

불완전한 사용현장 보증 데이터를 이용한 제품 신뢰도 추정

임태진*

승실대학교 산업·정보시스템공학과

Estimation of Product Reliability with Incomplete Field Warranty Data

Tae-Jin Lim

Department of Industrial & Information Systems Engineering, Soongsil University, Seoul, 156-743

As more companies are equipped with data acquisition systems for their products, huge amount of field warranty data has been accumulated. We focus on the case when the field data for a given product comprise with the number of sales and the number of the first failures for each period. The number of censored items and their ages are assumed to be given. This type of data are incomplete in the sense that the age of a failed item is unknown. We construct a model for this type of data and propose an algorithm for nonparametric maximum likelihood estimation of the product reliability. Unlike the nonhomogeneous Poisson process(NHPP) model, our method can handle the data with censored items as well as those with small population. A few examples are investigated to characterize our model, and a real field warranty data set is analyzed by the method.

Keywords: field data, warranty, nonparametric MLE, EM algorithm, reliability

1. 서론

가전제품과 같이 양산하여 판매하는 제품에 있어서는 개발단계에서뿐 아니라 시장에 출하한 후의 사용단계에서 신뢰성을 평가·확인하는 것이 중요하다. 개발단계에서 실험실에서 얻어지는 수명시험 데이터는 통제된 환경에서의 정밀한 결과를 제공하지만, 비용이 많이 소요될 뿐 아니라 실제 사용환경을 잘 반영하지 못하는 결점이 있다. 이에 반하여 사용현장 데이터는 불확실성이 크지만, 저렴하게 구할 수 있고 실제 사용환경을 반영하는 장점이 있다. 최근에는 제품 신뢰도 데이터 수집 시스템을 갖춘 기업들이 늘어남에 따라, 이러한 사용현장 데이터의 활용이 더욱 요구되고 있다.

사용현장 데이터는 주로 고객 서비스센터로부터 입수된 제품 고장 및 수리에 관한 보증 데이터로서, 본 연구에서는 각 기간별 총 판매대수와 각 기간별 총 고장대수로 정리된 데이터

를 대상으로 한다. 이러한 데이터는 고장 제품의 사용기간(age)을 알 수 없기 때문에 불완전 데이터로 분류된다. 고장 제품의 제품고유번호를 통하여 제조 시점은 알 수 있으나, 판매 시점, 혹은 사용개시 시점은 알 수 없는 경우가 대부분이다. 기존 업계에서는 제조 시점 1개월 후를 사용개시 시점으로 정하는 등의 검증되지 않은 가정을 통해 분석을 시도하기도 하였다. 이와 같이 불완전한 사용현장 보증 데이터에서 판매 시점에 대한 정보는 영업부문에 수집된 기간별 판매대수를 통해 얻을 수 있다.

본 논문에서는 이러한 불완전한 보증 데이터를 이용하여 제품 신뢰도를 추정하는 비모수적 방법을 제안한다. 특히 기존 방법으로는 다루지 못하던 사용 중단된 제품의 개수에 대한 데이터가 주어진 경우를 분석한다. 관련된 연구 현황 및 사용된 가정 및 기호는 다음과 같다.

기간별로 수집된 보증 데이터 분석에 관한 주목할 만한 연구는 비교적 최근에 발표되었다. Kalbfleisch *et al.*(1991)은 보

본 연구는 2001년도 LG연암문화재단 해외연구교수 지원을 받아 수행되었음.

* 연락처: 임태진 교수, 156-743 서울시 동작구 상도동 1-1 승실대학교 산업·정보시스템공학과, Fax: 02-825-1094, e-mail: tjlim@ssu.ac.kr
2001년 12월 접수, 1회 수정(15주 소요) 후, 2002년 9월 게재 확정.

증 데이터를 이용하여 보증 클레임 횟수의 기대값을 구하였다. 그들은 대수-선형 포아송(log-linear Poisson) 모형을 제안하여 문제를 간단히 표현하였으며, 고장 제품의 사용기간을 알고 동일 제품의 반복(recurrent) 고장을 허용하는 경우를 고려하였다. Baxter(1994)는 고장난 제품의 사용개시 시점을 모르는 경우의 수명표(life table)로부터 비모수적 방법으로 이산형 수명분포를 추정하였다. 이 방법은 매 시점마다 추가되는 제품수가 동일한 경우에만 적용 가능하다. Tortorella(1996)는 추가되는 제품수가 변하는 보다 일반적인 경우를 고려하였으나, 명확한 해를 제시하지 못하였다. 최근 Karim et al.(2001)은 각 기간별 고장 제품수와 판매 제품수만이 주어진 경우, Kalbfleisch et al.(1991)에서와 유사한 NHPP 모형을 제안하고, EM 알고리즘을 이용하여 각 기별 고장 강도(intensity)를 추정하였다. 기존 연구들은 모두 반복 고장을 허용하는 공통점을 갖는다. 이러한 모형은 수리 가능한 제품수준 데이터에는 적용 가능하지만, 수리 불가능한 부품의 경우, 특히 관측 중단이 있는 경우에는 적용하기 곤란한 문제점이 있다.

본 연구에서는 반복 고장을 제외한 첫 번째 고장 데이터를 대상으로 하여 제품이나 부품의 신뢰도를 추정하는 것을 목적으로 한다. Karim et al.(2001)도 수리 불가능한 경우의 데이터에 대해서 언급하였으나 관측 중단이 있는 경우에는 적용할 수 없으며, 그들의 실험결과는 NHPP 모형에서와 거의 동일한 것으로 나타났다. 따라서 본 연구에서는 관측 중단 데이터를 포함한 첫 번째 고장에 대한 사용현장 보증 데이터를 분석하기 위한 새로운 모형 및 추정 알고리즘을 제안하고자 한다.

가정

사용현장 데이터에 영향을 미치는 변수는 매우 다양하기 때문에 분석에 앞서서 많은 가정이 요구된다. 본 연구에서 사용한 가정을 정리하면 다음과 같다.

- ① 고장 데이터 및 판매 데이터는 일, 주, 월 등의 정해진 기(period) 단위로 집계되며, 고장률, 고장확률, 신뢰도 등의 제품 신뢰성 척도는 기별로 추정한다.
- ② 첫 번째 고장만을 고려하며, 동일한 제품의 반복되는 고장은 집계에서 제외한다.
- ③ 판매 시점을 사용개시 시점으로 간주한다.
- ④ 대상 제품의 수명분포는 제조·판매시기와 상관없이 동일하다.
- ⑤ 동일한 기간에 판매되고 고장난 경우의 수명은 0으로 간주한다.
- ⑥ 모든 고장은 서비스센터에 즉시 보고된다.
- ⑦ 관측 중단수는 판매 시기와 중단 시기에 따라 집계되며, 관측 중단은 기말에 발생하는 것으로 간주한다.

기호

- S : 총 판매기간(기수)
- T : 총 관측기간(기수) ($S \leq T$)

- N_i : i 기에 판매된 총 제품수 ($i = 1, \dots, S$)
- d_j : j 기에 집계된 총 고장대수 ($j = 1, \dots, T$)
- $d_{i,t}$: i 기에 판매된 제품 중 t 기 경과 후 고장난 제품수 ($t = 0, \dots, T-i, i = 1, \dots, S$)
- 즉, t 는 제품의 수명을 나타내며, $d_{i,t}$ 는 관측되지 않음
- $c_{i,t}$: i 기에 판매된 제품 중 t 기 경과 후 사용 중단된 제품수 ($t = 0, \dots, T-i, i = 1, \dots, S$)
- r_i : i 기에 판매된 제품 중 T 기까지 작동 중인 제품수 ($i = 1, \dots, S$)
- $n_{i,t}$: i 기에 판매된 제품 중 수명 t 기 직전까지 작동 중인 제품수 ($t = 0, \dots, T-i, i = 1, \dots, S$)
- X : 제품 수명
- R_t : 시점 t 에서의 신뢰도 ($t = 0, \dots, T-1$)
- q_t : 시점 t 에서의 고장 확률 ($t = 0, \dots, T-1$)
- h_t : 시점 t 에서의 고장률 ($t = 0, \dots, T-1$)

2장에서는 사용현장 보증 데이터의 구성과 완전 데이터에 근거한 우도함수를 유도하고, 3장에서는 기대화-최대화(EM : expectation maximization) 원리(Dempster et al., 1977)에 기초한 최대우도(maximum likelihood) 추정 알고리즘을 제안한다. 4장에서는 모의실험을 통하여 제안된 방법의 특성을 살펴보고, 사례를 통하여 실제 사용현장 보증 데이터를 분석한다. 5장에서는 결론 및 추후 연구방향을 제시한다.

2. 사용현장 보증 데이터 모형

먼저 연구의 대상이 되는 사용현장 보증 데이터의 구성을 알아보고, 관측 데이터의 우도함수를 표현할 수 없는 상황에서 완전 데이터가 주어진 경우를 가정하여 다양한 각도에서 우도함수를 유도한다.

2.1 불완전 데이터

만약 고장난 제품의 사용개시 시점을 모두 알 수 있다면, 모든 $d_{i,t}$ 를 알 수 있으며, 이러한 경우를 완전 데이터로 정의한다. 본 연구에서 다루는 데이터에서는 각 셀(i, t)의 고장대수 $d_{i,t}$ 를 알 수 없다. 다만, 서비스 부문으로부터 기별 고장대수 d_i 와 영업 부문으로부터 기별 판매대수 N_i 는 알 수 있다. 이러한 불완전 데이터를 주변 데이터(marginal data)라고 부르며(Karim et al., 2001), 그 구성은 <표 1>과 같다. 단, 어두운 영역의 데이터는 관측되지 않는 완전 데이터를 의미한다. 또한 가정 7에 의하여 각 기마다 $d_{i,t}$ 개의 고장이 발생한 후 $c_{i,t}$ 개의 관측 중단이 발생하는 것으로 간주한다.

이러한 데이터의 구성에 따라 각 관측기별로 집계된 고장대수는 식 (1)과 같고,

표 1. 완전 데이터 및 불완전(주변) 데이터의 구성

관측기 판매기	1	2	3	...	j	...	S	...	T	관측 중단	판매수량
1	d_{10}	d_{11}	d_{12}	...	$d_{1,j-1}$...	$d_{1,s-1}$...	$d_{1,T-1}$	r_1	N_1
	c_{10}	c_{11}	c_{12}	...	$c_{1,j-1}$...	$c_{1,s-1}$...	$c_{1,T-1}$		
2		d_{20}	d_{21}	...	$d_{2,j-2}$...	$d_{2,s-2}$...	$d_{2,T-2}$	r_2	N_2
		c_{20}	c_{21}	...	$c_{2,j-2}$...	$c_{2,s-2}$...	$c_{2,T-2}$		
3			d_{30}	...	$d_{3,j-3}$...	$d_{3,s-3}$...	$d_{3,T-3}$	r_3	N_3
			c_{30}	...	$c_{3,j-3}$...	$c_{3,s-3}$...	$c_{3,T-3}$		
⋮				⋮	⋮		⋮		⋮	⋮	⋮
i				$d_{i,0}$...	$d_{i,j-i}$...	$d_{i,s-i}$...	$d_{i,T-i}$	r_i	N_i
				$c_{i,0}$...	$c_{i,j-i}$...	$c_{i,s-i}$...	$c_{i,T-i}$		
⋮							⋮		⋮	⋮	⋮
S							$d_{s,0}$...	$d_{s,T-s}$	r_s	N_s
							$c_{s,0}$...	$c_{s,T-s}$		
총 고장수	d_1	d_2	d_3	...	d_j	...	d_s	...	d_T	d	N

$$d_j = \sum_{i=1}^{\min(S,j)} d_{i,j-i} \quad (j = 1, \dots, T) \quad (1)$$

단, $R_{-1} \equiv 1$.

최종적으로 관측 중단되는 T 기까지 작동 중인 제품수는 식 (2)와 같으며,

$$r_i = N_i - \sum_{t=0}^{T-i} (d_{i,t} + c_{i,t}) \quad (i = 1, \dots, S) \quad (2)$$

각 판매 시기별로 수명 t 기 직전까지 작동 중인 제품수는 식 (3)과 같이 표현된다.

$$n_{i,0} = N_i, \quad n_{i,t} = N_i - \sum_{k=0}^{t-1} (d_{i,k} + c_{i,k}) \quad (t = 1, \dots, T-i, \quad i = 1, \dots, S) \quad (3)$$

2.2 완전 데이터 우도함수(likelihood function)

<표 1>에서 기별 고장수(d_1, \dots, d_T)의 결합분포는 표현하기에 너무 복잡하다. 그러나 모든 셀 고장수($d_{i,t}$)를 안다고 가정하면 우도함수를 유도할 수 있다. 제품(컴포넌트)의 수명을 X 라 하고, 그 분포함수를 $F(x)$ 라 하자. 사용 개시 후 t 기에서의 신뢰도 함수를 $R_t = \Pr(X > t) = 1 - F(t)$ 로 나타내면 주어진 가정에 근거하여 우도함수는 상수항을 제외하고 다음과 같이 표현할 수 있다.

$$L(R_0, \dots, R_{T-1}) = \prod_{i=1}^S \left[R_{T-i}^{r_i} \prod_{t=0}^{T-i} \left\{ (R_{t-1} - R_t)^{d_{i,t}} R_t^{c_{i,t}} \right\} \right] \quad (4)$$

NHPP 모형과의 비교를 위하여 다른 표현을 유도해 보자. t 기에서의 고장 확률을 q_t 라 하면, $R_t = 1 - \sum_{k=0}^t q_k$ 이므로 우도함수는 다음과 같이 나타낼 수 있으나, 분석하기 어려운 형태임을 알 수 있다.

$$L(q_0, \dots, q_{T-1}) = \prod_{i=1}^S \left[\left(1 - \sum_{t=0}^{T-i} q_t \right)^{r_i} \prod_{t=0}^{T-i} \left\{ q_t^{d_{i,t}} \left(1 - \sum_{k=0}^t q_k \right)^{c_{i,t}} \right\} \right] \quad (5)$$

이에 반하여 Karim et al.(2001)이 제안한 NHPP 모형의 우도함수는 다음과 같이 간단한 형태를 갖는다. 단, 이때의 q_t 는 고장 강도(intensity)의 의미를 갖는다.

$$L(q_0, \dots, q_{T-1}) = \prod_{i=1}^S \prod_{t=0}^{T-i} \left[(N_i q_t)^{d_{i,t}} \exp(-N_i q_t) \right]$$

분석이 용이한 형태의 우도함수를 유도하기 위하여 다음과 같이 조건부 고장확률, 혹은 고장률을 정의하자.

$$h_t = P(X = t | X > t-1) = 1 - R_t / R_{t-1} \quad (t = 0, \dots, T-1) \quad (6)$$

단, $R_{-1} \equiv 1$.

이와 같이 새로 정의된 고장률 모수를 통하여 우도함수를 나타낸 결과는 [정리 1]과 같다.

[정리 1] 식 (4)나 (5)의 완전 데이터 우도함수는 식 (7)과 같

이 나타낼 수 있다.

$$L(h_0, \dots, h_{T-1}) = \prod_{i=1}^S \prod_{t=0}^{T-i} [(1-h_t)^{n_{i,t}-d_{i,t}} h_t^{d_{i,t}}] \quad (7)$$

단, $n_{i,t}$ 는 식(3)과 같다.

(증명) 부록 참조.

식 (7)로부터 완전 데이터 대수우도함수는 다음과 같이 미분하기 쉬운 식으로 표현된다.

$$\begin{aligned} \ln L(h_0, \dots, h_{T-1}) \\ = \sum_{i=1}^S \sum_{t=0}^{T-i} [(n_{i,t}-d_{i,t}) \ln(1-h_t) + d_{i,t} \ln(h_t)] \quad (8) \end{aligned}$$

3. 모수 추정 알고리즘

앞 장에서 살펴본 바와 같이 관측 데이터의 우도함수는 식으로 표현할 수 없고, 완전 데이터의 우도함수는 표현 가능한 경우에는 EM 알고리즘을 통해 최우추정량을 구하는 것이 바람직하다. EM 원리는 초기해에서 시작해서 다음의 기대화(expectation) 단계와 최대화(maximization) 단계를 반복적으로 수행해 가면서 확률모수의 고정점(fixed point)을 찾아 최우추정치를 구한다(Dempster et al., 1977).

3.1 기대화 단계

주어진 주변 데이터 $\vec{d} = (d_1, \dots, d_T)$, $\vec{N} = (N_1, \dots, N_S)$ 와 이전 단계의 모수 추정치 $\vec{h}^{(m)} = (h_0^{(m)}, \dots, h_{T-1}^{(m)})$ 을 바탕으로 대수우도함수의 조건부 기대값을 계산한다. 식 (8)에 주어진 대수우도함수를 살펴보면, 셀 고장수 $d_{i,t}$ 의 조건부 기대값이 요구됨을 알 수 있다.

먼저 고장률 $\vec{h}^{(m)}$ 과 기별 판매대수 \vec{N} 이 주어진 경우, 셀 고장수 $d_{i,t}$ 의 기대값은 다음과 같이 순차적으로 계산할 수 있다.

$$\begin{aligned} E[n_{i,0} | \vec{N}, \vec{h}^{(m)}] &= N_i, \quad (i = 1, \dots, S) \\ E[d_{i,0} | \vec{N}, \vec{h}^{(m)}] &= N_i \cdot h_0^{(m)}, \quad (i = 1, \dots, S) \\ E[n_{i,t} | \vec{N}, \vec{h}^{(m)}] \\ &= N_i - \sum_{k=0}^{t-1} (E[d_{i,k} | \vec{N}, \vec{h}^{(m)}] + c_{i,k}) \quad (9) \\ E[d_{i,t} | \vec{N}, \vec{h}^{(m)}] &= E[n_{i,t} | \vec{N}, \vec{h}^{(m)}] \cdot h_t^{(m)} \end{aligned}$$

각 관측기별 고장대수 $\vec{d} = (d_1, \dots, d_T)$ 가 주어지면 식 (1)에 의하여 j -열에 속한 셀 고장수 $d_{i,j-i}$ 는 제약식 $d_j = \sum_{i=1}^{\min(S,j)} d_{i,j-i}$ 을 만족해야 한다. 일반적으로 $d_{i,j-i}$ 의 조건부 기대값은 표현하기 어렵다(Karim et al., 2001). 그러나 판매대

수 \vec{N} 가 고장대수 \vec{d} 보다 충분히 크다면, 동일한 관측기 내에서 고장대수 간의 상관관계는 무시할 만큼 작으므로 다음과 같이 근사적으로 계산할 수 있다.

$$\begin{aligned} d_{i,j-i}^{(m)} &\equiv E[d_{i,j-i} | \vec{N}, \vec{d}, \vec{h}^{(m)}] \\ &\simeq d_j \left(\frac{E[d_{i,j-i} | \vec{N}, \vec{h}^{(m)}]}{\sum_{i=1}^{\min(S,j)} E[d_{i,j-i} | \vec{N}, \vec{h}^{(m)}]} \right) \quad (10) \\ &(i = 1, \dots, \min(S, j), \quad j = 1, \dots, T) \end{aligned}$$

식 (2)와 (3)으로부터 $n_{i,t}$ 와 r_i 의 조건부 기대값은 아래와 같이 계산한다.

$$n_{i,t}^{(m)} \equiv E[n_{i,t} | \vec{N}, \vec{d}, \vec{h}^{(m)}] = N_i - \sum_{k=0}^{t-1} (d_{i,k}^{(m)} + c_{i,k}) \quad (11)$$

$$\begin{aligned} r_i^{(m)} \equiv E[r_i | \vec{N}, \vec{d}, \vec{h}^{(m)}] &= N_i - \sum_{t=0}^{T-i} (d_{i,t}^{(m)} + c_{i,t}) \\ &(t = 0, \dots, T-i, \quad i = 1, \dots, S) \end{aligned}$$

3.2 최대화 단계

식 (8)의 대수우도함수에서 $d_{i,t}$ 와 $n_{i,t}$ 를 각각 조건부 기대값 $d_{i,t}^{(m)}$ 과 $n_{i,t}^{(m)}$ 로 치환하면 조건부 대수우도함수의 기대값을 얻는다. 이를 확률 모수에 대하여 최대화하여 다음 단계에 사용될 모수값을 계산한다. 식 (8)로부터

$$\begin{aligned} Q(\vec{h} | \vec{h}^{(m)}) &\equiv E[\ln L(\vec{h}) | \vec{N}, \vec{d}, \vec{h}^{(m)}] \\ &= \sum_{i=1}^S \sum_{t=0}^{T-i} [(n_{i,t}^{(m)} - d_{i,t}^{(m)}) \ln(1-h_t) + d_{i,t}^{(m)} \ln(h_t)] \end{aligned}$$

이를 확률 모수 h_t 에 대해서 편미분하면,

$$\begin{aligned} \frac{\partial Q}{\partial h_t} &= \sum_{i=1}^{\min(S, T-t)} \left[\frac{-(n_{i,t}^{(m)} - d_{i,t}^{(m)})}{1-h_t} + \frac{d_{i,t}^{(m)}}{h_t} \right] \equiv 0 \\ &(t = 0, \dots, T-1) \end{aligned}$$

해를 구하면 다음과 같이 간단한 형태의 고장률 추정치를 얻는다.

$$h_t^{(m+1)} = \frac{\sum_{i=1}^{\min(S, T-t)} d_{i,t}^{(m)}}{\sum_{i=1}^{\min(S, T-t)} n_{i,t}^{(m)}}, \quad (t = 0, \dots, T-1) \quad (12)$$

이는 마치 구간 생존 데이터에 Kaplan-Meier 추정량(Kaplan and Meier, 1958)을 일반화하여 적용한 형태를 갖는다.

3.3 EM 알고리즘

이상의 과정을 알고리즘으로 정리하면 다음과 같다.

초기화 단계

고장률 모수 값 $h_0^{(0)}, h_1^{(0)}, \dots, h_{T-1}^{(0)}$ 할당.

[대안 1] 주변 데이터를 이용하여 총 판매대수 $N = \sum_{i=1}^S N_i$ 와 총 고장대수 $d = \sum_{i=0}^{T-1} d_i$ 및 총 관측 중단대수 $c = \sum_{i=1}^S \sum_{t=0}^{T-i} c_{i,t}$ 로부터 $h_i^{(0)} = d/[T(N-c)]$, ($0 \leq t \leq T-1$)로 균일하게 할당한다.

[대안 2] 각 셀의 고장수 $d_{i,t}$ 가 기대값 $n_{i,t} h_t$ 와 같다는 이상적인 경우를 가정하여 다음과 같이 계산한다. 단, 식 (14)에서 $\tilde{d}_{1,t}$ 는 음수가 나올 수 있기 때문에 그런 경우에는 식 (13)과 같이 평균값으로 지정한다.

$$h_i^{(0)} = \begin{cases} \tilde{d}_{1,t} / \tilde{n}_{1,t}, & \text{if } \tilde{d}_{1,t} \geq 0, \\ d/[T(N-c)], & \text{if } \tilde{d}_{1,t} < 0, \end{cases} \quad (t = 0, \dots, T-1) \quad (13)$$

단, $\tilde{d}_{1,0} = d_1, \tilde{n}_{1,0} = N_1$,

$$\tilde{d}_{i,j-i} = \begin{cases} d_j - \sum_{l=2}^{\min(S,j)} \tilde{d}_{l,j-l}, & i=1, \\ \tilde{n}_{i,j-i} h_{j-i}^{(0)}, & i=2, \dots, \min(S,j) \end{cases} \quad (j = 2, \dots, T) \quad (14)$$

$$\tilde{n}_{i,t} = N_1 - \sum_{k=0}^{t-1} (\tilde{d}_{i,k} + c_{i,k}) \quad (t = 0, \dots, T-i, i = 1, \dots, S).$$

반복연산

다음의 기대화 단계와 최대화 단계를 수렴조건이 성립될 때까지 수행한다.

기대화 단계

- ① 전 단계에서 계산된 고장률 $\tilde{h}^{(m)}$ 에 기초하여 식 (9)와 같이 셀 고장수 및 초기 생존수 계산(비제약적).
- ② 식 (10)과 같이 각 셀 고장수의 근사적 조건부 기대값 $d_{i,t}^{(m)}$ 계산(제약적).
- ③ 식 (11)과 같이 기별 초기 생존수 $n_{i,t}^{(m)}$ 재계산.

최대화 단계

- ④ 식 (12)와 같이 고장률 계산

수렴조건 검토

다음과 같이 확률 모수값의 변화가 미미하면 고정점으로 수렴한 것으로 판단한다.

$$\frac{|\tilde{h}^{(m+1)} - \tilde{h}^{(m)}|^{1/2}}{|\tilde{h}^{(m)}|^{1/2}} < \epsilon, \quad t = 0, \dots, T-1$$

4. 모의실험

본 연구의 가장 큰 특징은 관측 중단을 고려한 것이라 할 수 있다. 관측 중단 데이터는 존재하기 마련이지만, 관측하기 어려운 것이 사실이다. 그렇다고 해서 관측 중단을 고려하지 않고 분석하면 낙관적인 예측을 하게 되어 신뢰도를 과대평가하는 결과를 초래한다. 제품의 사용비율이 알려진 경우에 대하여 다음과 같이 관측 중단 데이터가 주어졌다고 전제하고 다양한 경우에 대한 모의실험을 실시하였다.

$$c_{i,t} = N_i(1 - \gamma_{t-i}),$$

단 γ_k = 판매 이후 k 기에서의 제품 사용비율

모든 경우에 있어서 판매 기수 $S = 3$, 관측 기수 $T = 5$ 로 하였고, 기별 판매 제품수 $\vec{N} = (10000, 15000, 25000)$ 으로 정하였다. 또한 모의실험은 각각 1000회 반복으로 실시하였다. 기별 고장률 및 사용비율은 다음과 같이 설정하여 고장대수를 생성하였다.

$$h_t = (t+1)^{\alpha-1} \times 10^{-\delta}, \quad (t = 0, \dots, 4)$$

$$\gamma_t = 1 - t \times \gamma, \quad (t = 0, \dots, 4)$$

입력모수에 따른 실험번호는 <표 2>와 같다.

여기서 입력모수 α 는 고장률의 형태를 결정하며, α 가 1보다 작으면 감소형, 1이면 일정형, 1보다 크면 증가형 고장률을 나타낸다. 입력모수 δ 는 고장률의 크기를 결정하며, δ 값이 커질수록 작은 고장률을 나타낸다. 입력모수 γ 는 제품의 사용비율을 결정하며, γ 값이 커질수록 사용비율이 낮아지고, 관측 중단 효과가 커지게 된다.

표 2. 모의실험에 사용된 입력모수 및 실험번호

실험번호	alpha	delta	gamma
111	0.5	2	0.0
222	1.0	3	0.01
333	2.0	4	0.05
4*4	3.0	-	0.1
**5	-	-	0.15

4.1 관측 중단을 고려하지 않은 경우의 편이(bias)

실제로 관측 중단 데이터가 존재하는 경우에 관측 중단을 고려하지 않고 분석함으로써 발생하는 고장률 추정치의 편이를 모의실험을 통해 알아보았다. δ 는 2로 하고, α 는 0.5, 1, 2 각각의 경우에 대해 γ 를 0부터 0.15까지 5단계로 변화시켜 가며 실험하여 얻은 결과는 <표 3>과 같으며, 상대적 편이는 <그림 1>에서 <그림 3>까지와 같다. 표에서 'TRUE'는 실제 고장

표 3. 관측 중단을 고려하지 않은 경우의 고장률 추정 결과

	h_0	h_1	h_2	h_3	h_4		h_0	h_1	h_2	h_3	h_4
TRUE	0.0100	0.0071	0.0058	0.0050	0.0045						
111	0.0099	0.0073	0.0054	0.0049	0.0056	s111	0.0008	0.0015	0.0027	0.0041	0.0058
112	0.0100	0.0074	0.0052	0.0048	0.0055	s112	0.0008	0.0015	0.0026	0.0038	0.0059
113	0.0100	0.0075	0.0048	0.0042	0.0048	s113	0.0008	0.0014	0.0026	0.0037	0.0054
114	0.0099	0.0075	0.0045	0.0033	0.0037	s114	0.0007	0.0013	0.0023	0.0032	0.0046
115	0.0099	0.0076	0.0039	0.0027	0.0030	s115	0.0007	0.0012	0.0020	0.0029	0.0042
b111	-0.6%	3.9%	-6.9%	-1.4%	25.2%	cv111	0.0786	0.2011	0.4934	0.8256	1.0333
b112	-0.4%	4.0%	-9.5%	-3.5%	23.1%	cv112	0.0778	0.2046	0.4881	0.7905	1.0666
b113	-0.5%	5.6%	-16.6%	-15.9%	7.5%	cv113	0.0781	0.1927	0.5333	0.8804	1.1226
b114	-0.5%	5.4%	-21.7%	-33.4%	-18.1%	cv114	0.0736	0.1686	0.5016	0.9623	1.2515
b115	-0.9%	8.1%	-32.7%	-46.7%	-32.4%	cv115	0.0719	0.1523	0.5248	1.0764	1.3914
TRUE	0.0100	0.0100	0.0100	0.0100	0.0100						
211	0.0099	0.0102	0.0099	0.0096	0.0106	s211	0.0009	0.0019	0.0031	0.0053	0.0081
212	0.0099	0.0102	0.0098	0.0093	0.0102	s212	0.0010	0.0019	0.0032	0.0053	0.0082
213	0.0099	0.0103	0.0092	0.0081	0.0082	s213	0.0009	0.0018	0.0031	0.0052	0.0076
214	0.0099	0.0104	0.0085	0.0064	0.0056	s214	0.0008	0.0016	0.0030	0.0047	0.0063
215	0.0098	0.0107	0.0077	0.0047	0.0038	s215	0.0008	0.0014	0.0027	0.0040	0.0053
b211	-0.5%	1.5%	-0.6%	-3.7%	6.2%	cv211	0.0913	0.1862	0.3156	0.5456	0.7606
b212	-0.7%	1.7%	-2.1%	-6.8%	2.4%	cv212	0.0962	0.1884	0.3266	0.5707	0.7962
b213	-1.1%	2.8%	-7.8%	-19.0%	-17.6%	cv213	0.0928	0.1786	0.3365	0.6375	0.9202
b214	-1.2%	4.2%	-14.7%	-36.0%	-44.1%	cv214	0.0837	0.1518	0.3554	0.7318	1.1234
b215	-1.6%	6.8%	-23.0%	-53.3%	-62.1%	cv215	0.0809	0.1308	0.3530	0.8497	1.3857
TRUE	0.0100	0.0200	0.0300	0.0400	0.0500						
311	0.0100	0.0200	0.0299	0.0400	0.0500	s311	0.0010	0.0024	0.0043	0.0074	0.0132
312	0.0101	0.0199	0.0295	0.0391	0.0464	s312	0.0010	0.0024	0.0040	0.0074	0.0126
313	0.0099	0.0202	0.0283	0.0335	0.0342	s313	0.0010	0.0024	0.0042	0.0072	0.0124
314	0.0100	0.0201	0.0269	0.0268	0.0178	s314	0.0010	0.0022	0.0041	0.0071	0.0113
315	0.0098	0.0207	0.0245	0.0195	0.0059	s315	0.0009	0.0018	0.0039	0.0066	0.0072
b311	0.3%	0.2%	-0.5%	0.0%	0.0%	cv311	0.0985	0.1220	0.1434	0.1859	0.2648
b312	0.9%	-0.7%	-1.5%	-2.1%	-7.1%	cv312	0.0997	0.1208	0.1371	0.1879	0.2715
b313	-0.7%	1.0%	-5.7%	-16.3%	-31.7%	cv313	0.1015	0.1195	0.1478	0.2155	0.3638
b314	-0.3%	0.3%	-10.2%	-33.0%	-64.4%	cv314	0.0976	0.1107	0.1507	0.2634	0.6369
b315	-1.5%	3.6%	-18.2%	-51.2%	-88.1%	cv315	0.0908	0.0848	0.1570	0.3387	1.2180

Relative Bias (alpha=0.5, delta=2)

Relative Bias (alpha=2.0, delta=2)

그림 1. 감소 고장률하에서 관측 중단을 고려하지 않은 경우의 상대적 편이.

그림 3. 증가 고장률하에서 관측 중단을 고려하지 않은 경우의 상대적 편이.

그림 2. 일정 고장률하에서 관측 중단을 고려하지 않은 경우의 상대적 편이.

4.2 관측 중단 비율에 따른 추정의 정확도 및 정밀도

물, 'ijk'는 실험번호 'ijk'에 해당하는 고장률 추정치의 평균값, 'sijk'는 그 표준편차, 'bijk'는 상대적 편이, 'cvijk'는 변동계수(coefficient of variation)를 나타낸다.

모의실험을 통해 각각의 고장률 유형에 대하여 관측 중단 비율에 따른 추정의 정밀도 및 정확도를 살펴보았다. $\delta = 2$ 로 고정시키고, $\alpha = 0.5, 1, 2$ 각각의 경우에 대하여 $\gamma = 0, 0.0, 0.1, 0.05, 0.1$ 로 변화시켜 가며 실험한 결과를 <표 4>에 정리하였다. 표에서 'Uijk'와 'Lijk'는 각각 실험번호 'ijk'에 해당하는 고장률에 대한 90% 근사 신뢰구간의 상한과 하한을 나타낸다.

먼저 $\alpha = 0.5$ 인 감소 고장률하에서의 상대적 편이를 살펴보면(<그림 1>) 관측 중단이 없더라도(b111) 마지막 기의 고장률 h_4 는 20% 이상 과대평가된다. 그러나 사용비율이 줄어들며 따라 중기 이후의 고장률 h_2, h_3, h_4 는 모두 과소평가 되며, 특히 h_3 는 그 현상이 심하게 나타났다.

대체적으로 후반부로 갈수록 고장률에 대한 편이와 변동계수가 모두 증가하였는데, 이는 후반부의 관측 데이터 수가 적어지기 때문인 것으로 생각된다. 특히 고장률이 감소형인 경우 이러한 현상은 심각하게 나타났으며, 고장률이 증가형인 경우에는 관측 중단의 영향을 상대적으로 덜 받는 것으로 나타났다.

다음으로 $\alpha = 1$ 인 일정 고장률하에서의 결과를 살펴보면(<그림 2>) 관측 중단이 없는 경우에서의(b211) 마지막 기의 고장률 h_4 에 대한 과대평가 현상은 10% 미만으로 감소하였으며, 사용비율의 감소에 따른 후반기의 고장률 h_3 와 h_4 에 대한 과소평가 현상이 더욱 심하게 나타났다.

추정의 정밀도와 정확도를 동시에 비교하기 위해 고장률 추정치의 평균과 근사적 95% 상측 신뢰한계를 $\alpha = 0.5, 1, 2$ 각각의 경우에 대하여 <그림 4>에서 <그림 6>까지 나타내었다. 모든 경우에 있어서 관측 중단 비율의 증가에 따라 후반기 고장률에 대한 추정의 편이가 크게 나타나고, 정밀도 또한 크게 감소하는 것으로 나타났다.

마지막으로, $\alpha = 2$ 인 증가 고장률하에서의 결과를 살펴보면(<그림 3>) 관측 중단이 없는 경우에서의(b311) 모든 기의 고장률에 대한 편이는 거의 나타나지 않았으나, 사용비율의 감소에 따른 후반기의 고장률 h_3 와 h_4 에 대한 과소평가 현상이 심하게 나타났으며, 특히 마지막 기의 고장률 h_4 의 경우에는 γ 값이 5%(b313)에 불과하더라도 30% 이상 과소평가하게 되는 것으로 나타났다.

Approximate 95% Upper Confidence Interval (alpha=0.5, delta=2)

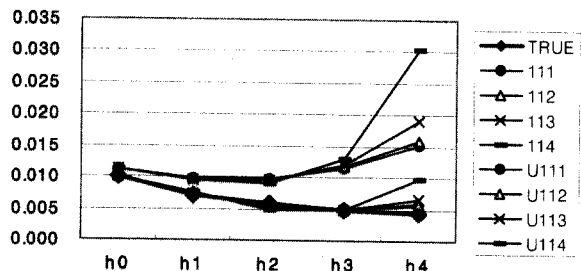


그림 4. 감소형 고장률하에서 사용비율에 따른 근사적 95% 신뢰상한.

표 4. 관측 중단 비율에 따른 모의실험 결과

	h0	h1	h2	h3	h4						
TRUE	0.0100	0.0071	0.0058	0.0050	0.0045						
111	0.0099	0.0073	0.0054	0.0049	0.0056	s111	0.0008	0.0015	0.0027	0.0041	0.0058
112	0.0100	0.0073	0.0053	0.0050	0.0059	s112	0.0008	0.0015	0.0026	0.0040	0.0061
113	0.0099	0.0074	0.0053	0.0049	0.0067	s113	0.0008	0.0014	0.0026	0.0044	0.0076
114	0.0099	0.0075	0.0049	0.0049	0.0097	s114	0.0008	0.0013	0.0025	0.0049	0.0125
U111	0.0112	0.0098	0.0097	0.0116	0.0151	b111	0.61%	3.90%	6.94%	1.41%	25.21%
U112	0.0113	0.0097	0.0096	0.0116	0.0160	b112	0.01%	3.55%	8.38%	0.40%	31.78%
U113	0.0113	0.0097	0.0095	0.0121	0.0192	b113	0.74%	4.62%	8.73%	2.68%	50.11%
U114	0.0112	0.0096	0.0091	0.0129	0.0303	b114	0.59%	6.03%	14.28%	2.14%	117.12%
L111	0.0087	0.0049	0.0010	0.0000	0.0000	cv111	0.0786	0.2011	0.4934	0.8256	1.0333
L112	0.0087	0.0049	0.0010	0.0000	0.0000	cv112	0.0770	0.2002	0.4903	0.8126	1.0421
L113	0.0086	0.0051	0.0010	0.0000	0.0000	cv113	0.0817	0.1857	0.4884	0.9020	1.1284
L114	0.0087	0.0054	0.0008	0.0000	0.0000	cv114	0.0773	0.1733	0.5144	0.9923	1.2878
TRUE	0.0100	0.0100	0.0100	0.0100	0.0100						
211	0.0099	0.0102	0.0099	0.0096	0.0106	s211	0.0009	0.0019	0.0031	0.0053	0.0081
212	0.0100	0.0102	0.0097	0.0100	0.0109	s212	0.0009	0.0018	0.0032	0.0056	0.0085
213	0.0099	0.0103	0.0098	0.0094	0.0119	s213	0.0009	0.0018	0.0031	0.0057	0.0107
214	0.0100	0.0103	0.0095	0.0096	0.0145	s214	0.0008	0.0016	0.0033	0.0068	0.0166
U211	0.0114	0.0133	0.0151	0.0183	0.0239	b211	0.54%	1.52%	0.56%	3.67%	6.19%
U212	0.0115	0.0132	0.0150	0.0192	0.0249	b212	0.15%	1.74%	2.98%	0.23%	9.09%
U213	0.0114	0.0132	0.0149	0.0188	0.0296	b213	0.89%	2.73%	1.91%	5.97%	18.87%
U214	0.0113	0.0130	0.0150	0.0209	0.0418	b214	0.50%	3.28%	5.44%	3.95%	44.85%
L211	0.0085	0.0070	0.0048	0.0010	0.0000	cv211	0.0913	0.1862	0.3156	0.5456	0.7606
L212	0.0085	0.0071	0.0044	0.0008	0.0000	cv212	0.0918	0.1813	0.3296	0.5595	0.7787
L213	0.0085	0.0073	0.0047	0.0000	0.0000	cv213	0.0895	0.1747	0.3158	0.6106	0.9034
L214	0.0086	0.0077	0.0039	0.0000	0.0000	cv214	0.0838	0.1561	0.3540	0.7128	1.1458
TRUE	0.0100	0.0200	0.0300	0.0400	0.0500						
311	0.0100	0.0200	0.0299	0.0400	0.0500	s311	0.0010	0.0024	0.0043	0.0074	0.0132
312	0.0101	0.0198	0.0300	0.0403	0.0494	s312	0.0010	0.0023	0.0041	0.0074	0.0139
313	0.0100	0.0199	0.0300	0.0403	0.0491	s313	0.0010	0.0025	0.0043	0.0087	0.0188
314	0.0100	0.0201	0.0299	0.0393	0.0518	s314	0.0010	0.0023	0.0045	0.0104	0.0322
U311	0.0117	0.0241	0.0369	0.0522	0.0718	b311	0.32%	0.22%	0.46%	0.00%	0.01%
U312	0.0117	0.0237	0.0369	0.0524	0.0723	b312	0.59%	0.85%	0.15%	0.65%	1.11%
U313	0.0117	0.0240	0.0371	0.0546	0.0800	b313	0.24%	0.29%	0.12%	0.82%	1.86%
U314	0.0115	0.0239	0.0374	0.0564	0.1047	b314	0.42%	0.66%	0.20%	1.77%	3.59%
L311	0.0084	0.0160	0.0228	0.0278	0.0282	cv311	0.0985	0.1220	0.1434	0.1859	0.2648
L312	0.0084	0.0160	0.0232	0.0281	0.0266	cv312	0.0995	0.1184	0.1377	0.1830	0.2813
L313	0.0083	0.0159	0.0229	0.0261	0.0182	cv313	0.1029	0.1241	0.1439	0.2151	0.3825
L314	0.0084	0.0164	0.0225	0.0222	0.0000	cv314	0.0971	0.1135	0.1510	0.2651	0.6208

Approximate 95% Upper Confidence Interval (alpha=1.0, delta=2)

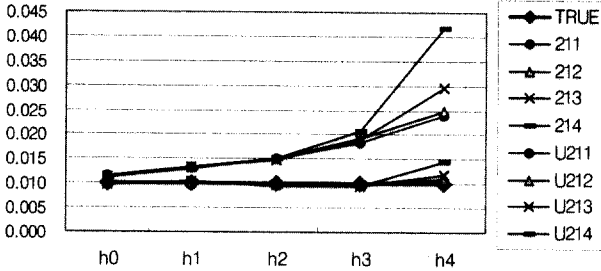


그림 5. 일정형 고장률하에서 사용비율에 따른 근사적 95% 신뢰상한.

Approximate 95% Upper Confidence Interval (alpha=2, delta=2)

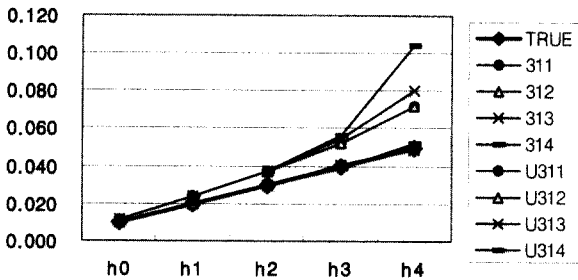


그림 6. 증가형 고장률하에서 사용비율에 따른 근사적 95% 신뢰상한.

4.3 고장률의 크기에 따른 추정의 정밀도 및 정확도

다음으로 모의실험을 통해 고장률의 상대적 크기가 추정의 정밀도 및 정확도에 미치는 영향을 살펴보았다. $\alpha = 2$,

한 결과는 <표 5>와 같다. 표를 살펴보면 고장률이 작을수록 (δ 가 커질수록) 상대적 편이와 변동계수가 모두 증가하여 추정의 정밀도 및 정확도가 모두 낮아짐을 알 수 있다. 이는 판매량이 고정된 경우, 고장률이 낮을수록 발생하는 고장 데이터의 수가 감소하기 때문인 것으로 판단된다.

4.4 사례 연구

본 연구의 결과가 실제 데이터에 적용 가능한지를 알아보기 위해 국내 모 기업에서 생산한 진공청소기 모터에 대한 사용현장 보증 데이터를 이용하여 분석하였다. 시간 단위는 개월로 정하였으며, 생산기간 및 관측기간 모두 12개월로 한정하여 데이터를 정리하였다.

$$\vec{N} = (34360, 21637, 27238, 31993, 20803, 16016, 24803, 22794, 47781, 27137, 38307, 29756)$$

$$\vec{d} = (0, 1, 6, 11, 10, 15, 18, 31, 26, 38, 62, 90)$$

관측 중단에 대한 데이터는 구할 수 없어서 다음과 같이 다양한 경우를 가정하여 분석한 결과를 <표 6>에 정리하였다. <그림 7>에는 관측 중단 유형에 따른 고장률 추정치의 추이를 나타내었다.

- ① $c_{i,t} = 0$, 관측 중단이 없는 경우 ($t = 0, \dots, T - i$, $i = 1, \dots, S$)
- ② $c_{i,t} = N_i/1000$, 매월 판매량의 0.1%가 관측 중단되는 경우
- ③ $c_{i,t} = N_i/100$, 매월 판매량의 1.0%가 관측 중단되는 경우
- ④ $c_{i,t} = N_i/20$, 매월 판매량의 5.0%가 관측 중단되는 경우

표 5. 고장률의 크기에 따른 모의실험 결과

	h0	h1	h2	h3	h4						
TRUE	0.0100	0.0200	0.0300	0.0400	0.0500						
312	0.0101	0.0198	0.0300	0.0403	0.0494	s312	0.0010	0.0023	0.0041	0.0074	0.0139
U312	0.0117	0.0237	0.0369	0.0524	0.0723	b312	0.59%	0.85%	0.15%	0.65%	1.11%
L312	0.0084	0.0160	0.0232	0.0281	0.0266	cv312	0.0995	0.1184	0.1377	0.1830	0.2813
TRUE	0.0010	0.0020	0.0030	0.0040	0.0050						
322	0.0010	0.0020	0.0030	0.0039	0.0053	s322	0.0003	0.0007	0.0013	0.0021	0.0036
U322	0.0015	0.0032	0.0051	0.0073	0.0113	b322	0.16%	0.90%	0.46%	3.31%	5.97%
L322	0.0005	0.0009	0.0009	0.0004	0.0000	cv322	0.3034	0.3509	0.4322	0.5463	0.6851
TRUE	0.0001	0.0002	0.0003	0.0004	0.0005						
332	0.0001	0.0002	0.0003	0.0004	0.0006	s332	0.0001	0.0002	0.0002	0.0004	0.0007
U332	0.0002	0.0005	0.0007	0.0011	0.0017	b332	11.94%	8.57%	10.27%	3.85%	17.84%
L332	0.0000	0.0000	0.0000	0.0000	0.0000	cv332	0.9209	0.7407	0.9217	1.0287	1.1173

$\gamma = 0.01$ 로 고정시키고 $\delta = 2, 3, 4$ 로 변화시켜 가며 실험

표 6. 관측 중단 유형에 따른 고장모수 추정 결과($h, q: \times 10^{-3}$)

유형	모수	0	1	2	3	4	5	6	7	8	9	10	11
①	h	0.00000	0.02908	0.15585	0.19931	0.01565	0.10559	0.15039	0.49201	0.00000	0.21759	0.68731	1.13021
	q	0.00000	0.02908	0.15585	0.19927	0.01564	0.10555	0.15032	0.49168	0.00000	0.21734	0.68637	1.12789
	R	1.00000	0.99997	0.99982	0.99962	0.9996	0.99950	0.99934	0.99885	0.99885	0.9986	0.99795	0.99682
②	h	0.00000	0.02911	0.15616	0.19991	0.01571	0.10612	0.15130	0.49548	0.00000	0.21957	0.69426	1.14281
	q	0.00000	0.02911	0.15616	0.19987	0.01570	0.10608	0.15123	0.49515	0.00000	0.21931	0.69331	1.14044
	R	1.00000	0.99997	0.99982	0.99962	0.99960	0.99949	0.99934	0.99885	0.99885	0.99863	0.99793	0.99679
③	h	0.00000	0.02937	0.15903	0.20548	0.01630	0.11115	0.16000	0.52907	0.00000	0.23913	0.76379	1.27022
	q	0.00000	0.02937	0.15903	0.20544	0.01629	0.11111	0.15992	0.52871	0.00000	0.23884	0.76268	1.26741
	R	1.00000	0.99997	0.99981	0.99961	0.99959	0.99948	0.99932	0.99879	0.99879	0.99855	0.99779	0.99652
④	h	0.00000	0.03061	0.17317	0.23449	0.01956	0.14081	0.21490	0.75720	0.00000	0.39598	1.37649	2.51793
	q	0.00000	0.03061	0.17316	0.23444	0.01955	0.14074	0.21477	0.75659	0.00000	0.39536	1.37379	2.50952
	R	1.00000	0.99997	0.99980	0.99956	0.99954	0.99940	0.99919	0.99843	0.99843	0.99804	0.99666	0.99415
⑤	h	0.00000	0.02911	0.15632	0.20051	0.01580	0.10720	0.15362	0.50619	0.00000	0.22785	0.72737	1.2103
	q	0.00000	0.02911	0.15631	0.20048	0.01580	0.10716	0.15354	0.50585	0.00000	0.22759	0.72635	1.20777
	R	1.00000	0.99997	0.99982	0.99961	0.99960	0.99949	0.99934	0.99883	0.99883	0.99860	0.99788	0.99667
⑥	h	0.00000	0.02911	0.15663	0.20214	0.01613	0.11174	0.16546	0.57216	0.00000	0.30445	1.11853	2.29271
	q	0.00000	0.02911	0.15663	0.20210	0.01612	0.11169	0.16537	0.57177	0.00000	0.30407	1.11679	2.28659
	R	1.00000	0.99997	0.99981	0.99961	0.99960	0.99948	0.99932	0.99875	0.99875	0.99844	0.99733	0.99504
⑦	h	0.00000	0.02943	0.15952	0.20611	0.01633	0.11115	0.15949	0.52511	0.00000	0.23449	0.74312	1.22471
	q	0.00000	0.02943	0.15951	0.20608	0.01632	0.11110	0.15941	0.52475	0.00000	0.23421	0.74205	1.22203
	R	1.00000	0.99997	0.99981	0.99961	0.99959	0.99948	0.99932	0.99879	0.99879	0.99856	0.99782	0.99660
⑧	h	0.00000	0.03397	0.21204	0.31391	0.02825	0.21558	0.34125	1.21601	0.00000	0.59896	1.94089	3.23230
	q	0.00000	0.03397	0.21204	0.31383	0.02823	0.21545	0.34098	1.21462	0.00000	0.59755	1.93515	3.21649
	R	1.00000	0.99997	0.99975	0.99944	0.99941	0.99920	0.99886	0.99764	0.99764	0.99704	0.99511	0.99189

관측중단 유형에 따른 고장률 추정치

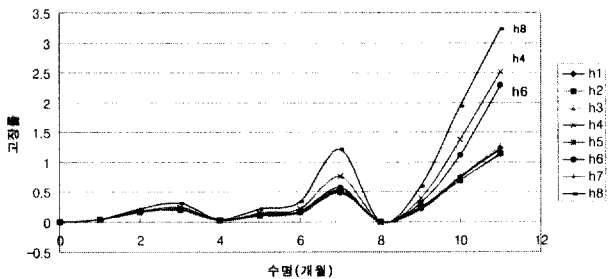


그림 7. 관측 중단 유형에 따른 고장률 추정치의 추이.

- ⑤ $c_{i,t} = N_i(t+1)/1000$, 관측 중단수가 관측기수에 비례하여 증가하는 경우
- ⑥ $c_{i,t} = N_i(t+1)^2/1000$, 관측 중단수가 관측기수의 제곱에 비례하여 증가하는 경우
- ⑦ $c_{i,t} = N_i(12-t)/1000$, 관측 중단수가 관측기수에 비례하여 감소하는 경우

⑧ $c_{i,t} = N_i(12-t)^2/1000$, 관측 중단수가 관측기수에 제곱에 비례하여 감소하는 경우

초기 고장률은 관측 중단 유형에 큰 영향을 받지 않으며, 후반부로 갈수록 영향이 커짐을 알 수 있다. 전반적으로 유형 ①, ②, ③은 추정결과에 큰 차이가 없다. 유형 ④의 경우는 후반부의 고장률이 유형 ③에 비해 두 배 가까이 증가했다. 이는 관측 중단이 매월 일정 비율로 이루어지는 경우, 중단 비율이 심각하게 크지 않은 경우에는 추정에 큰 영향을 미치지 않음을 의미한다.

유형 ⑤와 ⑦을 비교해 보면 큰 차이를 보이지 않으나, 유형 ⑥에 비해 유형 ⑧의 고장률 추정치가 유의하게 큰 것을 볼 수 있다. 이는 초기에 관측 중단이 집중적으로 발생할수록 고장률에 미치는 영향이 크다는 것을 의미한다.

이 제품은 11개월 사용 후 약 99.2~99.7%의 신뢰도를 가지는 것으로 추정되며, 고장률은 대체적으로 증가하는 경향을 보이며 11개월 제의 고장률은 대략 $1.1 \sim 3.2 (10^{-3}/\text{월})$ 로 추

정된다. 만약 단조 고장률 특성을 갖는 와이불(Weibull) 분포를 기초로 하여 추정한다면 적합도가 높지 않을 것으로 예상된다.

5. 결론 및 추후연구

본 연구에서는 기별 고장대수와 기별 판매대수 및 관측 중단수로 구성된 사용현장 보증 데이터에 대한 비모수적 분석 방법을 제시하였다. 기존의 NHPP 모형으로는 다루지 못하던 관측 중단 데이터를 포함하여 차별화된 추정 알고리즘을 구성하였으며, 특히 EM 알고리즘의 최대화 단계에서는 Kaplan-Meier 유형의 추정식을 유도하였다. 예제를 통하여 제안된 방법의 특성을 고찰하고, 사례를 통하여 활용 가능성을 보였다. 본 연구의 결과는 제품 고장에 대한 상세한 정보가 주어지지 않은 사용현장 데이터에 쉽게 적용될 수 있을 것으로 예상된다. 추후 연구로는 제품의 기별 사용비율을 추정하고, 이를 관측 중단수로 전환함에 따른 불확실성을 고려한 분석 방법 등이 기대된다.

부록 : [정리 1]의 증명

식 (4)로부터 식 (7)이 유도됨을 보인다.
식 (6)으로부터

$$R_0 = 1 - h_0, R_1 = (1 - h_0)(1 - h_1) \dots, \\ R_t = (1 - h_0)(1 - h_1) \dots (1 - h_t) \quad (A1)$$

$$\Rightarrow R_{t-1} - R_t = R_{t-1}h_t \\ = (1 - h_0)(1 - h_1) \dots (1 - h_{t-1})h_t \quad (A2)$$

식 (A1)과 (A2)를 식 (4)에 대입하면,

$$L(h_0, \dots, h_{T-1}) = \prod_{i=1}^S \prod_{t=0}^{T-i} [(1 - h_i)^{r_i} (1 - h_0) \dots (1 - h_{t-1})h_t]^{d_{i,t}} \\ \times \{(1 - h_0) \dots (1 - h_t)\}^{c_{i,t}} \quad (A3)$$

이를 다시 정리하면,

$$L(h_0, \dots, h_{T-1}) = \prod_{i=1}^S \prod_{t=0}^{T-i} [h_t^{d_{i,t}}] \times A \quad (A4)$$

$$\text{단, } A = [(1 - h_0) \dots (1 - h_{T-i})]^{r_i} (1 - h_0)^{d_{i,1}} \\ \dots [(1 - h_0) \dots (1 - h_{T-i-1})]^{d_{i,T-i}} \\ \times (1 - h_0)^{c_{i,0}} [(1 - h_0)(1 - h_1)]^{c_{i,1}} \\ \dots [(1 - h_0) \dots (1 - h_{T-i})]^{c_{i,T-i}} \\ = (1 - h_0)^{x_0} (1 - h_1)^{x_1} \dots (1 - h_{T-i})^{x_{T-i}} \quad (A5)$$

단, 식 (2)와 (3)으로부터

$$x_0 = r_i + (d_{i,1} + \dots + d_{i,T-i}) + (c_{i,0} + \dots + c_{i,T-i}) \\ = n_{i,0} - d_{i,0} \\ x_1 = r_i + (d_{i,2} + \dots + d_{i,T-i}) + (c_{i,1} + \dots + c_{i,T-i}) \\ = n_{i,1} - d_{i,1}, \\ \vdots \\ x_t = r_i + \sum_{k=t+1}^{T-i} d_{i,k} + \sum_{k=t}^{T-i} c_{i,k} \\ = N_i - \sum_{k=0}^{t-1} (d_{i,k} + c_{i,k}) + \sum_{k=t+1}^{T-i} d_{i,k} + \sum_{k=t}^{T-i} c_{i,k} \\ = N_i - \sum_{k=0}^{t-1} (d_{i,k} + c_{i,k}) - d_{i,t} = n_{i,t} - d_{i,t} \quad (A6) \\ \vdots \\ x_{T-i} = r_i + c_{i,T-i} = n_{i,T-i} - d_{i,T-i}$$

식 (A6)를 (A5)에, 이를 다시 (A4)에 대입하면 식 (7)을 얻는다.

참고문헌

Baxter, L. A. (1994), Estimation from Quasi Life Tables, *Biometrika*, **81**(3), 567-577.
Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of Royal Statistical Society, Series B*, **39**, 1-22.
Kalbfleisch, J. D., Lawless, J. F. and Robinson, J. A. (1991), Methods for the Analysis and Prediction of Warranty Claims, *Technometrics*, **33**(3), 273-285.
Kaplan, E. and Meier, P. (1958), Nonparametric Estimation from Incomplete Observation, *Jour. Amer. Statist. Assoc.*, **53**, 457-481.
Karim, M. R., Yamamoto, W. and Suzuki, K. (2001), Statistical Analysis of Marginal Count Failure Data, *Lifetime Data Analysis*, **7**(2), 173-186.
Tortorella, M. (1996), Life Estimation from Pooled Discrete Renewal Counts, in N. P. Jewell et al. (eds.) *Lifetime Data: Models in Reliability and Survival Analysis*, 331-338.