

고장 보고율을 이용한 현장 수명자료 분포의 모수추정

박태웅·김영복·이창훈

서울대학교 산업공학과

E-mail : tw5860@hanmail.net

Estimating parameters of the lifetime distribution is investigated when field failure data are not completely reported. To take into account the reality and the accuracy of the estimates in such a case, the failure reporting probability is incorporated in estimating parameters. Firstly, method of maximum likelihood estimate(MLE) is used to estimate parameters of the lifetime distribution when failure reporting probability is known. Secondly, Expectation and Maximization(EM) algorithm is used to estimate the failure reporting probability and parameters of the lifetime distribution simultaneously when failure

reporting probability is unknown. For both case, procedures of estimation are illustrated for single Weibull distribution and mixed Weibull distribution. Simulation results show that MLE obtained by the proposed method is more accurate than the conventional MLE.

Keywords : field failure data, failure reporting probability, MLE, EM algorithm, single Weibull distribution, mixed Weibull distribution

1. 서 론

현장시스템(field system)으로부터 얻어지는 현장고장자료(field failure data)는 실제 사용량과 환경스트레스를 절대적으로 설명하기 때문에 신뢰도를 추정하기 위한 자료로써 매우 바람직하다. 그러나 현장고장자료는 시간과 비용이 많이 소요되고, 고장자료 획득의 어려움 그리고 사용현장에서 얻어지는 자료의 중요성에 대한 인식부족으로 인하여 부품의 수명과 관련된 자료들은 실험실에서 수행되는 수명시험(life testing)과 가속수명시험(accelerated life testing)을 통해 얻어지는 것이 보통이다. 이들 자료는 통제된 환경에서의 정밀한 결과를 제공하지만 시험시간(testing time)을 단축하기 위하여 실제 사용되는 조건보다 더 가혹한 조건에서 수행되는 것이 보통이고, 부품이 설치된 사용 환경을 모두 반영하여 실험하는 것은 한계가 있기 때문에 사용현장에서의 부품에 대한 정보를 왜곡하여 나타낼 수 있는 위험이 있다. 따라서 부품의 수명에 대한 올바른 정보를 얻기 위하여 실제 사용현장에서 얻어진 현장고장자료를 분석하는 방법과 연구는 매우 중요하다.

현장고장자료는 미리 정해진 분석시점(time for analysis) 동안 부품에 고장이 발생했을 때 수리자(repair man) 및 보수자(maintenance personnel)는 해당 시스템을 찾게 되고, 고장보고를 통해서 부품 고장이 자료수집시스템으로 보고되며 고장 보고로부터 고장시간(time to failure), 고장원인(cause of failure) 그리고 관측기간 동안 고장이 나지 않은 부품으로 구성된다. 이를 이용하여 부품의 수명분포(life distribution) 모수(parameter)에 대한 최우추정값(MLE : Maximum Likelihood Estimate)이 구하여 진다. 그러나 현장고장자료는 모든 부품의 고장이 완벽하게 추적되어 기록되지 않고, 고장 시간 및 고장원인과 같은 상세한 보고가 누락이 되며, 부품의 고장보고가 이루어지더라도 부정확하게 기록되는 단점을 내포하고 있다. 이러한 단점 항목 중 본 논문에서 중점이 되는 사항은 부품 고장의 보고 누락이다. 고장보고의 누락 관점에서 수집된 자료에는 다양한 자료 완벽성(data completeness) 정도가 존재한다. 몇몇 자료는 완벽한 기록을 가졌고, 설치 날짜부터 현재 시간까지 작동특성에 대한 통계를 보여 줄 수 있다. 매우 종종 완벽하게 추적되는 유일한 부품은 시스템을 구성하는 부품들 중 중요도(importance)와 비용(cost)이 높은 부품으로써 예를 들면 쿨링타워(cooling tower), 보일러(boiler) 등이다. 그러나 상대적으로 중요도 및 비용이 낮은

밸브(valve)와 필터(filter)와 같은 부품들은 보통 기록되지 않는다[Hale2001]. 추가적으로 보수자는 그들의 주요한 책임은 시스템을 유지하는 것이지 자료를 기록하기 위한 것이 아니라고 이해한다[Coit99]. 즉, 고장이 발생한 부품이지만 수리자에 의해 보고가 되지 않으면, 그 부품에 대한 정확한 수명분포 모수를 추정하기 위한 정보를 얻을 수 없다는 것이다.

부품의 현장고장자료의 분석에 관한 기존연구로써 Coit는 매 고장시마다 각각의 고장시간 기록이 누락되고 단지 부품의 전체 고장수와 누적 시간자료가 가용할 때 자료를 분석하는 방법을 제시하였다[Coit99], [Coit2000]. 그리고 보증기간(warranty time) 동안 수집된 불완전한 보증자료를 바탕으로 신뢰도를 추정하는 비모수방법을 제안한 연구[Lim2002]와 보증기간 후에 추가적으로 수집된 자료를 추가하여 현장보증자료를 분석하였다[Oh2001]. 실제로 현장에 설치된 부품의 고장은 여러 원인에 의해 고장이 발생한다. 이들 부품의 수명분포를 모델링하기 위하여 혼합 와이불분포(mixed Weibull distribution)는 부품 혹은 시스템의 고장이 하나의 고장모드(failure mode)보다 더 많은 고장모드에 의해 원인이 될 때 전기적 혹은 기계적 부품의 수명에 대해 훌륭한 모델을 제공한다[Jiang92]. 이에 대한 고장자료의 원인별 수명을 추정하기 위한 연구로써 Jiang 과 Murthy는 와이불 확률지(WPP : Weibull Plotting Paper)의 관점에서 2-혼합 와이불분포를 특성화한 그래프 제도 방법(graphical plotting approach) [Jiang95]과 2-혼합 와이불 분포에 대해서 모수 값에 따른 8개의 고장률함수(failure rate function)의 형태를 제시했다 [Jiang98]. Jiang과 Kececioglu는 EM 알고리즘에 기초해 MLE를 따르는 새로운 알고리즘을 연구를 했고, 이는 사후자료(postmortem data)와 사전자료(Non-postmortem data)의 수명자료에 대해 분석을 했다[Jiang92].

본 연구에서는 주어진 부품에 대하여 관측기간 동안 첫 번째 고장에 대한 현장고장자료에 대해 중점을 두고 부품에 대한 고장 모드가 단일한 경우와 여럿인 경우에 고장 원인별 고장보고가 누락되는 현실성을 반영하기 위하여 고장 원인별 고장 보고율(failure reporting probability)과 혼합분포(mixture distribution)로 모델화하는 모수 추정(parameter estimation)절차를 제안한다. 이때 고장 원인별 수명분포는 와이불분포(Weibull distribution)로 가정한다. 와이불분포는 척도모수(scale parameter) α 와 형태모수(shape parameter) β 의 다양한 값에 대해 도달하는 많은 모양 때문에 신뢰성 공학

(Reliability engineering)에서 가장 일반적으로 사용되는 분포 중의 하나이고 매우 다양한 자료와 수명특성을 모델화 할 수 있다[Bucar2004]. 이를 위하여 2장에서는 고장 보고율이 알려진 경우 주어진 현장고장자료와 고장 보고율에 근거한 우도함수(likelihood function)를 유도하여 최우추정법(method of maximum likelihood estimate)에 기초한 모수 추정 절차를 제시하고, 3장에서는 고장 보고율이 알려지지 않은 경우 주어진 현장고장자료와 EM(expectation and Maximization) 알고리즘을 사용하여 새로운 우도함수를 유도하고 동시에 고장 보고율과 수명분포의 모수를 추정하는 절차를 제시한다. 4장에서는 실험예제를 통해 고장 보고율을 고려한 모수추정 방법에 대한 타당성을 검증하고 비교 분석한다. 5장에서는 결론 및 추후연구 방향을 제시한다.

2. 고장 보고율이 알려진 경우

본 장에서는 고장 원인별 고장 보고율이 알려진 경우에 대한 최우추정량(MLE : Maximization Likelihood Estimator)을 구하는 방법론 개발을 목적으로 한다. 고장 보고율이 알려진 경우는 공학적 판단이 요구되는데 고장 보고율은 현장에 설치된 시스템 및 부품에 관여하는 숙련된 설비관리자, 수리자 그리고 보수자의 공정일지, 기억 그리고 의견을 통해 추정될 수 있다. 이 처럼 엔지니어의 판단은 종종 수명분포의 모수를 결정하기 위해서 사용된다.

이처럼 고장 원인별 고장 보고율이 알려진 경우에 대해서 수집된 현장고장자료와 고장 원인별 고장 보고율을 사용하여 부품의 수명분포가 특히 와이블분포일 때 MLE를 추정하기 위한 유도 절차를 제시하고자 한다.

<가 정>

현장고장자료를 이용시 영향을 미치는 변수는 다양하기 때문에 분석에 앞서서 자료의 형태와 조사 방법에 따른 가정은 다음과 같다.

- ① 부품의 고장은 첫 번째 고장만을 고려한다.
- ② 부품의 고장시간은 모두 달려시간이다.
- ③ 부품에는 m 개의 고장원인이 있으며, 고장 원인별 수명은 서로 독립이다.
- ④ 분석 시점 안에 고장 발생한 부품의 고장 원인별 고장 보고율은 $P_j, j=1,2,\dots,m$ 이다.
- ⑤ 보고된 부품의 고장시간은 확률 1 로 정확하게 관측된다.

<기 호 정 의>

본 논문에서 사용되는 기호는 다음과 같다.

- N : 총 부품 설치개수
- T : 부품 수명의 분석시점
- $f_j(t; \alpha_j, \beta_j)$: 원인 j 에 의한 부품의 와이블 확률밀도함수, $j=1,2,\dots,m$
- $F_j(t; \alpha_j, \beta_j)$: 원인 j 에 의한 부품의 와이블 누적분포함수, $j=1,2,\dots,m$
- $R_j(t; \alpha_j, \beta_j)$: 원인 j 에 의한 부품의 와이블 신뢰성함수, $j=1,2,\dots,m$
- P_j : 기간 $(0, T]$ 동안 원인 j 에 의한 고장 보고율, $P_j < 1, j=1,2,\dots,m$
- π_j : 원인 j 에 의한 고장비율, $j=1,2,\dots,m$, 단, $\pi_1 + \dots + \pi_m = 1$
- D_j : 원인 j 에 의해 관측기간 $(0, T]$ 동안 고장난 후, 보고된 부품의 집합
- D_{NR} : 기간 $(0, T]$ 동안 보고되지 않은 부품의 집합
- D_f : 기간 $(0, T]$ 동안 고장 났으나, 보고되지 않은 부품의

집합

- D_s : 기간 $(0, T]$ 동안 고장 나지 않은 부품의 집합
- n_j : 집합 D_j 에 속한 부품의 수, $j=1,2,\dots,m, NR, f, s$

2.1 현장고장자료의 MLEs 추정 절차

현장에 설치된 부품이 하나의 고장 모드로 발생할 수 있으나 실제로 대부분의 고장은 여러 가지 고장 모드에 의해 고장이 발생한다. 이는 부품의 고장원인별 수명분포가 와이블 분포 일 때 혼합 와이블분포로 모델화 할 수 있다. 분석 시점 $(0, T]$ 안에 고장 발생한 부품이 자료수집 시스템에 보고되는 알려진 원인별 고장 보고율은 $P_j, j=1,2,\dots,m$ 이고 각 원인별 고장 비율 $\pi_j, j=1,2,\dots,m$ 은 이미 알려져 있다고 가정하며 $\pi_1 + \pi_2 + \dots + \pi_m = 1$ 이다. 알려지지 않은 고장 보고율 $P_j, j=1,2,\dots,m$ 는 다음 장에서 고려된다. 만일 부품의 고장모드가 하나인 경우 수명분포는 단일 와이블분포이고 고장 비율은 $\pi_1 = 1$ 이다. 우도함수는 다음과 같이 유도되어 구성될 수 있다.

(i) 분석 시점 $(0, T]$ 안에 고장 발생한 부품은 고장 원인별 고장 보고율 $P_j, j=1,2,\dots,m$ 로 보고되고, 고장자료는 집합 $D_j, j=1,2,\dots,m$ 에 속하며 우도(likelihood)에 각각 $P_j \pi_j f_j(t; \alpha_j, \beta_j), j=1,2,\dots,m$ 로 기여한다.

(ii) T 시점까지 보고되지 않은 부품은 분석 시점 $(0, T]$ 안에서 고장 발생했으나, $1 - P_j, j=1,2,\dots,m$ 로 보고되지 않았거나 혹은 분석 시점 T 까지 고장 발생하지 않은 부품들이다. 따라서 T 시점까지 보고되지 않은 자료는 집합 D_{NR} 에 속하고 이들은 우도에 $(1 - P_1)\pi_1 F_1(T; \alpha_1, \beta_1) + \dots + (1 - P_m)\pi_m F_m(T; \alpha_m, \beta_m) + \pi_1 R_1(T; \alpha_1, \beta_1) + \dots + \pi_m R_m(T; \alpha_m, \beta_m)$ 로 기여한다.

따라서 고장 원인별 보고율과 고장비율이 알려지고, 수집된 고장자료로부터 유도된 우도함수는 다음과 같이 식 (1)로 나타난다.

$$L(\bar{\alpha}, \bar{\beta}; t) = \prod_{j \in D_j} P_j \pi_j f_j(t; \alpha_j, \beta_j) \times \prod_{i \in D_{NR}} \left\{ 1 - \sum_{j=1}^m P_j \pi_j F_j(T; \alpha_j, \beta_j) \right\} \quad (1)$$

이를 이용하여 식 (1)의 양변에 로그(log)를 취하면, 대수 우도함수(log likelihood function)는 아래와 같이 식 (2)로 얻는다.

$$\log L(\bar{\alpha}, \bar{\beta}; t) = \sum_{j=1}^m \sum_{i \in D_j} \{ \log P_j + \log \pi_j + \log f_j(t; \alpha_j, \beta_j) \} + \sum_{i \in D_{NR}} \log \left\{ 1 - \sum_{j=1}^m P_j \pi_j F_j(T; \alpha_j, \beta_j) \right\} \quad (2)$$

식 (2)에서, 분석시점 T 까지 원인별 모든 고장이 완벽하게 보고된다면, $P_j = 1, j=1,2,\dots,m$ 이고 대수우도함수는 간략화 될 수 있으나, 본 연구에서는 고장 보고율을 $P_j < 1, j=1,2,\dots,m$ 로서 가정하였다. 만약, 부품의 원인별 수명분포가 척도모수(scale parameter) α 와 형태모수(shape parameter) β 를 가지는 와이블분포를 따른다면, 이때 부품의 확률밀도함수와 신뢰성 함수는 아래 식 (3)과 식(4)와 같이 각각 나타난다.

$$f(t_j; \alpha, \beta) = \alpha \beta t_j^{\beta-1} \exp(-\alpha t_j^\beta), \quad t_j > 0 \quad (3)$$

$$R(t_j; \alpha, \beta) = \exp(-\alpha t_j^\beta), \quad t_j > 0 \quad (4)$$

식 (2)는 부품의 수명분포가 와이블분포일 때 아래 식 (5)와 같이 나타난다.

$$\begin{aligned} \log L(\bar{\alpha}, \bar{\beta}; t_j) &= \sum_{j=1}^m n_j (\log P_j + \log \pi_j) \\ &+ \sum_{j=1}^m \sum_{i \in D_j} \{ \log \alpha_j + \log \beta_j + (\beta_j - 1) \log t_{i,j} - \alpha_j t_{i,j}^{\beta_j} \} \\ &+ n_{NR} \log \left[1 - \left\{ \sum_{j=1}^m P_j \pi_j (1 - \exp(-\alpha_j T^{\beta_j})) \right\} \right] \end{aligned} \quad (5)$$

위 대수우도함수의 모수 $\alpha_j, \beta_j, j=1, 2, \dots, m$ 의 MLEs $\hat{\alpha}_j, \hat{\beta}_j, j=1, 2, \dots, m$ 를 구하기 위해서 각각의 모수 $\alpha_j, \beta_j, j=1, 2, \dots, m$ 에 관해 일차 편미분하면 부록 1. 식 (A1)과 식 (A2)로 나타나고, 식 (A1)과 식 (A2)를 동시에 0으로 하면서 대수우도함수를 최대화하는 MLEs $\hat{\alpha}_j$ 과 $\hat{\beta}_j$ 는 Newton-Raphson 방법과 같은 수치적 방법(numerical method)을 이용하여 구할 수 있다.

3. 고장 보고율이 알려지지 않은 경우

본 장에서는 고장 원인별 고장 보고율이 알려지지 않은 경우에 대해서 연구한다. 고장 보고율이 알려지지 않은 경우는 부품이 현장에 설치된 후 시간이 얼마 경과되지 않은 부품 혹은 엔지니어에 의해 고장 보고율의 제안이 어려운 상황일 때 적용될 수 있다. 원인별 고장 보고율이 알려지지 않은 경우에 수집된 현장고장자료와 EM 알고리즘을 사용하여 고장 원인별 고장 보고율을 추정하는 절차를 제시한다. 고장 원인별 부품의 수명 분포가 와이블분포인 경우, 모수를 추정하기 위한 새로운 우도함수를 유도하고 MLE 추정 절차를 제시한다.

3.1 현장고장자료 MLEs 추정 절차

분석 시점 $(0, T]$ 안에 고장 발생한 부품이 자료 수집 시스템에 보고되는 알려지지 않은 고장 보고율은 $P_j, j=1, 2, \dots, m$ 이다. 고장 보고율과 수명분포의 MLE는 EM 알고리즘 원리에 기초하여 추정된다. 여기서 각 원인별 고장 비율 $\pi_j, j=1, 2, \dots, m$ 은 알려져 있다고 가정한다. 단 부품의 고장모드가 단일한 경우 $\pi_1=1$ 이다. 고장 보고율이 알려지지 않은 경우 혼합 와이블분포의 대수우도함수는 식 (6)으로 나타나고, 이는 모수공간(parameter space)을 제외한 식 (2)과 같다.

$$\begin{aligned} \log L(\bar{\alpha}, \bar{\beta}, \bar{P}; t_j) &= \\ &\sum_{j=1}^m \sum_{i \in D_j} \{ \log P_j + \log \pi_j + \log f_j(t_{i,j}; \alpha_j, \beta_j, P_j) \} \\ &+ \sum_{i \in D_{NR}} \log \left\{ 1 - \sum_{j=1}^m P_j \pi_j f_j(T; \alpha_j, \beta_j, P_j) \right\} \end{aligned} \quad (6)$$

식 (2)와 식 (6)의 대수우도함수에서, 분석 시점 T 까지 보고되지 않은 집합 D_{NR} 의 우도에 대한 기여는 분석 시점 T 까지 부품이 고장이 발생했으나 보고되지 않은 부품의 집합 D_f 와 분석 시점 T 까지 고장 나지 않은 부품의 집합 D_s 의 기여에 대신하여 사용되었다. 왜냐하면, n_f 와 n_s 는

분석시점 T 까지 분실되거나 혹은 관측되지 않은 부품들이기 때문이다. 수집된 자료에 결석 자료(missing data)가 포함되어 있을 때, EM 알고리즘은 MLE를 반복적으로 계산하기 위해 사용되어 질 수 있다. EM 알고리즘의 각각의 반복에는 기대화 단계와 최대화 단계 두 가지 단계로 구성되는데 기대화 단계에서, 결석 자료는 수집된 관측 자료를 조건부 기대화에 의해 대체한다. 최대화 단계에서, 모수의 최우추정값은 수집된 관측 자료와 기대화 단계에서 계산된 결석 자료의 조건부 기대화를 가지고 계산된다. EM 알고리즘에 기초하여 새로운 우도함수를 유도하는 절차는 다음과 같다.

초기단계

알려지지 않은 고장 원인별 고장 보고율 $\bar{P}^{(0)}$ 와 혼합분포의 초기추정값을 $\bar{\alpha}^{(0)}, \bar{\beta}^{(0)}$ 로 설정한다.

기대화 단계

분석 시점 T 까지 고장 났으나 보고되지 않은 부품의 결석변수 n_f 는 이항분포 (binomial distribution) $B(n_{NR}, p_{NR}^{(k)})$ 를 따른다. 부가 설명을 하면, 분석 시점 T 까지 보고되지 않은 부품 n_{NR} 에서 고장 발생한 부품이 보고되지 않은 사건 n_f 를 성공이라 했을 때, 이는 독립적인 베르누이(Bernoulli) 실험을 n_{NR} 번 시도하여 성공이 n_f 회 발생하는 이항분포를 따른다. 이때 성공확률이 $p_{NR}^{(k)}$ 라고 하면, 확률 $p_{NR}^{(k)}$ 는 다음의 식 (7)을 통해 구하여 진다.

$$p_{NR}^{(k)} = \frac{\sum_{j=1}^m (1 - P_j) \pi_j f_j(T)}{1 - \sum_{j=1}^m P_j \pi_j f_j(T)} \quad (7)$$

식 (7)에서, 분모는 분석 시점 T 까지 보고되지 않은 부품의 확률이고, 분자는 분석 시점 T 까지 고장 발생한 부품이 보고되지 않은 확률이다. 이렇게 이항분포의 성공 확률 $p_{NR}^{(k)}$ 을 사용하여, k 번째 반복단계에서 주어진 n_{NR} 와 n_f 의 기대값은 이항분포의 평균 공식을 사용하여 아래 식 (8)과 같이 구하여 진다.

$$E^{(k)}(n_f) = n_{NR} \times p_{NR}^{(k)} \quad (8)$$

여기서 $E^{(k)}(n_f)$ 는 $E^{(k)}(n_f) = E_1^{(k)}(n_f) + E_2^{(k)}(n_f) + \dots + E_m^{(k)}(n_f)$ 로서 고장 원인별 기대값의 합으로 나타난다. $E_j^{(k)}(n_f), j=1, 2, \dots, m$ 의 값은 고장 원인별 고장 비율 $\pi_j, j=1, 2, \dots, m$ 가 이미 알려져 있다고 가정하였으므로 아래의 식 (9)와 같이 구하여 진다.

$$\begin{bmatrix} E_1^{(k)}(n_f) = E^{(k)}(n_f) \times \pi_1 \\ E_2^{(k)}(n_f) = E^{(k)}(n_f) \times \pi_2 \\ \vdots \\ E_m^{(k)}(n_f) = E^{(k)}(n_f) \times \pi_m \end{bmatrix} \quad (9)$$

그리고 n_{NR} 는 $n_{NR} = n_f + n_s$ 이기 때문에 n_s 의 기대값은 위의 식 (8)을 통해 구하여진 $E^{(k)}(n_f)$ 를 이용하여 다음의 식 (10)과 같이 구하여 진다.

$$E^{(k)}(n_s) = n_{NR} - E^{(k)}(n_f) \quad (10)$$

최대화 단계

식 (2) 대수우도함수는 결석 자료 n_r 와 n_s 자료를 대신 하여 분석 시점에서 관측된 n_{NR} 을 사용하여 유도된다. 그러나 여기서 식 (8)과 식 (10)을 사용하여 획득된 $E^{(k)}(n_r)$ 와 $E^{(k)}(n_s)$ 는 집합 D_{NR} 을 대신하여 집합 D_r 와 집합 D_s 로 우도에 기여하는데 사용될 수 있다. 따라서 k 번째 반복 단계에서 대수우도함수는 식 (2)의 대수우도함수와는 다른 새로운 대수우도함수를 아래와 같이 식 (11)로 나타낼 수 있다.

$$\begin{aligned} \log L(\bar{\alpha}, \bar{\beta}, \bar{P}; t_i) = & \sum_{j=1}^m \sum_{i \in D_j} \{ \log P_j + \log \pi_j + \log f_j(t_i; \alpha_j, \beta_j, P_j) \} \\ & + E^{(k)}(n_r) \log \left\{ \sum_{j=1}^m (1-P_j) \pi_j F_j(T; \alpha_j, \beta_j, P_j) \right\} \\ & + E^{(k)}(n_s) \log \left\{ \sum_{j=1}^m \pi_j R_j(T; \alpha_j, \beta_j, P_j) \right\} \quad (11) \end{aligned}$$

앞에서 구한 새로운 대수우도함수를 최대화하는 모수 $(\alpha_j, \beta_j, P_j), j=1, 2, \dots, m$ 의 MLEs $(\alpha_j^{(k)}, \beta_j^{(k)}, P_j^{(k)}), j=1, 2, \dots, m$ 을 구하기 위한 방법은 고장 보고율이 알려진 경우의 2.1절 경우와 같은 수치적 방법을 통해 구하여진다. 특별히 고장 보고율 $P_j^{(k)}, j=1, 2, \dots, m$ 는 식 (12)를 통하여 구하여진다.

$$\begin{aligned} P_1^{(k)} &= \frac{n_1}{n_1 + E_1^{(k)}(n_r)} = \frac{n_1}{n_1 + \pi_1 E^{(k)}(n_r)} \\ P_2^{(k)} &= \frac{n_2}{n_2 + E_2^{(k)}(n_r)} = \frac{n_2}{n_2 + \pi_2 E^{(k)}(n_r)} \\ &\vdots \\ P_m^{(k)} &= \frac{n_m}{n_m + E_m^{(k)}(n_r)} = \frac{n_m}{n_m + \pi_m E^{(k)}(n_r)} \end{aligned}$$

단, $E^{(k)}(n_r) = E_1^{(k)}(n_r) + \dots + E_m^{(k)}(n_r)$

4.1 고장 보고율이 알려진 경우

고장이 누락된 현장고장자료를 모의실험을 통해서 얻고 고장 보고율은 알려져 있다.

4.1.1 단일 와이블분포의 MLEs와 MSEs

실험 조건

$$\alpha_1=0.05 \quad \beta_1=3.0 \quad P_1=0.2, 0.5 \quad T=5$$

부품의 수명분포 모수가 위의 실험조건과 같을 때, 고장 보고율 P_1 은 알려져 있고, 모의실험을 통해 추정된 α_1 과 β_1 의 MLEs $\hat{\alpha}_1$ 과 $\hat{\beta}_1$ 결과는 다음의 [Table 1]과 같다.

모의실험을 통한 MLEs와 $\hat{\alpha}_1$ 과 $\hat{\beta}_1$ 의 MSEs를 나타내고 있다. 보고가 누락되는 현실성을 고려하기 위해 고장 보고율을 도입하여 모수를 추정 했을 때, MLEs $\hat{\alpha}_1$ 와 $\hat{\beta}_1$ 는 모수 α_1, β_1 에 거의 일치하는 결과를 나타내고 있다. MSEs는 MLEs $\hat{\alpha}_1, \hat{\beta}_1$ 와 모수 α_1, β_1 와의 거리를 나타내는데 $MSE(\hat{\alpha}_1)$ 는 10^{-5} 정도의 작은 값을 나타내고 있고, $MSE(\hat{\beta}_1)$ 역시 10^{-2} 정도의 작은 값을 나타내고 있다. 이 결과는 추정된 MLEs가 일치추정량(consistent estimator)에 거의 일치하는 결과를 보임을 알 수가 있다.

특히, 고장 원인별 수명분포가 와이블분포일 경우 대수우도함수는 식 (13)과 같고, 각 모수에 관하여 일차 편미분하면 부록 1 (A3)과 (A4)로 나타내어진다.

$$\begin{aligned} \log L(\bar{\alpha}, \bar{\beta}, \bar{P}; t_i) = & \sum_{j=1}^m n_j (\log P_j + \log \pi_j) \\ & + \sum_{j=1}^m \sum_{i \in D_j} \{ \log \alpha_j + \log \beta_j + (\beta_j - 1) \log t_i - \alpha_j t_i^{\beta_j} \} \\ & + E^{(k)}(n_r) \log \left[\sum_{j=1}^m (1-P_j) \pi_j \{ 1 - \exp(-\alpha_j T^{\beta_j}) \} \right] \\ & + E^{(k)}(n_s) \log \left[\sum_{j=1}^m \pi_j \exp(-\alpha_j T^{\beta_j}) \right] \quad (13) \end{aligned}$$

EM 알고리즘의 기대화 단계와 최대화 단계 반복이 진행됨에 따라, 고장 보고율 $P_j, j=1, 2, \dots, m$ 과 수명분포 모수 $\alpha_j, \beta_j, j=1, 2, \dots, m$ 의 추정값은 안정해에 수렴하게 된다. 만약에 우도 함수가 단봉이라면, EM 알고리즘의 안정해는 유일한 최우추정값이다.

4. 실험 및 분석

본 장에서는 모의실험을 통하여 고장 원인별 고장 보고율과 주어진 수명자료를 바탕으로 부품의 수명분포 모수추정에 대한 타당성을 검증하여 본다. 이를 위해서, 부품의 고장 원인별 수명분포는 와이블분포로 가정하였고, 고장 보고율이 알려진 경우와 고장 보고율이 알려지지 않은 경우에 고장 모드가 하나인 단일 와이블분포와 고장 모드가 두 가지인 경우의 2-혼합 와이블분포에 대해서 모수추정을 하였다. 마지막으로 고장 보고율을 고려하지 않은 경우의 결과와 비교 분석해 보기로 한다.

4.1.2 2-혼합 와이블분포의 MLEs와 MSEs

부품의 고장 모드가 두 가지인 경우 2-혼합 와이블분포의 모수추정을 하기 위한 실험 조건은 다음과 같다.

실험 조건

$$\begin{aligned} \alpha_1=0.05 & \quad \beta_1=3.0 & \quad P_1=0.2, 0.5, 0.8 \\ \alpha_2=0.1 & \quad \beta_2=3.0 & \quad P_2=0.2, 0.5, 0.8 \\ \pi_1=0.3 & \quad \pi_2=0.7 & \quad T=5 \end{aligned}$$

부품의 두 가지 고장 원인에 따른 고장 보고율 P_1, P_2 과 고장 비율 π_1, π_2 은 이미 알려져 있다고 가정을 하였고, 모수를 추정된 결과는 다음의 [Table 2]와 같다.

앞선 단일 와이블분포의 결과와 동일하게 2-혼합 와이블분포에서의 각 원인별 수명분포의 모수의 MLEs는 입력모수와 거의 일치하게 나타났다. 따라서 각 원인별 고장 보고율을 도입하여 추정된 MLEs는 추정의 정확성을 증가시킨다는 것을 알 수 있다.

4.2 고장 보고율이 알려지지 않은 경우

고장 보고율이 알려지지 않은 경우에서도, 현장고장자료에서 고장이 누락된 자료를 모의실험을 통해서 얻는다. 단, 고장 보고된 자료만을 분석자가 가지고 있을 뿐, 각 원인별 고장 보고율은 알려지지 않았다.

4.2.1 단일 와이블분포의 MLEs와 MSEs

고장 보고율이 알려지지 않은 경우 실험조건은 고장 보고율이 알려져 있는 경우와 같으나, 고장이 보고된 자료를 바탕으로 고장 보고율 P_1 을 EM 알고리즘에 기초해서 추정하고, 동시에 부품의 수명분포 모수를 추정한다.

[Table 3]에서 나타나는 바와 같이 추정된 고장 보고율은 입력된 고장 보고율과 거의 일치하게 추정되었으며, 수명분포의 MLEs 역시 입력된 모수에 일치하는 결과를 보이고 있다. 위 실험 결과는 수명자료 분석시, 고장 보고율이 알려지지 않은 자료가 주어졌을 때, 본 논문에서 제안된 방법으로 알려지지 않은 고장 보고율과 수명 분포의 모수를 동시에 추정한다면 MLEs의 정확도가 증가함을 알 수 있다.

4.2.2 2-혼합 와이블분포의 MLEs와 MSEs

고장 보고율 P_1, P_2 이 알려지지 않은 2-혼합 와이블분포의 실험조건 또한 고장 보고율이 알려진 경우의 실험조건과 같다. 단, 알려지지 않은 고장 보고율 P_1, P_2 을 EM 알고리즘에 기초한 추정 알고리즘을 통하여 추정하고 이와 동시에 2-혼합 와이블분포의 모수를 추정하였다. 실험 결과는 다음의 [Table 4]와 같다. 추정된 고장 보고율을 통한 2-혼합 와이블분포의 MLEs $\hat{\alpha}_1, \hat{\beta}_1$ 와 $\hat{\alpha}_2, \hat{\beta}_2$ 는 모수 α_1, β_1 와 α_2, β_2 에 거의 일치하는 결과를 나타내고 있음을 알 수 있다.

Table 1. 알려진 고장 보고율 P_1 을 고려한 MLEs와 MSEs

모 수 값			MLEs		$MSE(\hat{\alpha}_1)$	$MSE(\hat{\beta}_1)$
α_1	β_1	P_1	$\hat{\alpha}_1$	$\hat{\beta}_1$		
0.05	3.0	0.2	0.05204	3.00079	0.00010	0.02863
		0.5	0.05208	2.97410	0.00004	0.01048
		0.8	0.05105	3.00529	0.00002	0.00593

Table 2. 알려진 고장 보고율 P_1, P_2 를 고려한 MLEs와 MSEs

모 수 값						MLEs			
α_1	β_1	P_1	α_2	β_2	P_2	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$
0.05	3.0	0.2	0.1	3.0	0.2	0.05100	3.04662	0.10224	2.97916
		0.5			0.5	0.04466	3.08106	0.10195	3.01227
		0.8			0.8	0.04919	3.02593	0.09918	3.00626
α_1	β_1	P_1	α_2	β_2	P_2	$MSE(\hat{\alpha}_1)$	$MSE(\hat{\beta}_1)$	$MSE(\hat{\alpha}_2)$	$MSE(\hat{\beta}_2)$
0.05	3.0	0.2	0.1	3.0	0.2	0.00035	0.11506	0.00053	0.07053
		0.5			0.5	0.00014	0.04611	0.00015	0.01554
		0.8			0.8	0.00007	0.02121	0.00010	0.00871

Table 3. 알려지지 않은 고장 보고율 P_1 을 고려한 MLEs와 MSEs

모 수 값			MLE			$MSE(\hat{\alpha}_1)$	$MSE(\hat{\beta}_1)$	$MSE(\hat{P}_1)$
α_1	β_1	P_1	$\hat{\alpha}_1$	$\hat{\beta}_1$	\hat{P}_1			
0.05	3.0	0.2	0.05218	2.99710	0.19958	0.00010	0.02909	0.00016
		0.5	0.05254	2.96226	0.50048	0.00005	0.01162	0.00022
		0.8	0.05184	2.98466	0.80129	0.00003	0.00675	0.00015

Table 4. 알려지지 않은 고장 보고율 P_1, P_2 를 고려한 MLEs와 MSEs

모 수 값						MLEs					
α_1	β_1	P_1	α_2	β_2	P_2	$\hat{\alpha}_1$	$\hat{\beta}_1$	\hat{P}_1	$\hat{\alpha}_2$	$\hat{\beta}_2$	\hat{P}_2
0.05	3.0	0.2	0.1	3.0	0.2	0.05073	3.05539	0.19878	0.10071	3.00220	0.20070
		0.5			0.04591	3.08136	0.49760	0.10195	3.01227	0.49975	
		0.8			0.04916	3.03031	0.79575	0.09918	3.00626	0.79847	
α_1	β_1	P_1	α_2	β_2	P_2	$MSE(\hat{\alpha}_1)$	$MSE(\hat{\beta}_1)$	$MSE(\hat{P}_1)$	$MSE(\hat{\alpha}_2)$	$MSE(\hat{\beta}_2)$	$MSE(\hat{P}_2)$
0.05	3.0	0.2	0.1	3.0	0.2	0.00035	0.11602	0.00038	0.00040	0.04107	0.00020
		0.5			0.00014	0.04629	0.00053	0.00015	0.01554	0.00028	
		0.8			0.00007	0.02114	0.00023	0.00010	0.00871	0.00016	

4.3 비교분석

본 절에서는 부품의 수명분포가 단일 와이블분포인 경우, 고장 보고율을 고려하지 않은 경우와 고장 보고율을 고려한 결과를 비교 분석하고 이와 동시에 고장 보고율 증감에 따른 결과를 살펴본다. 수명분포의 모수를 추정한 결과는 아래 [Table 5]와 같다.

[Table 5]는 예상했던 바와 같이, 고장 보고율을 고려하지 않고 주어진 고장자료만으로 부품의 수명분포 모수를 추정한 결과는 입력 모수와는 매우 동떨어진 결과를 나타내고 있고 고장자료의 수가 증가함에 따라 MLEs는 입력 모수와 근접하는 추세를 보이고 있지만 차이가 크다는 것을 알 수 있다. 반면에 주어진 고장자료와 알려지지 않은 고장 보고율을 추정하여 부품의 수명분포 모수를 동시에 추정한 결과는 고장 보고율이 매우 낮더라도, 입력모수와 매우 근접한 MLEs를 나타내고 있다. 또한 고장 보고율이 알려지지 않은 경우와 고장 보고율이 알려진 경우의 MLEs를 비교하여 보면 매우 근소한 차이지만 고장 보고율이 알려져 있을 때의 MLEs가 입력모수와 좀 더 일치하는 결과를 보이고 있고 고장 보고율의 값이

증가함에 따라서 추정된 MLEs의 정확도는 점점 증가하는 추세를 보이고 있다.

다음의 [Table 6]은 위에서 추정된 MLEs에 대한 MSEs 결과를 나타내고 있다. [Table 6]은 MLEs의 MSEs 값이 고장 보고율이 증가함에 따라 점점 작아지는 추세를 보여 주고 있다. 예상한 바와 같이 고장 보고율을 고려하지 않은 경우의 MSEs 값은 매우 크게 도출 되었다. 위 결과를 각각 $\hat{\alpha}_1$ 와 $\hat{\beta}_1$ 에 관해 그래프로 나타내면 다음의 [Figure 1]과 [Figure 2]와 같다. 나타나는 바와 같이 고장 보고율을 고려하지 않은 MSE와 고장 보고율을 고려한 MSE의 차이는 고장 보고율이 증가함에 따라 현저히 큰 차이를 보임을 알 수 있다. 이는 고장 보고율이 알려져 있지 않더라도 본 연구의 모수추정 알고리즘을 이용한다면 실제에 가까운 MLE를 구할 수 있고 알려진 고장 보고율을 이용하여 MLE를 구한다면 더욱 정확도가 증가한 모수 추정값을 구할 수 있음을 알 수가 있다.

Table 5. P_1 을 고려하지 않은 경우와 P_1 을 고려한 MLEs 비교 분석

모 수 값			P_1 고려하지 않은 <i>MLEs</i>		알려지지 않은 P_1 고려한 <i>MLEs</i>		알려진 P_1 고려한 <i>MLEs</i>	
α_1	β_1	P_1	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_1$	$\hat{\beta}_1$
0.05	3	0.1	1.01901	-0.44812	0.06419	2.82840	0.06407	2.83109
		0.2	0.02757	1.31219	0.05707	2.90546	0.05686	2.91050
		0.3	0.03988	1.38347	0.05459	2.91923	0.05424	2.92784
		0.4	0.05279	1.44168	0.05560	2.90445	0.05514	2.91585
		0.5	0.06443	1.51709	0.05528	2.91395	0.05474	2.92737
		0.6	0.07426	1.61780	0.05429	2.93065	0.05366	2.94660
		0.7	0.08160	1.74879	0.05409	2.92804	0.05329	2.94831
		0.8	0.08675	1.91867	0.05522	2.91784	0.05427	2.94159
		0.9	0.08224	2.21079	0.05525	2.91273	0.05405	2.94277

Table 6. P_1 을 고려하지 않은 경우와 P_1 을 고려한 MSEs 비교 분석

모 수 값			P_1 고려하지 않은 <i>MSEs</i>		알려지지 않은 P_1 고려한 <i>MSEs</i>		알려진 P_1 고려한 <i>MSEs</i>	
α_1	β_1	P_1	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_1$	$\hat{\beta}_1$
0.05	3	0.1	2.76706	11.88980	0.00048	0.07931	0.00047	0.07776
		0.2	1.82859	2.84884	0.00018	0.03805	0.00018	0.03650
		0.3	1.82819	2.61326	0.00011	0.02718	0.00010	0.02521
		0.4	1.82809	2.42842	0.00009	0.02350	0.00009	0.02087
		0.5	1.82829	2.19907	0.00008	0.01893	0.00007	0.01628
		0.6	1.82867	1.91050	0.00005	0.01362	0.00005	0.01111
		0.7	1.82908	1.56555	0.00005	0.01312	0.00004	0.00997
		0.8	1.82943	1.16931	0.00006	0.01367	0.00005	0.00966
		0.9	1.82912	0.62288	0.00005	0.01374	0.00004	0.00873

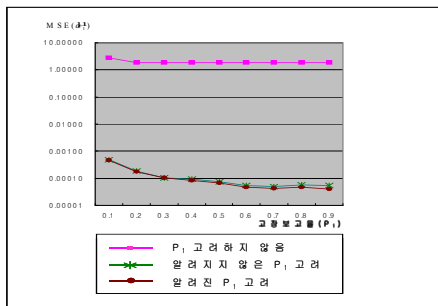


Figure 1. P_1 증가에 따른 $MSE(\hat{\alpha}_1)$ 변화

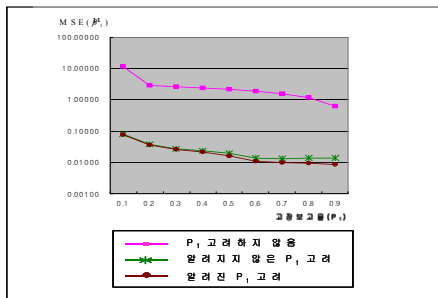


Figure 2. P_1 증가에 따른 $MSE(\hat{\beta}_1)$ 변화

5. 결론

본 논문에서는 현장에 설치된 부품의 고장이 자료 수집 시스템에 고장 보고가 누락되는 현실성을 고려하기 위하여 부품의 고장 원인별 고장 보고율을 이용한 모수추정 알고리즘을

제시하였다. 제시된 모수추정 방법론은 고장 보고율을 이용할 때 보고가 누락된 자료를 추가 반영하는 효과를 가짐으로써 부품에 대한 MLE의 정확도를 증가시키는 추정 방법론이다. 또한 부품의 고장 보고율이 알려진 경우뿐만 아니라 고장 보고율이 알려지지 않은 경우에 대해서도 EM 알고리즘과 MLE에 기초한 추정 알고리즘을 통해 고장 보고율을 추정하고 동시에 수명분포의 모수를 추정하는 절차를 제시하였다. 특히 부품의 고장 원인별 수명분포가 와이블분포인 경우에 고장모드가 하나인 경우뿐만 아니라 여럿인 경우에 대해 고장 원인별 고장 보고율을 2-혼합 와이블분포에 적용한 실험예제를 통하여 제시된 방법론의 타당성을 검증하여 보았다.

본 논문의 결과는 주어진 고장자료와 고장 보고율을 고려하여 모수를 추정하면 MLE가 실제와 일치하였고 고장 보고율이 증가함에 따라 MLE의 정확도가 증가함을 알 수 있었다. 추후 연구 방향으로 와이블분포 대신에 대수정규분포, 감마분포 그리고 지수분포 등과 같은 고장 원인별 다른 형태의 수명분포를 가진 혼합 모델에 대한 연구가 필요하다.

참고 문헌

Abernethy, R. B. (1988), "The New Weibull Handbook 3rd ed", Reliability Analysis Center, Florida.
 Barlow, R. E. and Proschan, F. (1979), "Inference and Data Analysis for Reliability and Life Testing", California.
 Bucar, T., Nagode, M. and Fajdiga, M. (2004), "Reliability approximation using finite Weibull mixture distribution", *Reliability Engineering and System Safety*, 84, 241-251.
 Coit, D. W. and Dey, K. A. (1988), "Analysis of grouped

data from field failure reporting systems”, *Reliability Engineering and System Safety*, 65, 95-101.

Coit, D. W. and Jin, T. (2000), “Gamma distribution parameter estimation for field reliability data with missing failure times”, *IIE Transactions*, 32, 1161-1166.

Hale, P. S. and Arno, R. G. (2001), “Survey of Reliability and Availability Information for Power Distribution, Power Generation, and HVAC components for Commercial, Industrial and Utility Installations”, *IEEE Transaction on Industry Applications*, 37(1), 191-196.

Jiang, S. and Kececioglu, D. (1992), “Maximum Likelihood Estimates, from Censored Data, for Mixed-Weibull Distributions”, *IEEE Transactions on Reliability*, 41(2), 248-255.

Jiang, R. and Murthy, D.N.P. (1995), “Modeling Failure-Data by Mixture of 2 Weibull Distribution : A

Graphical Approach”, *IEEE Transactions on Reliability*, 44(3), 477-488.

Jiang, R. and Murthy, D.N.P. (1995), “Reliability modeling involving two Weibull distributions”, *Reliability Engineering and System Safety*, 47, 187-198.

Jiang, R. and Murthy, D.N.P. (1988), “Mixture of weibull distributions-parametric characterization of failure rate function”, *Applied Stochastic Models and Data Analysis*, 14, 47-65.

Lim, T. J. (2002), “Estimation of Product Reliability with Incomplete Field Warranty Data”, *Journal of the Korean Institute of Industrial Engineers*, 28(4), 368-378.

Oh, Y. S. and Bai, D. S. (2001), “Field data analyses with additional after warranty failure data”, *Reliability Engineering and System Safety*, 72, 1-8.

부 록 1. 편미분

$$\frac{\partial \log L}{\partial \alpha_j} = \frac{n_j}{\alpha_j} - \sum_{i \in D_j} t_i^{\beta_j} + n_{NR} \frac{G_{\alpha_j}}{G}, \quad j=1,2,\dots,m$$

$$G = \left[1 - \left\{ \sum_{j=1}^m P_j \pi_j (1 - \exp(-\alpha_j T^{\beta_j})) \right\} \right]$$

$$G_{\alpha_j} = -P_j \pi_j T^{\beta_j} \exp(-\alpha_j T^{\beta_j})$$
(A1)

$$\frac{\partial \log L}{\partial \beta_j} = \frac{n_j}{\beta_j} + \sum_{i \in D_j} (\log t_i - \alpha_j t_i^{\beta_j} \log t_i) + n_{NR} \frac{G_{\beta_j}}{G}, \quad j=1,2,\dots,m$$

$$G_{\beta_j} = -P_j \pi_j \alpha_j T^{\beta_j} \log T \exp(-\alpha_j T^{\beta_j})$$
(A2)

$$\frac{\partial \log L}{\partial \alpha_j} = \frac{n_j}{\alpha_j} - \sum_{i \in D_j} t_i^{\beta_j} + E^{(k)}(n_p) \frac{D_{\alpha_j}}{D} + E^{(k)}(n_s) \frac{Q_{\alpha_j}}{Q}, \quad j=1,2,\dots,m$$

$$D = \sum_{j=1}^m (1 - P_j) \pi_j \{ 1 - \exp(-\alpha_j T^{\beta_j}) \}, \quad D_{\alpha_j} = (1 - P_j) \pi_j T^{\beta_j} \exp(-\alpha_j T^{\beta_j})$$

$$Q = \sum_{j=1}^m \pi_j \exp(-\alpha_j T^{\beta_j}), \quad Q_{\alpha_j} = -\pi_j T^{\beta_j} \exp(-\alpha_j T^{\beta_j})$$
(A3)

$$\frac{\partial \log L}{\partial \beta_j} = \frac{n_j}{\beta_j} + \sum_{i \in D_j} [\log t_i - \alpha_j t_i^{\beta_j} \log t_i] + E^{(k)}(n_p) \frac{D_{\beta_j}}{D} + E^{(k)}(n_s) \frac{Q_{\beta_j}}{Q}, \quad j=1,2,\dots,m$$

$$D_{\beta_j} = (1 - P_j) \pi_j \alpha_j T^{\beta_j} \log T \exp(-\alpha_j T^{\beta_j})$$

$$Q_{\beta_j} = -\pi_j \alpha_j T^{\beta_j} \log T \exp(-\alpha_j T^{\beta_j})$$
(A4)