

회전기기에 대한 데이터 주도 고장진단 및 예측 모형 개발*

박지현¹ · 송명준² · 안선웅^{2†}

¹한국조선해양, ²한양대학교 산업경영공학과

Developing the Fault Diagnostics and Prognostics Model of a Rotating Machinery

Jihyun Park¹ · Myeongjoon Song² · Suneung Ahn²

¹Korea Shipbuilding & Offshore Engineering

²Department of Industrial and Management Engineering, Hanyang University

■ Abstract ■

With the advent of the fourth industrial revolution in recent years, engineering systems require prognostics and health management techniques that can accurately diagnose and prognose the state of the system. In this study, we developed a state-of-the-art fault diagnostics model using real-time vibration signals in the engineering systems. We established a degradation model that follows the Weibull hazard function and performed Bayesian estimation based on Markov chain Monte Carlo simulation to develop a fault diagnostics model. From the fault diagnostics model, we developed a data-driven fault prognostics model by monitoring the condition of the system continuously. In the experiment, this model is applied in the actual Intelligent Maintenance System bearing data provided by the University of Cincinnati. Compared with the previous methods, the results showed better performance.

Keywords : Fault Diagnostics, Condition based Maintenance(CBM), Markov Chain Monte Carlo (MCMC), Prognostics and Health Management(PHM), Rotating Machinery

논문접수일 : 2019년 11월 08일 논문게재확정일 : 2019년 12월 15일

논문수정일 : 2019년 12월 13일

* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2018R1A2B6003232).

† 교신저자, sunahn@hanyang.ac.kr

1. 서 론

공학 시스템(engineering system)은 시간이 지나면서 성능이 열화(degradation)되는 특성이 있기 때문에 열화 수준이 특정 임계 수준에 도달하게 되면 고장이 발생하는 특성이 있다. 또한, 최근 4차 산업혁명이 도래하면서 공학시스템은 첨단화, 대형화, 복잡화 추세가 나타나고 있으며, 이로 인해 수집되는 데이터의 양이 방대해지고 있다[3]. 최근의 공학 시스템은 복잡해지면서 시스템을 구성하는 모든 부품과 하위 시스템을 반영한 모델링은 불가능하다. 이에 최근 변화하는 시대환경에 따라, 시스템의 상태 데이터를 이용해 고장을 진단하고 예측하는 건전성 예측 및 관리(PHM : prognostics health management)에 대한 연구가 활발히 이루어지고 있다.

본 연구의 대상이 되는 시스템은 공학시스템의 회전기기로 이는 베어링, 전동기, 엔진 형태를 포괄하며 제조, 항공, 조선등 다양한 산업에 활용되는 부품이다[1]. 본 연구에서는 회전기기중 베어링에서 실시간으로 수집되는 진동신호를 이용하여 고장진단 모형을 개발하고, 지속적인 고장예측을 위한 데이터 주도형(data-driven) 고장예측 모형을 개발한다.

고장 진단 및 예측 모형의 개발은 건전성 예측 및 관리에 관한 연구의 중요 부분으로 건전성 예측 및 관리에 관한 포괄적인 선행연구를 소개한다. 건전성 예측 및 관리 기술은 운용중인 시스템의 상태를 관찰하여 열화상태를 진단해 향후 고장 및 사용 불능에 이르는 시점을 예측하는 기술이다. 이상 징후를 진단하고 예측하는 과정에서 진동신호에 대한 신호처리 기법과 시스템의 손상 또는 열화 정도를 정밀하게 반영할 수 있는 수리 모형을 개발하는 것이 중요하다.

건전성 예측 및 관리 분야는 물리 기반(physics based) 방법과 데이터 주도 방법으로 분류할 수 있다. 물리 기반 방법은 고장 물리 메커니즘(physics of failure mechanism)기반의 수리 모형을 이용하여 고장을 진단하고 예측하는 방법으로 기존에 확립된 고장 물리 모형을 사용한다. 물리 기반 방법은 수명에

측의 정확도가 비교적 높고 적은 양의 고장데이터로도 고장 진단이 가능한 장점이 있으나, 확립된 고장 물리 모형이 적어 새로운 시스템에는 적용이 어려운 단점이 있다. 반면, 데이터 주도 방법은 시스템의 신뢰성, 건전성 정보를 이용하여 확률 및 통계적 모형을 개발하여 고장을 예측한다. 이 방법은 특정 고장 물리 모형이 없이도 양질의 데이터만 확보된다면 시스템의 고장을 예측할 수 있는 장점이 있다. 최근 많은 양의 데이터를 수집할 수 있는 환경 때문에, 기계학습(machine learning)기법은 방대한 데이터를 학습모델에 적용하여 비교적 정확한 예측력을 보여주고 있으며 빠르게 발전하고 있는 연구 분야이다[20]. 이러한 환경 때문에 최근에는 물리 기반 방법보다는 데이터 주도 방법이 주로 연구되고 있는 추세이고, 특히 데이터 주도 방법에서 기계학습이나 회귀분석 기반 방법이 주로 연구되고 있다[7].

데이터 주도 방법은 시스템에 대한 고장 물리 모형이 없는 경우 수집된 건전성 데이터를 활용하여 학습 과정에서 수리 모형을 확립하는 방법이다. 데이터 주도 방법으로는 기계학습기반 방법과 회귀분석 기반 방법이 있으며, 기계학습기반 방법에는 대표적으로 인공신경망(artificial neural network)을 이용하여 고장을 예측 하는 방법이 있다. 이에 관한 연구로 Tse and Atherton[22], He et al.[7], Ali et al.[4], Wu et al.[23]이 있다. 회귀분석 기반 방법을 사용한 연구로는 Meeker and Escobar[17], Lu and Meeker[16], Xu and Zhou[24]이 있다. 또한 최근에는 베이지안 접근법을 기반으로 한 파티클 필터(particle filter)와 마코프체인 몬테카를로(Markov chain Monte Carlo) 시뮬레이션을 이용한 고장예측에 대한 연구도 진행되고 있다[2, 24].

본 연구의 목적은 회전기기인 베어링을 대상으로 진동신호를 이용한 고장진단 모형을 개발하고, 지속적인 고장예측을 위한 데이터주도 고장예측 모형을 개발하는 것이다. 본 연구의 순서로는 제 2장에서 진동신호를 이용한 고장진단 모형에 대해 소개한 후 제 3장에서 데이터 주도 고장예측 모형을 소개한다. 제 4장에서는 수립한 모델을 실제 베어링 시스템에

적용하여 타당성을 검증하며, 마지막 제 5장에서는 본 연구의 결론을 서술한다.

2. 고장진단 모형

2.1 특징신호 추출

공학 시스템은 시간이 지나면서 고장강도가 증가하는 열화(degradation) 현상을 갖는다[10]. 이때 시스템의 열화 정도를 나타내는 계측 변수가 특징신호이다. 센서로부터 수집된 진동 신호를 이용하여 특징신호로 변환하는 과정을 신호 처리(signal processing)라 하며, 특징신호는 대표적으로 시간 영역(time domain)과 주파수 영역(frequency domain)으로 분류할 수 있다. 본 연구에서 활용하는 특징신호로서 시간 영역을 기준으로 진동신호를 변환한 첨도(kurtosis)를 제안한다[13].

첨도는 진동 파형이 얼마나 충격적인가 혹은 뾰족한가를 나타내며, 시스템에 결함이 발생하였을 때 발생하는 충격진동이 첨도에 민감하게 반영되는 특성이 있기 때문에 건전성 예측 및 관리 분야에서 고장 예측에 유용하게 사용되는 특징신호이다. n 개의 신호 데이터 ϕ_i 와 평균 $\bar{\phi}$ 에 대한 첨도는 식 (1)과 같이 계산한다.

$$Kurtosis = \frac{\sum_{i=1}^n (\phi_i - \bar{\phi})^4}{n\sigma^4}. \quad (1)$$

첨도는 일반적으로 기계시스템의 작동 초기에 상승하며 시스템이 일정 수준 진행된 후 안정상태를 지속하며 결함에 다다를수록 증가하는 경향을 나타낸다. 이러한 첨도의 특성은 시스템의 결함을 진단하고 예측하는데 유용하게 쓰일 수 있다. 이러한 이유를 근거로 본 연구에서는 첨도를 이용한 고장진단 및 예측 모형을 개발한다.

2.2 열화모형

특징신호는 공학시스템의 상태를 대표해주는 지

수로, 이를 활용하여 시스템의 열화 정도를 나타내는 수리 모형인 열화모형을 수립할 수 있다. 본 연구에서 열화모형은 와이블 위험함수(Weibull hazard function)를 따른다고 가정한다. 와이블 위험함수는 고장 강도가 멱수법칙 고장모형(power law failure model)을 따르는 열화시스템에 적합한 특성이 있으며, 이때 고장률 함수는 와이블 분포(Weibull distribution)의 위험 함수 형태와 같다는 성질을 이용한다[9].

와이블 분포는 공학시스템의 신뢰성 및 위험분석 시 가장 널리 활용되는 확률분포로 형상(shape) 모수, 척도(scale) 모수, 위치(location) 모수에 따라 다양한 분포를 표현할 수 있어 수명 분석에 유용하게 쓰인다[18]. 그러므로 본 연구에서는 열화모형에 형상 모수와 척도 모수만 갖는 2모수 와이블 위험함수를 사용한다.

본 연구는 열화모형의 독립변수는 시간이기 때문에 시계열 분석에 해당하고, 시간에 따른 특징신호의 변화를 나타내는 수리식은 회귀(regression)식으로 표현할 수 있다. 시간의 변화에 따른 특징신호의 변화가 와이블 위험함수를 따라 변한다고 가정하면 식 (2)와 같다. 즉, 공학 시스템의 시간 경과에 따른 특징신호의 변화는 와이블 위험함수를 따른다.

$$Y = a \frac{\alpha}{\theta^\alpha} x^{\alpha-1} + b + \epsilon, \quad (2)$$

$$\epsilon \sim N(0, \sigma^2).$$

식 (2)에서 a 는 회귀식의 척도모수이고 b 는 y 절편 값을 의미하고, ϵ 는 회귀모형의 오차(error)를 의미한다. 또한, 와이블 위험함수의 α 는 형상모수를 의미하고, θ 는 척도모수를 의미한다.

2.3 베이지안 추정법

본 연구에서는 열화모형의 회귀식을 추정하기 위해 베이지안 추정법을 적용한다. 베이지안 추정법은 우도함수(likelihood function)와 사전분포(prior distribution)의 곱이 사후분포(posterior distribution)를

따르는 우도원리(likelihood principle)를 이용해 회귀식의 우도함수로 사후분포를 추론하는 방법이다. 베이저안 추정법의 우도원리와 갱신(updating) 특성을 이용하여 고장진단 및 예측 모형에 적용하면 열화모형에 적합도가 증가하게 되며 결론적으로 모형의 정확도가 증가하는 장점이 있다.

베이저안 추정법에서는 관측 횟수 i 에 대한 관측 시간이 $\mathbf{x} = \{x_1, x_2, \dots, x_i\}$ 일 때, 관측치 Y_i 는 식 (3), 식 (4)와 같이 정규분포로 표현할 수 있다. 그리고 확률적으로 독립인 관측치의 집합 $\mathbf{Y} = \{Y_1, Y_2, \dots, Y_n\}$ 에 대한 우도함수는 식 (5)와 같이 표현할 수 있고, 식 (6)과 같은 정규분포 형태로 표현한다.

$$Y_i | \theta, \alpha, a, b, \sigma^2 \sim N(Y_i, \sigma_i^2). \quad (3)$$

$$Y_i | \theta, \alpha, a, b, \sigma^2 \sim N\left(a \frac{\alpha}{\theta^\alpha} x_i^{\alpha-1} + b, \sigma^2\right). \quad (4)$$

$$L(\theta, \alpha, a, b, \sigma^2 | \mathbf{Y}) = p(Y_1, Y_2, \dots, Y_n | \theta, \alpha, a, b, \sigma^2) = \prod_{i=1}^n p(Y_i | \theta, \alpha, a, b, \sigma^2). \quad (5)$$

$$L(\theta, \alpha, a, b, \sigma^2 | \mathbf{Y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(Y_i - a \frac{\alpha}{\theta^\alpha} x_i^{\alpha-1} - b\right)^2\right\}. \quad (6)$$

$$L(\theta, \alpha, a, b, \sigma^2 | \mathbf{Y}) = \frac{L(\theta, \alpha, a, b, \sigma^2 | \mathbf{Y}) \pi(\theta, \alpha, a, b, \sigma^2)}{\int_{\theta} \int_{\alpha} \int_a \int_b \int_{\sigma^2} L(\theta, \alpha, a, b, \sigma^2 | \mathbf{Y}) \pi(\theta, \alpha, a, b, \sigma^2) d\sigma^2 db da d\alpha d\theta}. \quad (7)$$

식 (6)과 같은 열화모형에 대한 우도함수와 사전분포가 주어진다면, 베이즈 공식을 통해 사후분포를 도출할 수 있으며 사후분포에 대한 수학적 표현은 식 (7)과 같다.

본 연구에서는 베이저안 추정법을 적용하기 위해 마코프체인 몬테카를로 시뮬레이션을 적용한다. 베이저안 추정법에서 우도함수가 닫히지 않은 형태(no-closed form)를 나타내면 수치 계산을 통한 사후분포 도출이 어렵게 되는데, 이 경우에는 추정법(approximation)을 적용해야 한다. 본 연구에서는 추정법의 대표적인 방법으로 조건부 확률분포에서 표본을 추출하는 마코프체인 몬테카를로(MCMC : Markov chain Monte Carlo) 시뮬레이션을 사용하여

이 같은 문제를 해결한다[11]. 마코프체인 몬테카를로 시뮬레이션은 모수 추정 방법의 하나로 임의의 분포에서 표본을 반복해서 생성 및 추출하여 모수를 추정하는 원리이다. 마코프체인 몬테카를로 시뮬레이션을 기반으로 샘플링을 하면 다른 샘플링과 달리 마코프체인의 특성(Markov property)을 이용해 빠른 샘플링이 가능하다. 이러한 특징을 이용하여 열화모형에 필요한 모수를 도출할 수 있다. 마코프체인 몬테카를로 시뮬레이션에 대표적인 방법으로 임의의 제안 분포를 설정하여 생성하는 메트로폴리스-헤스팅스(M-H : Metropolis-Hastings) 샘플링과 미리 주어진 제안 분포에서 생성하는 깁스 샘플링(Gibbs sampling)이 있다.

메트로폴리스-헤스팅스 샘플링은 마코프 성질의 평형 방정식(equilibrium equation)을 따르는 제안 분포(proposal distribution)를 설정하여, 해당 제안 분포에서 표본을 추출하여 모수 추정하는 방법이다. 베이저안 추정법에서 모수를 추출할 제안 분포는 사전분포를 의미하며 사전분포에서 추출한 표본을 통해 사후분포를 추정한다. 메트로폴리스-헤스팅스 샘플링은 j 번째 실행 횟수일 때 다음과 같은 M-H 절차로 모수 θ 를 추론한다. 마찬가지로, M-H 절차로 나머지 모수 α, a, b, σ^2 를 추론할 수 있다.

Procedure of M-H method

Step 1. Set $j=1$

Step 2. Generates an initial value for $\theta_{(0)}$

Step 3. Repeat

A. Set the proposal distribution $q(\theta^* | \theta)$

B. Generates a proposal for θ^* from $q(\theta^* | \theta)$

C. Evaluates the acceptance probability

$$\rho(\theta, \theta^*) = \min\left(1, \frac{p(\theta^* | Y, \alpha, a, b, \sigma^2) q(\theta | \theta^*)}{p(\theta | Y, \alpha, a, b, \sigma^2) q(\theta^* | \theta)}\right).$$

D. Generates a u from the *uniform*(0,1)

E. If $u \leq \rho(\theta, \theta^*)$, accept the proposal and set $\theta = \theta^*$

F. Conduct the next M-H step in $j=j+1$

Step 4. Until $j=m$

본 연구에서는 메트로폴리스 해스팅스 샘플링을 이용하여 모수 채택률이 0.8 이상인 값을 표본으로 추출한다. 마코프체인 몬테카를로 시뮬레이션을 통해 추출한 표본의 극한 분포에 대해 대수의 법칙(laws of large number)에 의해 모수 추정치를 도출할 수 있다. 이 과정에서 수렴성을 고려해야 하는데 이를 수렴성 진단이라 한다. 수렴성 진단은 부록에서 다룬다.

베이시안 추정법의 정확도를 높이기 위해서 적절한 사전분포를 사용하여야 한다. 본 연구에서는 무정보적 사전분포를 적용한다. 일양(uniform)분포와 같은 모호한 사전분포(vague prior)나 정규분포 같은 대칭적(symmetric) 분포를 설정하는 것이 추정의 정확도를 높일 수 있으므로 본 연구에서는 일양분포를 사전분포로 사용한다.

적절한 사전분포와 초기값 설정은 마코프체인 몬테카를로 시뮬레이션의 수렴(convergence) 가능성을 높일 수 있다[6]. 특히, 본 연구 모형과 같은 다차원의 열화모형에서는 적절한 초기값을 선택하는 것이 중요하다. 사전분포의 상위모수(hyper-parameter)에 대한 초기값은 최대 사후확률(maximum a posterior)방법을 적용하며 이는 사후확률이 최대가 되는 최적의 모수를 찾는 방법이다. 베이지 공식에 의해 사후확률은 우도함수와 사전분포의 곱에 비례하여 표현할 수 있다. 모수 집합 $\theta = \{\theta, \alpha, a, b, \sigma^2\}$ 에 대한 사전분포 $\pi(\theta)$ 와 관측치 \mathbf{Y} 에 대한 우도함수 $p(\mathbf{Y}|\theta)$ 를 통해 사후분포 $f(\theta|\mathbf{Y})$ 가 최대가 되는 값에 대한 최대사후확률은 식 (8)과 같이 표현할 수 있다.

$$\begin{aligned}\hat{\theta} &= \arg \max_{\theta} f(\theta|\mathbf{Y}) \\ &= \arg \max_{\theta} p(\mathbf{Y}|\theta)\pi(\theta).\end{aligned}\quad (8)$$

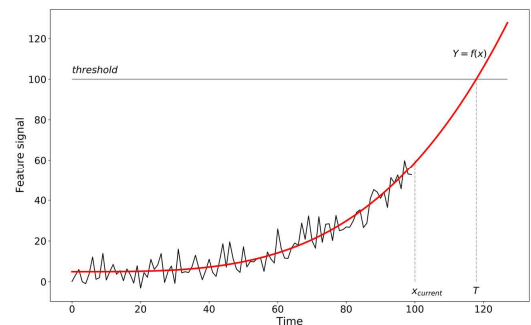
식 (8)에서 사전분포와 우도함수의 곱을 로그(log) 미분 후 0이 될 때의 모수 추정치가 최적값으로 도출된다. 그러나 일반적으로 로그 미분한 곱은 닫힌 형태가 아니기 때문에 최적값을 구하기 위한 최적화 알고리즘을 적용해야 한다. 본 연구에서는 최적화 알고리즘으로 준뉴턴(Quasi-Newton) 방법을 적용하여 도출되는 최적값을 마코프체인 몬테카를로 시뮬레이션의 초기값으로 사용한다.

3. 고장예측 모형

3.1 수명예측

공학시스템의 신뢰성 평가 실험을 위해서는 수명 시험을 통해 얻은 완전 고장(run-to-failure) 데이터를 주로 사용하지만, 이러한 완전 고장 데이터를 얻는 데에는 많은 어려움이 있다. 그러므로 최근 데이터가 방대해지는 환경을 통해 건전성 예측 및 관리 분야에서는 완전 고장 데이터가 아닌 실시간 관측 데이터를 이용하여 시스템의 신뢰성을 평가하고 고장을 예측한다. 시간이 지날수록 시스템의 성능이 저하되는 공학시스템의 특성을 분석하여, 본 연구에서는 시스템의 성능이 임계 수준에 도달할 경우 고장으로 판정하고 이에 따른 시스템의 고장 시간을 예측할 수 있다. 위와 같은 방법으로 도출한 고장 시간은 시스템의 수명 예측치를 의미한다.

[그림 1]을 통해 열화모형 $Y=f(x)$ 가 시간이 지남에 따라 시스템이 열화되며 임계수준에서 고장이 발생하고, 그 때 시간이 부품의 수명임을 확인할 수 있다. 또한, 잔존수명(residual lifetime)은 예측수명 T 에서 현재 시간 $X_{current}$ 을 뺀 값을 의미한다. 예를 들어, [그림 1]에서 100기까지 관측된 데이터를 이용해 열화모형이 임계 수준인 100기에 도달할 때 부품의 수명은 118.45로 추정한다. 잔존수명은 $118.45 - 100 = 18.45$ 이 도출된다.



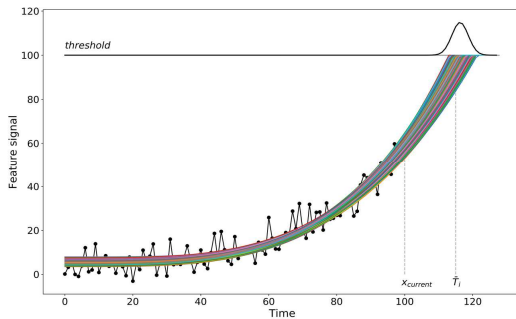
[그림 1] 특정 임계수준에 대한 열화모형

열화모형에 대해 베이지안 추정법을 적용하면 예측 수명이 분포 형태로 도출된다. 고장 수준을 나타내는 임계 수준은 y_{th} 로 표현하고 전체 m 번의 마코프체인 몬테카를로 시뮬레이션 중 j 번째 시뮬레이션에서 추정된 모수를 이용해 도출한 예측 수명은 식 (9)와 같다. 특정 시점 x_i 에서 열화모형을 통해 추정된 수명은 독립된 동일한 분포(independent identically distribution)를 따르는 표본이 추출되고, 중심 극한 정리(central limit theorem)에 따라 표본추출 원리로 인해 정규분포의 형태를 따른다.

$$x_{i,j} = \left(\left(\frac{y_{th} - b_{(j)}}{a_{(j)}} \right) \times \frac{\theta_{(j)}^{\alpha_{(j)}}}{\alpha_{(j)}} \right)^{\frac{1}{\alpha_{(j)} - 1}}. \quad (9)$$

특정 시점 x_i 에 m 번의 마코프체인 몬테카를로 시뮬레이션을 통해 도출한 예측 수명은 $T_i = \{t_{i1}, t_{i2}, \dots, t_{im}\}$ 이고, 예측 수명들에 대한 수리식은 식 (10)과 같다. 그리고 예측 수명은 [그림 2]와 같은 정규분포의 형태를 따르게 된다. 즉, 특정 시점 x_i 에 T_i 에 대한 함수는 $T_i = g(t_i)$ 로 표현되고 정규분포를 따른다.

$$T_i = \infty \{x : f(x|\theta, \alpha, a, b, \sigma^2) > y_{th}\}. \quad (10)$$

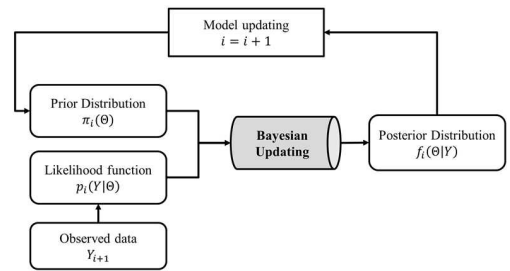


[그림 2] 특정 시점에서 예측 수명 분포

예측 수명의 분포인 정규분포의 평균은 시스템의 예측 수명을 의미한다. 기존 연구에서는 예측 수명에서 현재 관측 시점의 차이를 잔존수명이라 하며, 기존의 연구들은 잔존 수명 추정에 초점을 맞추어 연구되었다[7].

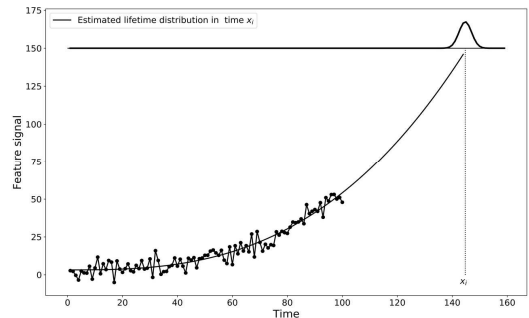
3.2 실시간 고장예측

본 연구에서는 수명 예측 방법을 이용해 데이터 주도형 실시간 고장예측 프레임워크를 개발한다. 시간이 지나며 데이터를 추가 관측하면서, 베이지안 갱신 과정을 통해 사후분포를 갱신해 나간다. 이 과정을 통해 학습과 예측을 반복하여 실시간 고장예측을 수행할 수 있다[4]. [그림 3]에서는 추정하고자 하는 모수 집합 $\theta = \{\theta, \alpha, a, b, \sigma^2\}$ 에 대한 관측치 Y 의 베이지안 갱신 과정이 나타난다.

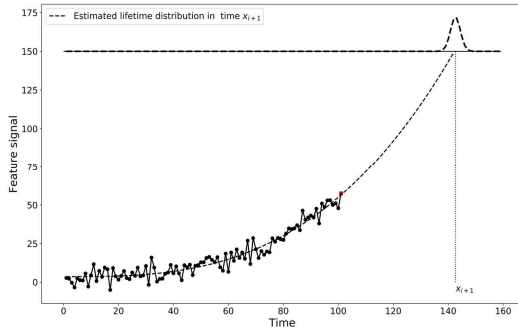


[그림 3] 데이터 주도 베이지안 갱신과정

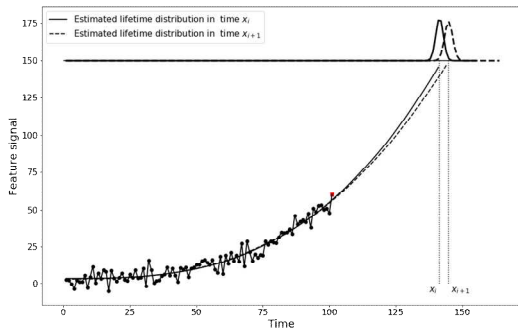
본 연구에서는 데이터가 추가로 관측되면서 마코프체인 몬테카를로 시뮬레이션 기반 베이지안 갱신 과정을 통해 실시간으로 고장에 대한 예측 수명에 대한 확률 분포가 갱신된다. 본 연구에서는 마코프체인 몬테카를로 시뮬레이션 방법으로 메트로폴리스 해스팅스 샘플링을 적용한다. [그림 4]는 특정시점 x_i 에 관측한 데이터를 통해 추정한 열화모형의 예측수명 분포이다. 이때, [그림 5]와 같이 x_{i+1} 시점에 추가 관측데이터가 들어오면 열화모형이 갱신되면서 새로운 예측수명 분포가 생성된다.



[그림 4] 특정 시점 x_i 에 예측 수명분포

[그림 5] 시점 x_{i+1} 에 예측 수명분포

종합적으로 [그림 6]과 같이 데이터가 추가 관측됨에 따라 새로운 예측 수명 분포가 갱신되는 것을 확인할 수 있다. 이와 같은 순서로 데이터 주도 방법을 통해 고장예측 모형이 갱신됨을 알 수 있다.



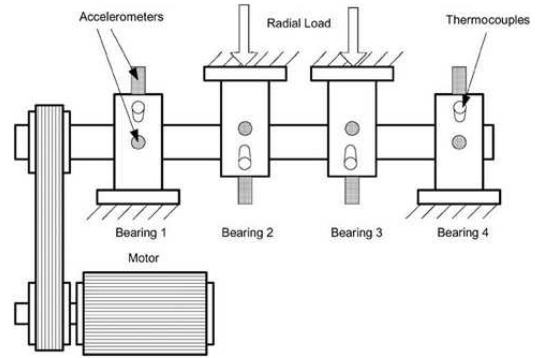
[그림 6] 예측 수명분포의 갱신에 대한 예시

4. 실험

4.1 적용 시스템 소개

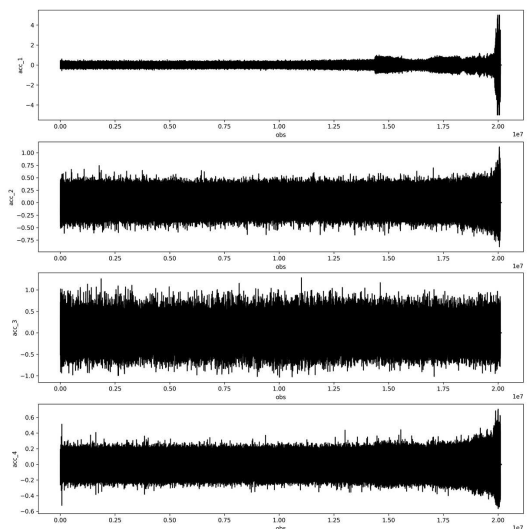
본 연구에서는 신시내티 대학교에서 제공하는 IMS (intelligent maintenance system) 베어링 결함 데이터를 대상으로 고장진단 및 예측 모형을 적용한다 [16]. IMS데이터는 NASA Ames prognostics data repository에서 구할 수 있다[12]. [그림 7]과 같이 Rexnord에서 제작한 ZA-2115모델 네 개의 베어링에 수평, 수직 방향으로 가속도계가 장착되어 있다. 진동 신호는 샘플링 주파수 20kHz로 10분 간격으로 1초간 측정되었고, 1주기(cycle)당 진동 신호 관측치는 2,048개이다. 총 관측 주기는 984주기이므

로 총 관측치는 2,015,232개이다. 총 측정 기간은 2004년 2월 12일부터 19일까지 8일로 고장이 발생할 때까지 측정하였다.



[그림 7] 베어링 구조 소개

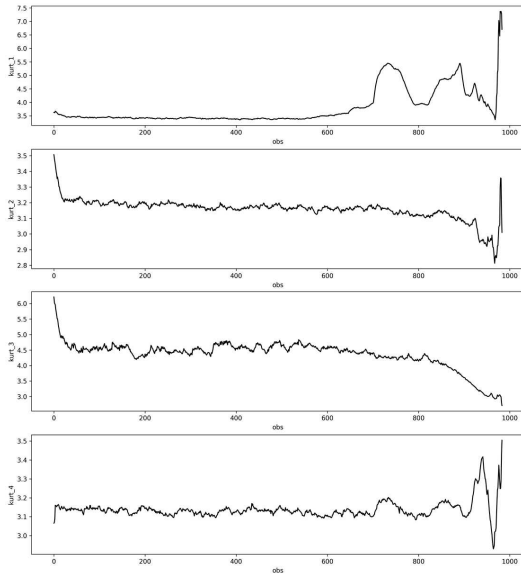
네 개의 베어링의 수직방향의 가속도계에서 측정 한 가속도 값은 [그림 8]과 같다. 데이터는 4개의 베어링중 하나의 베어링이 고장나면 고장으로 간주되고, 그 시점이 전체 시스템의 고장 시점이 된다. 그러므로 데이터는 하나의 베어링이 고장 날 때(run-to-failure)까지 실험한 것으로 나머지 세 개는 중도 절단(censored) 데이터가 된다. 기존 연구에서는 네 개의 베어링에 대한 잔존 수명 예측을 위한 고장진단 및 예측 모형을 개발하였다[16, 21, 22].



[그림 8] 네 개의 베어링에 대한 가속도

4.2 고장진단 및 예측 모형

IMS에서 제공하는 네 개의 베어링에 대하여 가속도 값을 특징신호로 변환한다. 특징신호로 침도를 사용하고, 네 개의 베어링에 대한 침도를 도출하면 [그림 9]와 같다.



[그림 9] 네 개의 베어링에 대한 침도

열화모형에 대한 우도함수를 식 (5)와 같이 표현하고, 각 모수에 대한 사전분포로 무정보적(noninformative) 사전분포로서 정규(normal)분포로 가정하면 식 (11), 식 (12), 식 (13), 식 (14), 식 (15)와 같다.

$$\pi(\theta) \sim \text{Normal}(\hat{\mu}_\theta, \sigma_\theta^2). \quad (11)$$

$$\pi(\alpha) \sim \text{Normal}(\hat{\mu}_\alpha, \sigma_\alpha^2). \quad (12)$$

$$\pi(a) \sim \text{Normal}(\hat{\mu}_a, \sigma_a^2). \quad (13)$$

$$\pi(b) \sim \text{Normal}(\hat{\mu}_b, \sigma_b^2). \quad (14)$$

$$\pi(\sigma) \sim \text{HalfNormal}(\tau). \quad (15)$$

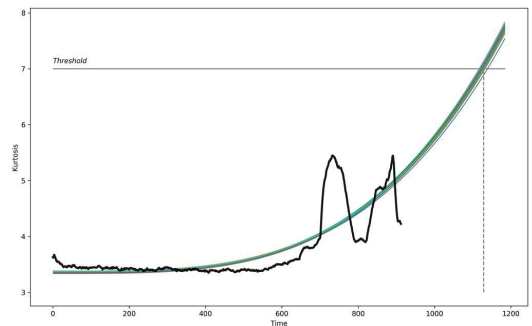
정규분포를 사전분포로 하고 마코프체인 몬테카를로 시뮬레이션을 시행하려면 각 사전분포의 상위 모수인 평균과 표준편차를 설정해야 한다. 마코프체인 몬테카를로 시뮬레이션의 수렴성을 높이기 위해 적절한 상위모수를 설정해야 하므로 최소자승 회귀

분석을 통해 도출한 모수 추정치를 각 사전분포에 대한 정규분포의 평균으로 설정한다. 또한, 정규분포의 분산은 계층적(hierarchical) 베이저안 추정법으로 분산의 범위를 일양분포로 설정하여 모수추정의 정확성을 높인다. 계층적 베이저안 접근을 적용하면 본 연구와 같은 모수가 많은 모형에서 과대적합(overfitting)문제를 해결할 수 있다[8]. 즉, 정규분포를 따르는 모수 $\theta = \{\theta, \alpha, a, b, \sigma^2\}$ 의 분산은 일양(uniform)분포로 식 (16)과 같이 설정한다.

$$\sigma_\theta \sim \text{Uniform}(a_\theta, b_\theta). \quad (16)$$

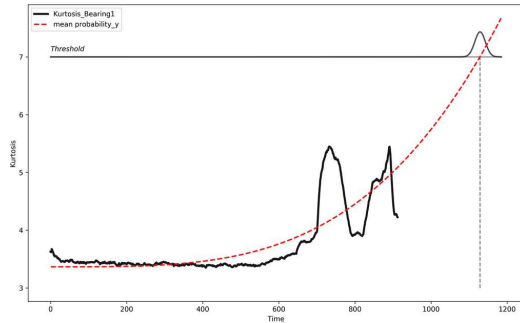
식 (5)에 열화모형에 대한 우도함수와 모수에 대한 사전분포를 베이즈 공식에 대입하면 식 (7)을 통해 사후분포 $f(\theta, \alpha, a, b, \sigma^2 | \mathbf{Y})$ 를 도출할 수 있다.

마코프체인 몬테카를로 시뮬레이션기반 베이저안 추정법을 적용하여 열화모형의 각 모수를 추정한다. 실험에서는 많은 관측치로 인한 분석 소프트웨어에 속도와 부하 문제가 발생할 수 있다. 그러므로 본 연구에서는 CUDA기반 GPU 병렬 프로세싱을 통해 분석 소프트웨어의 속도를 높였다. 사전분포의 초기값은 최대 사후확률 방법을 통해 도출한다. 5,000번의 마코프체인 몬테카를로 시뮬레이션을 시행하면 5,000개의 적합 회귀선을 추정할 수 있다. 또한, 마코프체인 몬테카를로 시뮬레이션의 수렴가능성을 높이기 위해 번인(burn-in) 기간을 전체 실행 횟수의 10%인 500번으로 설정한다. 그러므로 베어링 1에 대해 마코프체인 몬테카를로 시뮬레이션을 시행하면 [그림 10]과 같이 도출된다.



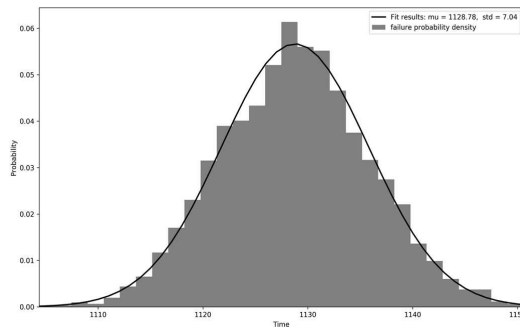
[그림 10] 베어링 1에 대한 MCMC 수행

기준에 주어진 IMS 베어링에 대한 임계수준과 동일하게 임계수준을 7로 설정하면, 베이지안 회귀 모형을 통해 예측 수명 값들을 도출할 수 있다. IMS 데이터 자료를 살펴보면 베어링 설계된 임계 수준은 7이라 하였다[7]. 임계 수준 7에 대한 베이지안 추정법을 적용한 열화모형은 [그림 11]과 같다.



[그림 11] 베어링 1에 대한 열화모형과 예측 수명

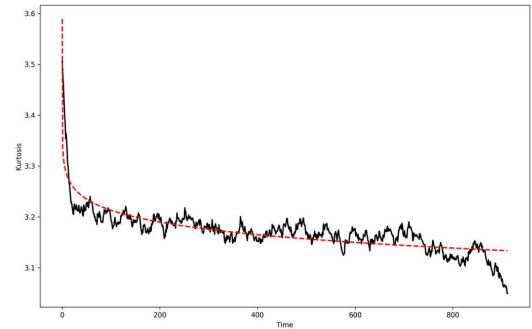
[그림 11]에서 나타나는 예측 수명 값들에 대한 확률 분포는 [그림 12]에 나타난다. 5,000개의 예측 수명에 대한 확률분포는 중심극한 정리에 따라 정규분포를 따르게 된다.



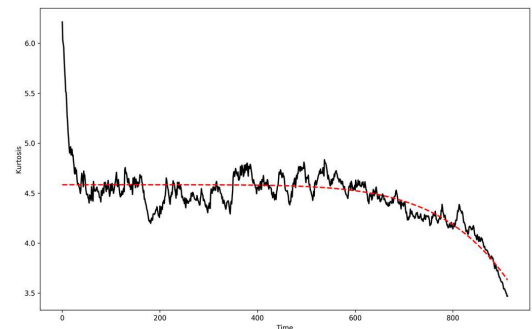
[그림 12] 베어링 1에 대한 예측 수명 분포

베어링 1의 예측 수명은 약 188시간(1128.78 관측주기)로 나타난다. 현재 관측 시점이 160시간(964 관측주기)임을 가정하면 약 28시간의 잔존수명을 갖고 있는 것으로 예측된다. 또한, [그림 13], [그림 14], [그림 15]를 살펴보면, 베어링 2, 3, 4는 열화에 진입하지 않은 것을 살펴볼 수 있다. 그러므로 베어링 1이 먼저 고장날 가능성이 높다. 실제

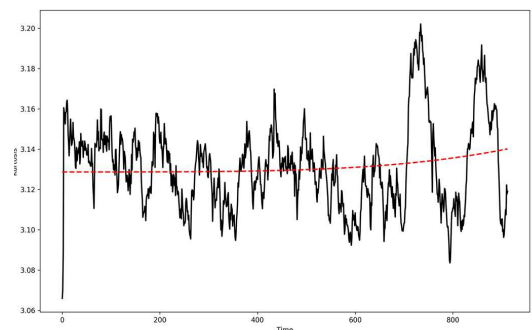
실험결과를 살펴보면 베어링 1이 외륜(outer race) 고장으로 인해 베어링 시스템이 멈춘 것으로 나타난다[16].



[그림 13] 베어링 2에 대한 열화모형



[그림 14] 베어링 3에 대한 열화모형



[그림 15] 베어링 4에 대한 열화모형

4.3 실험 비교분석

본 연구의 실험결과 검증을 위해 기존 연구와 비교 분석한다. 베어링 1에 대한 잔존 수명을 예측한 연구로 Ali et al.[4], Wu et al.[23] 연구가 있다. Ali et al.

[4]는 SFAM(simplified Fuzzy adaptive resonance theory map) 인공신경망을 이용하여 데이터 주도 예측 모델을 개발하였고, 이를 이용해서 수명을 예측하고 그에 따른 잔존 수명을 예측하였다. Wu et al.[23] 연구에서는 SFB(sensitive frequency band)라는 새로운 특징신호를 통해 RBF(radial basis function)를 이용한 인공신경망을 개발하여 잔존 수명을 예측하였다. 베어링 1에 대해 전체 984주기 중 964주기 때 절단(truncated)하여 학습데이터로 활용하였고, 임계 수준 40에 대한 고장 예측시간과 그에 따른 잔존 수명 및 예측 정확도를 계산하였다. 본 연구에서도 같은 실험 환경인 964주기 때 동일 절단 후 잔존 수명을 예측할 수 있다. 비교 결과는 <표 1>과 같이 도출된다.

<표 1> 잔존수명 추정의 정확도 비교분석

| | Ali et al. [4] | Wu et al. [23] | Proposed model |
|-----|-------------------|-------------------|-------------------|
| 정확도 | 78.3% | 80.4 | 84.6% |

예측 수명을 통한 예측 정확도를 계산한 결과 84.6%가 도출되었다. <표 1>의 정확도 비교를 살펴보면, 본 연구 모형이 약 84.6%로 가장 뛰어난 성능을 나타내었다.

5. 결 론

기술의 발전과 함께 공학시스템은 더욱 다양한 부품들로 구성되고 있으며, 많은 양의 데이터를 수집할 수 있는 환경이 조성되었다. 그러므로 복잡한 시스템의 경우 물리적 모델의 수립이 사실상 불가능하므로 상태를 나타내는 데이터를 활용하여 시스템의 상태를 예측하는 연구가 활발히 이루어지고 있다. 그러므로 본 연구에서는 회전기기중 베어링에서 실시간으로 수집되는 진동신호를 이용한 데이터 주도형 고장진단 모형을 개발하고, 지속적인 고장예측을 위한 데이터 주도 고장예측 모형을 개발하였다.

본 연구에서는 고장진단 모형을 개발하기 위한 특징신호로 첨도(kurtosis)를 사용하였다. 첨도를 이용

하여 시스템의 열화정도를 나타내는 수리모형인 열화모형을 정립하였다. 또한, 고장진단 모형에서 열화모형으로는 와이블 위험함수(Weibull hazard function)를 사용하였고, 마코프체인 몬테카를로(Markov chain Monte Carlo) 시뮬레이션을 기반으로 베이지안 추정법을 적용하여 베어링의 수명을 예측하였다.

고장진단 모형을 바탕으로 데이터 주도 고장예측 모형을 개발하였다. 본 연구에서는 시스템의 성능이 임계 수준에 도달할 경우 고장으로 판명되고, 이에 따른 시스템의 수명을 예측할 수 있는 원리를 이용한다. 본 연구의 대상인 베어링의 수명 예측에는 마코프체인 몬테카를로 시뮬레이션 기반 베이지안 추정법을 이용하였고, 수명은 중심극한정리를 통해 정규분포를 따르게 된다. 지속적인 데이터 관측을 이용한 베이지안 갱신 과정으로 사후부포를 도출하여 예측 수명이 갱신하는 실시간 고장예측 프레임워크를 개발하였다.

신시내티 대학교에서 제공하는 IMS 베어링 데이터를 이용하여 고장진단 및 예측 모형을 검증하였다. 네 개의 베어링으로 구성된 시스템을 고장진단 및 예측 모형을 적용한 결과 베어링 1에 고장이 먼저 발생하는 것으로 예측되었고, 실제로 베어링 1이 먼저 고장이 발생하는 것을 확인하였다. 또한, 베어링 1에 대한 잔존수명을 예측한 Ali et al.[4], Wu et al.[23]의 연구와 비교한 결과 본 연구 모형이 뛰어난 성능을 나타내었다.

본 연구에서 개발한 고장진단 및 예측 모형은 미래 고장을 예측하는 실시간 모니터링(monitoring) 시스템에 활용될 수 있다. 또한, 이와 같은 고장진단 모형은 시스템의 상태 기반 정비(condition based maintenance)로 정비 및 운용비용 절감을 실현할 수 있다. 본 연구는 연구의 대상인 베어링뿐만 아니라 전동기, 엔진 등 대형 회전기기에도 적용할 수 있다. 또한, 최근 대두되고 있는 빅데이터(big data) 환경에서 데이터를 분석하면서 분석 시간 및 수렴 문제가 발생할 수 있다. 그러므로 GPU기반 병렬 프로세싱 환경이나 전처리 방법에 대한 추가적인 연구가 필요하다.

참 고 문 헌

- [1] 이상우, 송지은, 배석주(2019), “건전성 지표 기반의 베어링 이상 사전 예측 모형에 관한 연구,” 한국경영과학회 춘계공동학술대회 논문집, 광주, 2019. 4. 10, 2709-2727.
- [2] 전치혁, 양희중, 장수영, 정의승, 임혜란(1993), “베이지안기법에 의한 임무 신뢰도 예측,” 한국경영과학회지, 18(1), 71-78.
- [3] 주강진, 이민화, 양희진, 류두진(2017), “4차 산업혁명과 인공지능-현황, 사례, 구제에 대한 개괄적 고찰,” 한국경영과학회지, 42(4), 1-14.
- [4] Ali, J.B., Chebel-Morello, B., Saidi, L., Malinowski, S., and Fnaiech, F.(2015), “Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network,” Mechanical Systems and Signal Processing, 56-57, 150-172.
- [5] An, D., Choi, J.H., and Kim, N.H.(2013), “Prognostics 101 : A tutorial for particle filter-based prognostics algorithm using Matlab,” Reliability Engineering and System Safety, 115, 161-169.
- [6] Andrieu, C., De Freitas, N., Doucet, A., and Jordan, M.I.(2003), “An introduction to MCMC for machine learning,” Machine learning, 50(1-2), 5-43.
- [7] He, Y., Tan, Y., and Sun, Y.(2004), “Wavelet neural network approach for fault diagnosis of analogue circuits,” IEEE Proceedings-Circuits, Devices and Systems.
- [8] Johnson, W., Branscum, A., Hanson, T.E., and Christensen, R.(2010), Bayesian ideas and data analysis : an introduction for scientists and statisticians, CRC Press.
- [9] Karklin, Y., and Lewicki, M.S.(2005), “A hierarchical Bayesian model for learning non-linear statistical regularities in nonstationary natural signals,” Neural Computation, 17(2), 397-423.
- [10] Kim, T., Park, C., and Ahn, S.(2008), “Determining the optimal maintenance action for a deteriorating repairable system,” Probabilistic Engineering Mechanics, 23(1), 95-101.
- [11] Kim, W., and Ahn, S.(2009), “Determining the periodic maintenance interval for guaranteeing the availability of a system with a linearly increasing hazard rate : industry application,” International Journal of Industrial Engineering : Theory, Applications and Practice, 16(2), 126-134.
- [12] Kuo, L., and Yang, T.Y.(1996), “Bayesian computation for nonhomogeneous Poisson processes in software reliability,” Journal of the American Statistical Association, 91(434), 763-773.
- [13] Lee, J., Qiu, H., Yu, G., and Lin, J.(2015), Bearing data set, University of Cincinnati, IMS, NASA Ames : Moffett Field, CA.
- [14] Lei, Y., and Zuo, M.J.(2009), “Gear crack level identification based on weighted K nearest neighbor classification algorithm,” Mechanical Systems and Signal Processing, 23(5), 1535-1547.
- [15] Lei, Y.G., Li, N.P., Guo, L., Li, N.B., Yan, T., and Lin, J.(2018), “Machinery health prognostics : A systematic review from data acquisition to RUL prediction,” Mechanical Systems and Signal Processing, 104, 799-834.
- [16] Lu, C.J., and Meeker, W.Q.(1993), “Using degradation measures to estimate a time-to-failure distribution,” Technometrics, 35(2), 161-174.
- [17] Meeker, W.Q., and Escobar, L.A.(2014), Statistical methods for reliability data, John

- Wiley & Sons.
- [18] Patil, A., Huard, D., and Fonnesbeck, C.J. (2010), "PyMC : Bayesian stochastic modeling in python," *Journal of Statistical Software*, 35(4), 1-81.
- [19] Qiu, H., Lee, J., Lin, J., and Yu, G.(2006), "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of Sound and Vibration*, 289(4-5), 1066-1090.
- [20] Rinne, H.(2008), *The Weibull distribution : a handbook*, Chapman and Hall/CRC.
- [21] Si, X.S., Wang, W., Hu, C.H., and Zhou, D.H. (2011), "Remaining useful life estimation-a review on the statistical data driven approaches," *European Journal of Operational Research*, 213(1), 1-14.
- [22] Tse, P.W., and Atherton, D.P.(1999), "Prediction of machine deterioration using vibration based fault trends and recurrent neural networks," *Journal of Vibration and Acoustics*, 121(3), 355-362.
- [23] Wu, B., Li, W., and Qiu, M.Q.(2017), "Remaining useful life prediction of bearing with vibration signals based on a novel indicator," *Shock and Vibration*.
- [24] Xu, Z., and Zhou, D.H.(2006), "A degradation measurements based real-time reliability prediction method," *IFAC Proceedings Volumes*, 39(13), 950-955.

〈부록〉수렴성 진단

마코프체인 몬테카를로 시뮬레이션을 통한 표본 추출에서는 마코프체인(Markov chain)의 정상분포(stationary distribution)가 우리가 원하는 사후분포로 수렴하는지 여부를 확인해야 한다. 만약 수렴하지 않았다면 수렴할 때까지 반복횟수를 늘리거나 사전분포를 재설정하는 작업이 필요하다. 그러나 반복횟수를 늘리게 되면 분석시간이 증가하고 수렴이 느려지게 되는 문제를 일으킨다. 본 장에서 수렴성의 진단(convergence diagnostics)은 수치적 접근인 겔만 루빈(Gelman-Rubin) 통계량과 경험적 접근인 트레이스 그림(trace plot)을 통해 확인한다.

겔만 루빈 통계량은 두 개 이상의 체인(chains)을 갖고 수렴성 진단을 하는 것으로 각 체인의 분산과 체인 간 분산을 비교하는데 기초를 두고 있다. 표본추출값은 $p = (1, \dots, h)$ 의 체인에서 $j = (1, \dots, m)$ 번째 반복 실행 값을 의미한다. 체인 간 분산(B)은 식 (A1)과 같다.

$$B = \frac{m}{k-1} \sum_{p=1}^h (\bar{y}_p - \bar{y})^2. \quad (A1)$$

식 (A1)에서 \bar{y}_p 과 \bar{y} 는 식 (A2)와 (A3)을 의미한다.

$$\bar{y}_p = \frac{1}{c} \sum_{j=1}^h y_{pj}. \quad (A2)$$

$$\bar{y} = \frac{1}{c} \sum_{p=1}^h \bar{y}_p. \quad (A3)$$

이를 통해 체인 내 분산을 구할 수 있다. 체인 내 분산(W)은 식 (A4)와 같이 표현할 수 있고, s_p^2 는 식 (A5)와 같다. 식 (A5)는 d 번째 체인에서 표본 분산을 의미한다.

$$W = \frac{1}{c} \sum_{p=1}^h s_p^2. \quad (A4)$$

$$s_p^2 = \frac{1}{m-1} \sum_{j=1}^m (y_{pj} - \bar{y}_p)^2. \quad (A5)$$

전체 표본값 y 에 대한 분산은 체인 간의 분산과 체인 내 분산으로 식 (A6)과 같이 표현할 수 있다.

$$\widehat{var}(y) = \frac{m}{m-1} W + \frac{1}{m} B. \quad (A6)$$

최종적으로 수렴성 진단에서 구하고자 하는 기준 측정치 \hat{R} 는 식 (A7)과 같다. 기준 측정치는 잠재적 척도축소인자(potential scale reduction factor)라고 한다.

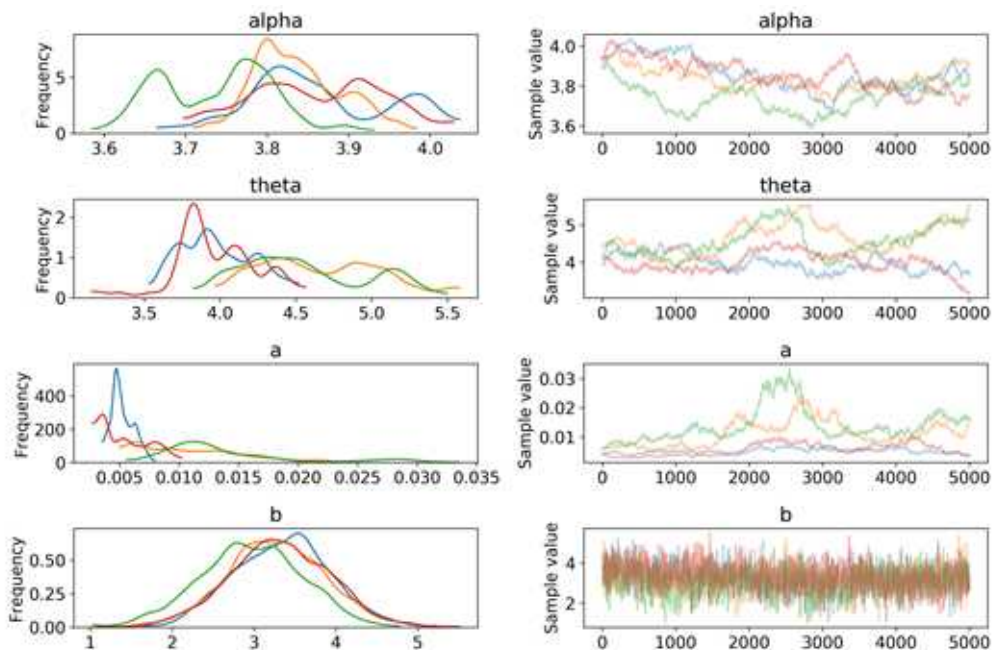
$$\sqrt{\hat{R}} = \sqrt{\frac{\widehat{var}(y)}{W}}. \quad (A6)$$

마코프체인 몬테카를로 시뮬레이션이 진행됨에 따라 \hat{R} 이 1값에 가까워지면 표본 추출된 y 가 동일한 분포로 수렴해가서 적절한 실험환경임을 의미한다. 본 연구에서 사용한 Python소프트웨어에서는 1.4 이하를 안정적 수렴이라 한다[19]. 본 연구에서 제 2장, 제 3장에서 예시로 보여준 데이터를 이용하여 수렴성 진단을 시행하면 <부록 표 1>과 같다.

<부록 표 1> 모수에 대한 수렴성 진단

| | mean | Rhat |
|-------|--------|--------|
| alpha | 4.5119 | 1.2227 |
| a | 0.0017 | 1.2881 |
| b | 5.5496 | 1.0505 |
| theta | 4.9157 | 1.2882 |
| sigma | 5.2177 | 1.0000 |

<표 2>를 살펴보면 모든 모수에서 \hat{R} 값이 1.4 이하로 나타난다. 이러한 결과는 본 연구의 마코프체인 몬테카를로 시뮬레이션이 통계적으로 유의미한 실험임을 의미한다. 마찬가지로 트레이스 그림은 [부록 그림 1]과 같다.



[부록 그림 1] MCMC 시뮬레이션의 트레이스 그림

총 4개의 체인에 대한 수렴 정도를 살펴보면 모든 모수에서 비교적 잘 수렴하는 경향이 나타난다. 이러한 결과는 본 연구에서 개발한 고장예측 모형이 타당함을 뒷받침할 수 있는 근거가 될 수 있다. 그러므로 고장 예측 모형을 적용하는 과정에서 수렴성 진단 여부를 고려한 모형의 개발이 필요하다.