음과 같다.

$$N'|\theta,\beta,D_{x_n} \sim P(\theta e^{-\beta x_n^2}),$$

$$\theta|N',\beta,D_{x_n} \sim Gamma(N'+a+n,\ b+1),$$

$$\beta|N',\theta,D_{x_n} \sim \Gamma(n+c,\ d+N'x_n^2+\sum_{i=1}^{n}x_i^2)$$

4.6.3 균등분포를 적용한 무정보적 베이지안 추론

Goel-Okumoto 모형에 대하여 공액사전분포와 같이 조건부 밀도함수에 대해 구체화하기 위해 N'을 도입하고 각각의 모수에 $\pi(\theta) \sim U(a,b),\quad a,b>0,$ $\pi(\beta) \sim U(c,d),\ c,d>0,\ \theta \perp \beta$라고 하고 각 모수에 적용한다. 두 모수에 대하여 동시에 균등분포를 적용하는 것은 Goel-Okumoto 모형의 우도함수와 동일하므로 여기서는 생략한다.

먼저 모수 θ에 대하여 균등분포를 적용하고, β에 대하여 공액사전부포인 감마분포를 적용하여 나타난 밀도함수는 다음과 같다.

$$L(\theta,\beta,N'|D_{x_n}) \propto L(D_{x_n}|\theta,\beta)\cdot\pi(\theta)\cdot\pi(\beta)\cdot\pi(N')$$

$$\propto \frac{(\theta e^{-\beta x_n})^{N'}}{N'!}\cdot e^{-\theta e^{-\beta x_n}}\cdot\frac{1}{b-a}\cdot\theta^n\cdot\beta^{n+c-1}\cdot e^{-\theta}\cdot e^{-(d+\sum_{i=1}^{n}x_i)\beta}\exp[-\theta(1-e^{-\beta x_n})]$$

$$\propto \frac{1}{N'!}\theta^{N'+n}\cdot\beta^{c+n-1}\cdot e^{-\theta(b+1)}e^{-\beta(d+N'x_n+\sum_{i=1}^{n}x_i)}$$

이 식을 장애모수 개념을 이용하여 깁스 샘플링 알고리즘에 사용되는 각 모수에 대한 사후 조건부 밀도는 다음과 같다.

$$N'|\theta,\beta,D_{x_n} \sim P(\theta e^{-\beta x_n}),$$

$$\theta|N',\beta,D_{x_n} \sim Gamma(N'+n+1,\ 1),$$

$$\beta|N',\theta,D_{x_n} \sim \Gamma(n+c,\ d+N'x_n+\sum_{i=1}^{n}x_i)$$

다음으로 θ에 대하여 공액사전분포인 감마분포를 적용하고, β에 대하여 균등분포를 적용하여 나타난 밀도함수는 다음과 같다.

$$L(\theta,\beta,N'|D_{x_n}) \propto L(D_{x_n}|\theta,\beta)\cdot\pi(\theta)\cdot\pi(\beta)\cdot\pi(N')$$

$$\propto \frac{(\theta e^{-\beta x_n})^{N'}}{N'!}\cdot e^{-\theta e^{-\beta x_n}}\cdot\theta^{n+a-1}\cdot\beta^{n}\cdot e^{-b\theta}\cdot e^{-(\sum_{i=1}^{n}x_i)\beta}\exp[-\theta(1-e^{-\beta x_n})]$$

$$=\frac{1}{N'!}\theta^{N'+n+a-1}\cdot\beta^{c+n-1}\cdot e^{-\theta(b+1)}e^{-\beta(d+N'x_n+\sum_{i=1}^{n}x_i)}$$

이 식을 장애모수 개념을 이용하여 깁스 샘플링 알고리즘에 사용되는 각 모수에 대한 사후 조건부 밀도는 다음과 같다.

$$N'|\theta,\beta,D_{x_n} \sim P(\theta e^{-\beta x_n}),$$

$$\theta|N',\beta,D_{x_n} \sim Gamma(N'+a+n,\ b+1),$$

$$\beta|N',\theta,D_{x_n} \sim \Gamma(n+1,\ N'x_n+\sum_{i=1}^{n}x_i)$$

Rayleigh 모형 역시 N'을 도입하고 각각의 모수에 $\pi(\theta)\sim U(a,b),\ a,b>0$, $\pi(\beta)\sim U(c,d),\ c,d>0,\ \theta\perp\beta$라고 하고 각 모수에 적용한다. 두 모수에 대하여 동시에 균등분포를 적용하는 것은 Rayleigh 모형의 우도함수와 동일하므로 생략한다.

먼저 모수 θ에 대하여 균등분포를 적용하고, β에 대하여 공액사전부포인 감마분포를 적용하여 나타난 밀도함수는 다음과 같다.

$$L(\theta,\beta,N'|D_{x_n}) \propto L(D_{x_n}|\theta,\beta)\cdot\pi(\theta)\cdot\pi(\beta)\cdot\pi(N')$$

$$\propto \frac{(\theta e^{-\beta x_n^2})^{N'}}{N'!}\cdot e^{-\theta e^{-\beta x_n^2}}\cdot\frac{1}{b-a}\cdot\theta^{n}\cdot\beta^{n+c-1}\cdot e^{-\theta}\cdot e^{-(d+\sum_{i=1}^{n}x_i^2)\beta}\exp[-\theta(1-e^{-\beta x_n^2})]$$

$$\cdot \prod_{i=1}^{n} x_i \propto \frac{1}{N'!}\theta^{N'+n} \cdot \beta^{c+n-1} \cdot e^{-\theta(b+1)} e^{-\beta(d+N'x_n^2+\sum_{i=1}^{n} x_i^2)} \cdot \prod_{i=1}^{n} x_i$$

이 식을 장애모수 개념을 이용하여 깁스 샘플링 알고리즘에 사용되는 각 모수에 대한 사후 조건부 밀도는 다음과 같다.

$$N'|\theta,\beta,D_{x_n} \sim P(\theta e^{-\beta x_n}),$$

$$\theta|N',\beta,D_{x_n} \sim Gamma(N'+n+1,\ 1),$$

$$\beta|N',\theta,D_{x_n} \sim \Gamma(n+c,\ d+N'x_n^2+\sum_{i=1}^{n} x_i^2)$$

다음으로 θ에 대하여 공액사전분포인 감마분포를 적용하고, β에 대하여 균등분포를 적용하여 나타난 밀도함수는 다음과 같다.

$$L(\theta,\beta,N'|D_{x_n}) \propto L(D_{x_n}|\theta,\beta) \cdot \pi(\theta) \cdot \pi(\beta) \cdot \pi(N')$$

$$\propto \frac{(\theta e^{-\beta x_n^2})^{N'}}{N'!} \cdot e^{-\theta e^{-\beta x_n^2}} \cdot \theta^{n+a-1} \cdot \beta^{n} \cdot e^{-b\theta} \cdot e^{-(\sum_{i=1}^{n} x_i^2)\beta} \exp[-\theta(1-e^{-\beta x_n^2})] \cdot \prod_{i=1}^{n} x_i$$

$$= \frac{1}{N'!}\theta^{N'+n+a-1} \cdot \beta^{c+n-1} \cdot e^{-\theta(b+1)} e^{-\beta(d+N'x_{n+}\sum_{i=1}^{n} x_i)} \cdot \prod_{i=1}^{n} x_i$$

이 식을 장애모수 개념을 이용하여 깁스 샘플링 알고리즘에 사용되는 각 모수에 대한 사후 조건부 밀도는 다음과 같다.

$$N'|\theta,\beta,D_{x_n} \sim P(\theta e^{-\beta x_n}),$$

$$\theta|N',\beta,D_{x_n} \sim Gamma(N'+a+n,\ b+1),$$

$$\beta|N',\theta,D_{x_n} \sim \Gamma(n+1,\ N'x_n+\sum_{i=1}^{n} x_i)$$

5. 수치적 예제 및 결론

5.1. 수치적 예제

다음의 자료는 고장간격 시간에 대해 실제로 관측된 데이터(NTDS data, Goel and Okumoto, 1979)이다.

<표 6> 소프트웨어 고장 시간

고장 번호	고장 간격	총 고장 시간	고장 번호	고장 간격	총 고장 시간
1	9	9	14	9	87
2	12	21	15	4	91
3	11	32	16	1	92
4	4	36	17	3	95
5	7	43	18	3	98
6	2	45	19	6	104
7	5	50	20	1	105
8	8	58	21	11	116
9	5	63	22	33	149
10	7	70	23	7	156
11	1	71	24	91	247
12	6	77	25	2	249
13	1	78	26	1	250

<표 6>의 자료를 분석하기 위해 R 프로그램을 이용하였으며, 고전적 추론 방법, 베이지안 추론(공액 사전 분포, 무정보적 사전 분포)을 통해 모수 추정을 실시하였다.

두 모수 (θ,β)의 초기값을 (33,0.005)로 하고 고전적 추론 방법에서 허용오차를 $1.0\times e^{6}$으로 하여 직전값과 추정값의 오차가 허용오차보다 높게 나타날 때까지 반복하였고, 공액사전분포와 무정보적 사전분포를 적용한 베이지안 모형에서는 공액사전분포를 적용할 경우 $\theta,\beta\sim\Gamma(1,0.0001)$를, 무정보적 사전 분포를 적용할 경우 $\theta,\beta\sim U(0,1)$를 적용하여 각 모수에 대한 조건부 사후분포를 구한 후 깁스 샘플링

알고리즘을 이용하여 2000개의 (θ, β)에 대한 추정치를 얻어냈으며 수렴성을 확인하기 위해한회의 깁스샘플링 과정에서의 샘플에 대한 그래프를 제시하였다. 또한 각 모형에 대한 효율성을 비교하기 위하여 다음과 같은 상대오차의 합을 이용하고자 한다. m_i를 실제 i시간에서의 고장수라 하고 $\hat{m}_i$를 i시간에서 예측된 고장수라고 하면 상대오차 C_{mv}는 다음과 같다.

$$C_{mv} = \sum_{i=1}^{n} \left| \frac{m_i - \hat{m}_i}{m_i} \right| \quad \text{단, } i = 1, 2, 3, \cdots, n$$

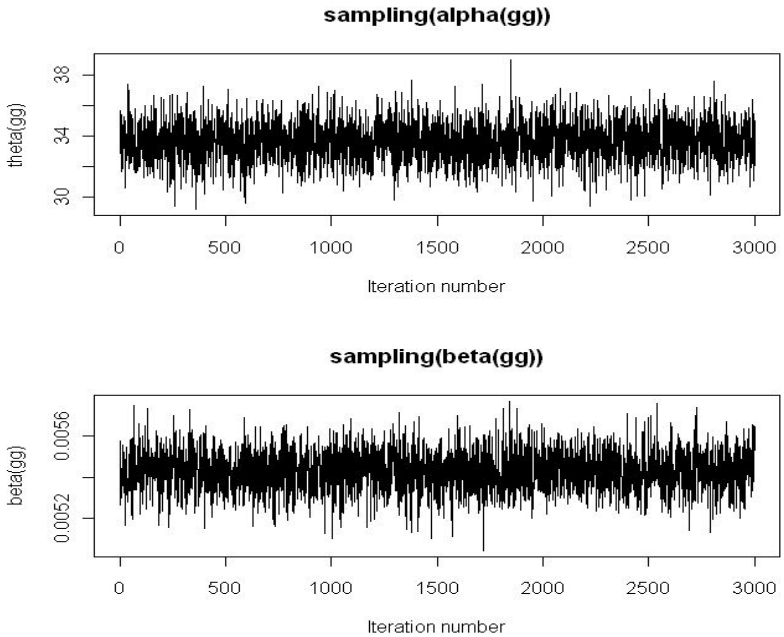
위의 <그림 4>는 NHPP모형에 대한 각 추론 방법의 그래프를 나타낸 것이다. 여기서 MLE는 고전적 추론 방법, GG는 공액사전분포, UG와 GU는 각각 한 모수 모수에 균등분포를 적용한 모형으로 각 방법에 대한 모수의 추정치는 다음과 같이 나타난다.

<표 7> θ에 대한 모수의 추정치

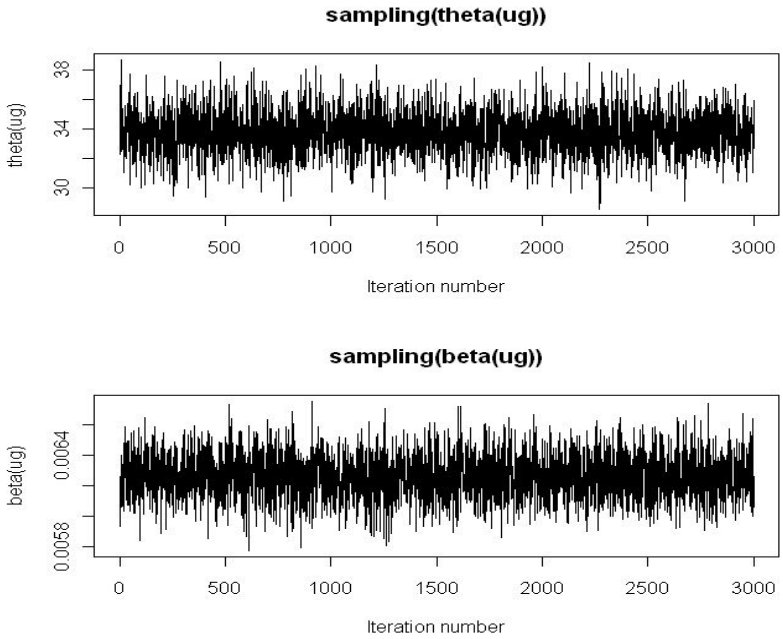
모형	$\hat{\theta}_{MLE}$	$\hat{\theta}_{GG}$	$\hat{\theta}_{UG}$	$\hat{\theta}_{GU}$
Goel-Okumoto	33.9932	33.6768	33.7797	33.8857
Rayleigh	27.3145	26.8168	26.3536	26.0557

<표 8> β에 대한 모수의 추정치

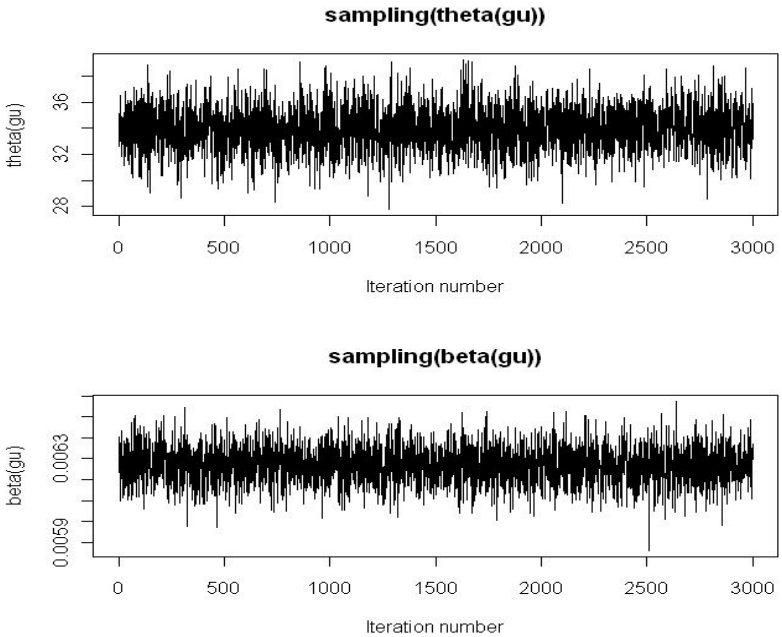
모형	$\hat{\beta}_{MLE}$	$\hat{\beta}_{GG}$	$\hat{\beta}_{UG}$	$\hat{\beta}_{GU}$
Goel-Okumoto	0.005790	0.005430	0.005423	0.005366
Rayleigh	0.000434	0.000424	0.000434	0.000438



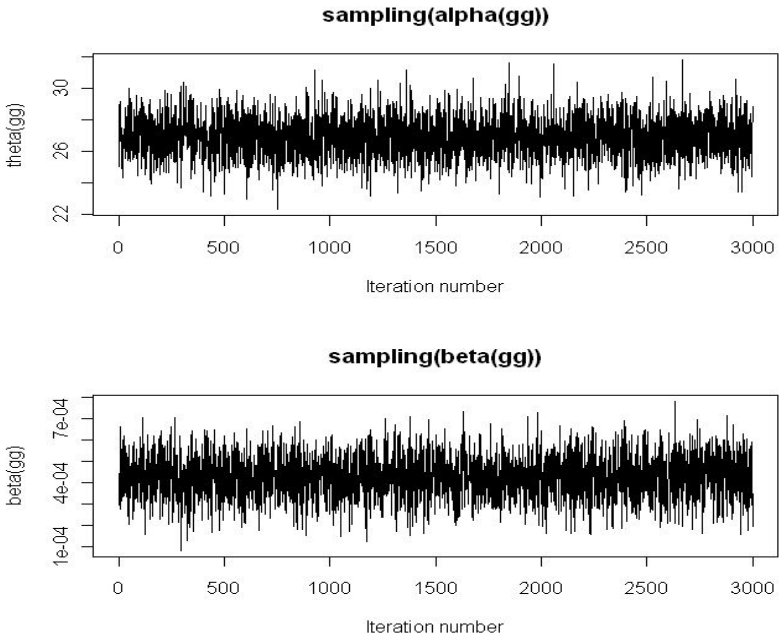
<그림 4> GO모델에서 사전분포 GG 적용 시의 수렴성



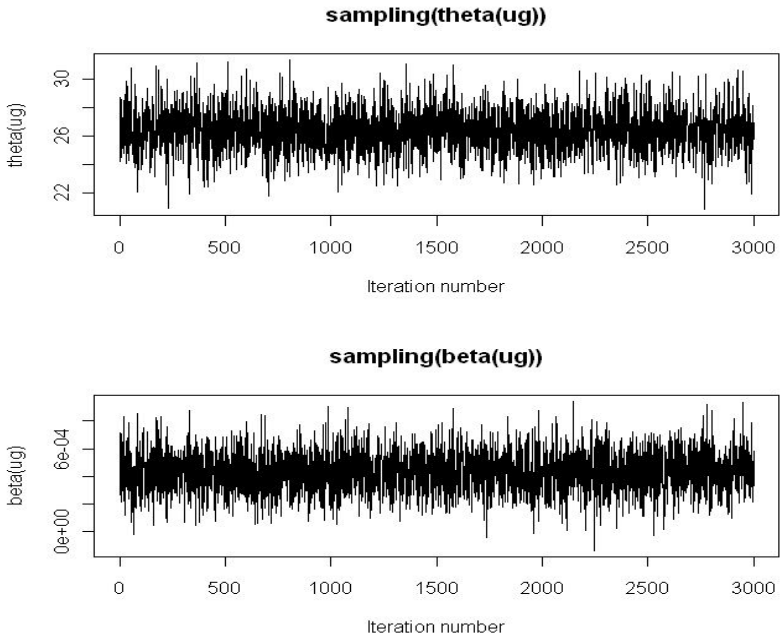
<그림 5> GO모델에서 사전분포 UG 적용 시의 수렴성



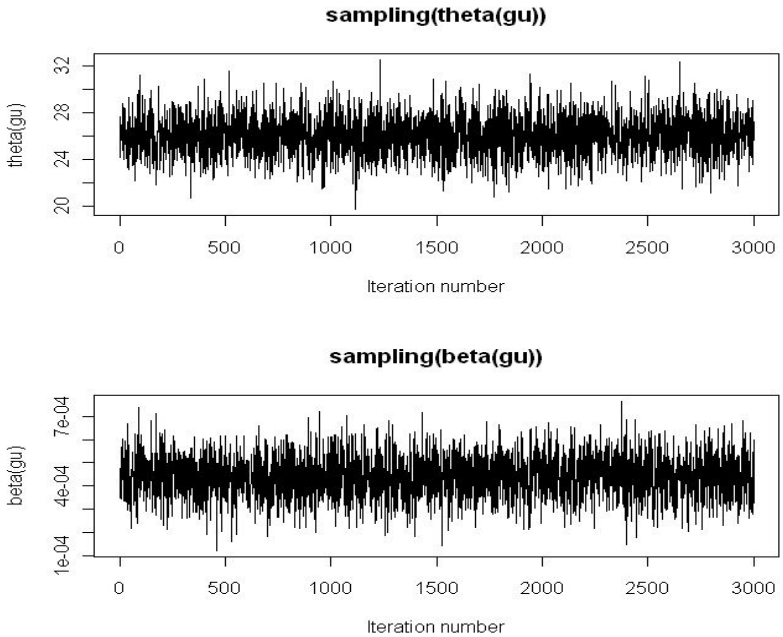
<그림 6> GO모델에서 사전분포 GU 적용 시의 수렴성



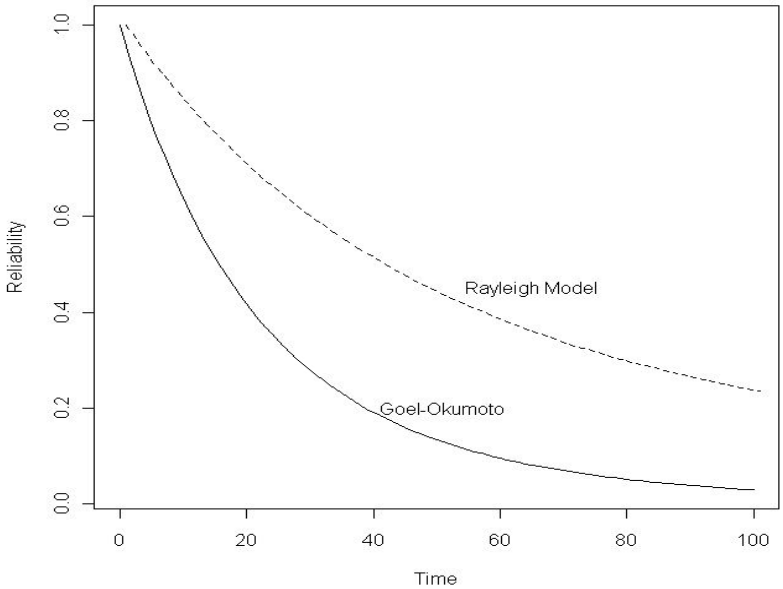
<그림 7> Rayleigh모델에서 사전분포 GG 적용 시의 수렴성



<그림 8> Rayleigh모델에서 사전분포 UG 적용 시의 수렴성



<그림> 9 Rayleigh모델에서 사전분포 GU 적용 시의 수렴성



<그림 10> 두 모형에 대한 신뢰도 그래프

<표 9> 상대오차에 의한 모형 비교

모형	MLE	GG	UG	GU
Goel-Okumoto	23.84558	22.58476	22.17881	21.79209
Rayleigh	6.353491	5.887581	5.890908	3.855959

<표 7>과 <표 8>의 각 모수에 대한 추정치를 보면 모든 방법에 대하여 추정치 $\hat{\theta}$와 $\hat{\beta}$가 유사하게 나타났다. <그림 4~9>는 각각의 사전분포가 정의된 상태에서 깁스샘플링을 실시하였을 때 나타나는 샘플들에 대해 Burn-in을 거쳐 생성된 샘플들이 특정 점에 대해 수렴하는 것을 나타낸 그래프이다. 그래프의 모양이 일정한 간격 안에 분포되어 있으므로 이 샘플링은 잘 생성된 샘플이라고 말 할 수 있다. <그림 10>은 Goel-Okumoto 모형과 Rayleigh 모형의 신뢰도 곡선에 대한 그래프이고, <표 9>는 각각의 모형에 대한 상대오차를 나타낸 표이다. 이 그래프와 표에서 나타난 바와 같이 Rayleigh 모형이 Goel-Okumoto 모형에 대해 효율적인 모형으로 나타났고, 사전분포를 적용하여 모수를 추론한 베이지안 방법이 고전적 방법보다 더 효율적인 것으로 나타났다.

5.2. 결론 및 제언

소프트웨어의 신뢰성은 개발의 최종단계에 있는 테스트 공정이나 사용단계에 있어서 소프트웨어 내에 존재하는 에러 수나 소프트웨어의 고장발생시간에 의해 효과적 평가를 할 수 있는 것으로 그 평가 기술이 중요하게 된다. 소프트웨어 개발의 테스트 공정이나 실제 사용단계에 있어서 에러 발생상황이나 소프트웨어 고장 발생현상을 수리적 모형화가 가능하다면 평가를 할 수 있다. 테스트의 개발 상황의 파악 테스트에 의해 미 발생되었던 고장에 대한 보수 코스트의 예측 등 구체적인 소프트웨어 개발의 보수관리문제에도 적용이 가능하다. 따라서 테스트 시간 혹은 실행시간과 발생된 고장 수나 소프트웨어 고장의 발생기간과의 관계를 소프트웨어 신뢰도 성장 과정이라고 볼 수 있다.

본 연구에서는 수리가 가능한 소프트웨어 고장수를 모형화하는데 비동질적 포아송 과정 형태를 이용하였으며, 기본적으로 널리 알려진 Goel-Okumoto 모형에서의 사전분포를 정의하여 베이지안 방법에 의해 모수를 추정하였다. 고전적인 추론 방법과 각각의 사전분포 적용에 대하여 효율적인 모형을 선택하기 위해 상대오차의 합을 계산하여 비교하였고, 비교한 결과에 대하여 고전적인 추론방법에 비교하여 베이지안 추론 방법이 상대적으로 좋은 모형임을 보였다.

하지만 라플라스에 의해 정의된 균등분포를 이용한 베이지안 추론은 특정 모수 θ에 대해 θ의 사전정보가 모호하면, θ의 변형형태의 사전정보 또한 동일하게 모호해야 하는데 주어진 라플라스의 사전분포는 이러한 불변성(invariance)이 결여되어있다는 한계점을 갖고 있다. 그러므로 이 불변성을 보완하기 위하여 제프리 사전분포, 준거사전분포, 확률일치 사전분포등이 연구되고 있다. 앞으로 다양한 사전분포를 정의하여 모형의 효율성을 높이는 연구가 필요한 것으로 생각된다.

참 고 문 헌

[1] A. Wood(1996), Software reliability growth models, Tandem Tech. Report

[2] Adrian E. Raftery(1987), Inference and Prediction for a General Order Statistic Model With Unknown Population Size, Journal of the American Statistical Association vol 82, 1163-1168

[3] B.H. Kim, I.H.Chang(2001), Bayesian estimation for the reliability in w-eibull stress-strength systems usning noninformative priors, Far East J. Theo. Stat. vol5, 2001

[4] B Littlewood, JL Verrall (1973), A Bayesian reliability growth model for computer software, Applied Statistics, 22 332-346

[5] Goel. Amrit L, Okumoto. Kazu(1979), Time-Dependent Error-Detection Rat-e Model for Software Reliability and Other Performance Measures, IEEE Transactions on Reliability, vol 28, 206-211

[6] H. Joe(1989), Statistical Inference for General-Order-Statistics and Nonhomogeneous-Poisson-Process Software Reliability Models, IEEE Transactions on Software Engineering archive vol 15, 1485 - 1490

[7] Hoang Pham,(2007), On Recent Generalizations of the Weibull Distribution, IEEE Transactions on Reliability vol 56, 454-458

[8] James O. Berger and Jose M. Bernardo(1989), Estimating a Product of Means: Bayesian Analysis with Reference Priors, Journal of the American Statistical Association vol 84, 200-207

[9] J.Albert(2007), Bayesian Computation with R, Springer

[10] KY Cai, DB Hu, CG Bai, H Hu, T Jing, Does software reliability growth behavior follow a non-homogeneous Poisson process, Information and Software Technology 50, 1232-1247

[11] L Kuo, TY Yang(1996), Bayesian Computation for Nonhomogeneous Poisson Processes in Software Reliability. Journal of the American Statistical Association 91 763-773

[12] L Kuo, JC Lee, K Choi, TY Yang (1997), Bayes inference for S-shaped so- ftware-reliability growth models, IEEE Transactions on Reliability 46, 76-80

[13] M. Barghout(2010) Predicting software reliability using an imperfect debugging Jelinski Moranda Non-homogeneous Poisson Process model vol5, 31-41

[14] M.S. Hamada, A.G.Wilson, C.S.Reese, H.F.Martz(2008), Bayesian Reliability, Springer

[15] RI Zequeira(2000), A for Bayesian software reliability analysis, Quality and Reliability Engineering 16, 187-193

[16] S. M. K. Quadri, N. Ahmad, Sheikh Umar Farooq, Software Reliability Growth modeling with Generalized Exponential testing - effort and optimal SOFTWARE RELEASE policy. Global Journal of Computer Science and Technology Vol.11 27-42

[17] Utkin LV. An uncertainty model of the stress-strength reliability with imprecise parameters of probability distributions. Zeitschrift fuer Angewandte Mathematik und Mechanik (Appl Math Mech) 2004;84:688 - 9

[18] Yamada. Shigeru, Ohba. Mitsuru, Osaki. Shunji(1983), S-Shaped Reliability Growth Modeling for Software Error Detection, Reliability, IEEE Transactions on, vol 32, 475 - 484

[19] Y. Hayakawa, T. Irony, M Xie(2001), System and Bayesian Reliability, World Scientific

[20] 간광현, 김희철, 이병수(2001), NHPP에 기초한 소프트웨어 신뢰도 모형에 대한 비교연구, 정보처리학회지 D8(4) 393-400

[21] 강기훈, 박홍선, 신기일, 신민웅, 정석오, 최대우(2005) 베이지안 통계학, 자유아카데미

[22] 김희철, 이동철(1999), 비동질적 포아송과정을 사용한 소프트웨어 베이지안 신뢰성 분석에 관한 연구, 공업경영학회지 22 23-31

[23] 김달호(2005), R과 WinBUGS를 이용한 베이지안 통계학, 자유아카데미