응용통계연구 제 8권 2호
1995년 9월 pp. 133-145.

수리가능한 제품의 사용현장 데이타 분석[1)]

배 도 선, 윤 형 제, 최 인 수[2)]

요 약

고장원인이 여럿인 수리불가능한 제품에 대하여 사용환경에서 얻어진 고장데이타와 추적조사에 의해 얻어진 설명변수에 관한 데이타를 이용하여 제품의 고장원인별 수명분포를 추정한 배도선 등(1995)의 연구를 수리가능한 제품의 경우로 확장하였다. 수명분포의 모수와 설명변수가 대수선형 관계일 때 비동질성 포아송과정을 이용하여 의사우도함수를 유도하고, 고장원인별 수명이 와이블 분포를 따를 때의 의사 최우추정량과 점근분산을 구하였다. 추적조사 방법으로는 보증기간 동안 고장이 발생하지 않은 제품의 일정비율을 추적조사하는 경우와 총 시험제품의 일정비율을 랜덤하게 선택하고 이들 중에서 보증기간 동안 한번도 고장이 발생하지 않은 제품에 대해서만 추적조사하는 경우를 고려하였다.

1. 서 론

생산되는 제품의 평균고장시간, 보증기간 동안 고장날 제품의 비율 및 이에 수반되는 품질비용, 시장에 출하한 제품 중 1% 또는 10% 등 일정비율이 고장나는 평균시점 등과 같은 제품의 수명에 관련된 품질특성치를 구하기위한 데이타들은 일반적으로 실험실에서 행해지는 수명시험이나 가속수명시험(accelerated life testing) 등으로 얻어지는 고장정보로 부터 수집된다. 그러나 실험실에서 이루어지는 수명시험에서는 제품의 실제 사용조건들을 모두 반영하는 것이 불가능하므로, 실험실에서 얻어지는 데이타들은 실제 사용환경에서의 제품수명에 대한 정보를 왜곡하여 나타날 수 있는 위험이 있다. 이러한 문제점의 해결을 위해서 제품을 사용현장에서 실제로 사용했을때의 고장데이타, 즉 사용현장 데이타를 얻고 이를 분석하는 방법들을 연구하는 것은 매우 중요한 일이다.

사용환경에서 얻어지는 고장데이타는 실험실에서 얻어지는 수명시험데이타와 다음과 같은 차이점으로 인하여 일반적인 수명시험데이타의 분석방법들을 이용할 수 없다.

(1) 사용현장 데이타는 아프터서비스를 통해서 얻어지고, 제품의 품질보증기간이 지난 후에는 이들 정보를 얻기 어렵다.
(2) 보증기간 이내에 고장이 나는 경우에도 아프터서비스를 받지 않는 경우에는 그 정보를 알 수 없다.
(3) 제품의 수명이 사용장소, 사용자의 특징 등과 같은 환경특성에 영향을 받는다.

특히 보증기간내에 고장이 한 번도 발생하지 않은 제품의 경우에는 제품의 수명에 대한 아무런 정보도 얻을 수 없게 되고, 보증기간 동안에 얻어지는 고장데이타만을 이용하여 제품의 수

1) 본 연구는 1993년 한국과학재단 연구과제 KOSEF 931-1000-004-2의 지원에 의하여 수행되었음.
2) (305-701) 대전시 유성구 구성동 373-1, 한국과학기술원 산업공학과.

명분포의 모수를 추정할 경우에는 일치추정량(consistent estimator)을 얻을 수 없게 된다 (Suzuki[1985a,b]). 사용현장 데이타의 분석에서는 이와같은 문제점을 해결하기 위해, 보증기간 동안의 고장데이타뿐만 아니라 고장이 발생하지 않은 제품에 대한 정보를 소비자로 부터 설문조사형식으로 얻어내어 제품의 수명을 추정하는데 사용한다.

사용환경에서 얻어진 고장데이타의 분석에 관한 기존연구로는 Hahn과 Meeker(1982), Lawless(1983), Suzuki(1985a,b), Kalbfleish와 Lawless(1988), 배도선 등(1995)이 있다. 이중에서 특히, Suzuki(1985a)는 사용환경에서 얻어진 고장데이타와 사용환경에 있는 총 시험제품 중 일정비율을 추적조사(follow-up)하여 얻은 제품의 운영시간을 이용하여 수명분포를 추정하는 문제를 다루었고, Suzuki(1985b)는 Suzuki(1985a)와 동일한 상황에서 비모수적 방법인 카플란 마이어(Kaplan-Meier)추정량을 이용하여 제품의 생존함수를 추정하는 문제를 다루었다. Kalbfleish와 Lawless(1988)는 제품의 수명에 영향을 주는 제품모델, 제조시간 등의 제조특성 및 사용자의 특징, 기후조건등의 환경특성 등과 같은 설명변수(covariate)가 존재할 때 사용환경에서 보증기간 동안 얻어지는 고장데이타와 사용환경에 있는 총 시험제품 중 고장이 발생하지 않은 제품의 일정비율을 추적조사하여 얻어진 설명변수에 대한 데이타를 이용하여 수명분포를 추정하였다. Suzuki(1985a,b)와 Kalbfleish와 Lawless(1988)는 제품에 고장을 일으키는 원인이 한가지라고 가정하였으나, 사용환경에서 사용되어지는 대부분의 제품은 여러가지 원인에 의하여 고장이 발생하게 된다. 고장원인이 여럿인 제품의 각 고장원인별 수명추정을 위해서는 다수고장원인(competing risk) 모형이 필요하며, 배도선 등(1995)은 Kalbfleish와 Lawless(1988)의 연구를 다수고장원인 모형으로 확장하였다.

사용환경 데이타의 분석에 관한 이와같은 기존연구들은 제품의 수리가 불가능하다는 가정하에 보증기간 동안 제품의 첫번째 고장만을 관측하는 경우에 대하여 다루었다. 그러나 실제 사용환경에서 쓰이는 제품들은 보증기간 동안 고장이 발생하면 수리하여 다시 사용되며, 수리된 제품은 남아있는 보증기간 동안 다시 고장이 발생할 수 있다. 예를 들어, 자동차가 정상적인 운행을 하지 못하는 상황을 고장이라고 하면, 고장은 엔진, 밧데리, 브레이크, 트랜스밋션 고장 등과 같은 여러가지 원인들 중 하나 이상의 결합에 의하여 발생된다. 만약 보증기간내에 자동차의 엔진에 고장이 발생했다면 구매자는 서비스센타에서 고장을 수리하여 사용할 것이고, 이 자동차는 수리이후에도 보증기간내에 엔진이나 다른 부위에 고장이 발생할 수 있으며 보증기간 동안 이와같은 과정이 반복될 것이다. 이런 경우 수리가 불가능하다고 가정한 모형에서는 첫번째 고장만이 분석을 위해 이용되고 그 이후의 고장정보는 무시되어 많은 정보의 손실을 가져오게 된다. 이와 같은 정보들을 제품의 수명분석에 적절히 이용하기 위해 수리가능한 제품에 대한 사용현장 데이타의 분석방법을 연구하는 것은 매우 중요한 일이다.

이 논문에서는 고장원인이 여럿인 수리불가능한 제품에 대한 배도선 등(1995)의 연구 결과를 확장하여, 고장원인이 여럿있고 수리가 가능한 제품에 대하여 사용환경에서 얻어진 고장데이타와 추적조사에 의해 얻어진 설명변수에 관한 데이타를 이용하여 제품의 고장원인별 수명분포를 추정한다. 수명분포의 모수와 설명변수가 대수선형관계일 때, 의사 우도함수(pseudo likelihood function)를 세워 의사 최우추정량(pseudo maximum likelihood estimator; PMLE)을 구하는 식을 유도하고, 특히 각 고장원인별 수명분포가 와이블분포일 때, 의사 최우추정량과 이들의 점근분산을 구한다. 또한 모의실험을 통하여 추적조사 비율에 따른 효과에 대하여 조사한다. 이 논문에서는 사용되는 기호는 다음과 같다.

N 총 시험제품의 수

k 제품의 고장원인 수

n 설명변수의 수

l_i 보증기간 동안 제품 i의 고장회수, $i=1,2,\cdots,N$

θ 제품 수명분포의 모수벡터

T_{ijm} 고장원인 m에 의한 제품 i의 j번째 고장시간

$i=1,2,\cdots,N;\ j=1,2,\cdots,l_i;\ m=1,2,\cdots,k$

T_{ij} 제품 i의 j번째 고장시간 ; $\min\{T_{ij1},\cdots,T_{ijk}\}$

x_i 제품 i에 대한 설명변수 열벡터 ; $(x_{i1}, x_{i2},\cdots,x_{in})'$

r_{ij} 제품 i의 j번째 고장의 고장원인

T^0 제품의 보증기간

D_1 보증기간 동안 한 번 이상 고장난 제품의 집합

D_1^c 보증기간 동안 한번도 고장이 발생하지 않은 제품의 집합

D_2 집합 D_1^c에서 샘플링된 제품의 집합

N_1, N_2 집합 D_1과 D_1^c에 속한 제품의 수 ; $N_1 + N_2 = N$

n_2 집합 D_1^c에서 샘플링된 제품의 수

p_2 추적조사비율 ; $p_2=\dfrac{n_2}{N_2}$

$f_m(t \mid x;\theta)$, $\overline{F}_m(t \mid x;\theta)$, $h_m(t \mid x;\theta)$ 설명변수 x 와 모수 θ가 주어졌을 때 고장원인 m에 의한 제품 수명의 확률밀도함수, 신뢰도함수 및 고장률 함수

2. 수리가능 제품의 사용현장 데이타의 분석

사용현장 데이타의 형태는 보증기간 동안 고장이 발생한 제품에 대하여 고장시간과 설명변수의 값들이 여러번 관측될 수 있다는 점을 제외하면 배도선등(1995)과 동일하다. 또한 이에 따른 가정은 배도선 등(1995)의 가정이외에 다음과 같은 추가가정이 필요하다.

<추가가정>

④ 고장난 제품은 수리된 후 고장이 발생하기 바로 전 상태로 된다.

⑤ 고장난 제품을 수리하는 시간은 보증기간에 비하여 무시할 수 있을 정도로 짧다.

추적조사를 위한 샘플링방법으로 총 N개의 시험제품에서 보증기간 동안 고장나지 않은 제품 N_2개 중 p_2비율의 제품을 단순 랜덤샘플링하는 Kalbfleish 와 Lawless(1988)와 배도선 등

(1995)의 접근방법을 사용하여, 제품의 고장원인별 수명분포의 모수를 추정하기 위한 의사(pseudo) 우도함수를 세우고, 이 함수를 이용하여 의사 최우추정량을 구하며, 고장원인별 수명분포가 와이블분포인 경우의 의사 최우추정량과 점근분산을 구한다. 또한 총 N개의 시험제품에서 p^*의 비율로 제품을 단순 랜덤샘플링하고 이들 중에서 보증기간 동안 고장이 한 번도 발생하지 않는 제품에 대해서만 추적조사를 하는 Suzuki(1985a,b)의 접근방법에 대해서는 제2.3절에서 다룬다.

2.1 의사 최우추정량 및 추정량의 점근성질

수리 가능한 제품의 시간 t까지의 고장회수 $N(t)$는 다음과 같은 성질을 갖는 비동질성 포아송 과정(Non-homogeneous Poisson Process : NHPP)으로 표현될 수 있다. (Asher와 Feingold, 1984)

① $N(0) = 0$

② $\{ N(t), \ t \geq 0 \}$는 독립증분(independent increment)의 성질을 갖는다.

③ $P\{ N(t+\Delta) - N(t) \geq 2 \} = o(\Delta)$

④ $P\{ N(t+\Delta) - N(t) = 1 \} = h(t) \cdot \Delta + o(\Delta)$

단, $o(\Delta)$는 $\lim_{\Delta \to 0} \frac{z(\Delta)}{\Delta} = 0$인 함수 $z(\Delta)$를 나타낸다. Crow(1974)는 $\{ N_i(t), \ t \geq 0 \}$가 NHPP이고, 제품 i의 고장시간들이 $T_{i1} < T_{i2} < \cdots < T_{il_i} < T^0$일때, $Y_{ij} = T_{ij} - T_{i,j-1}$라고 하면 Y_{ij}의 누적분포함수가 다음과 같음을 보였다.

$$F_{ij}(y) = \frac{F(t_{i,j-1}+y) - F(t_{i,j-1})}{1 - F(t_{i,j-1})}, \quad i = 1, 2, \cdots, N, \ j = 1, 2, \cdots, l_i \qquad (1)$$

식 (1)을 이용하여 고장원인이 하나인 경우에 보증기간 T^0까지 한 번 이상 고장난 제품의 우도함수에 대한 기여는

$$\begin{aligned} & f(t_{i1}) \cdot \frac{f(t_{i2})}{1 - F(t_{i1})} \cdots \frac{f(t_{i,l_i})}{1 - F(t_{i,l_i} - 1)} \cdot \frac{1 - F(T^0)}{1 - F(t_{i,l_i})} \\ & = f(t_{i1}) \cdot f(t_{i2} | T_{i2} \geq t_{i1}) \cdots f(t_{i,l_i} | T_{i,l_i} \geq t_{i,l_i-1}) \cdot F(T^0 | T_{i,l_i+1} \geq T^0) \end{aligned} \qquad (2)$$

가 되고 보증기간 T^0까지 한 번도 고장나지 않은 제품의 경우에는

$$1 - F(T^0) = \overline{F}(T^0) \qquad (3)$$

가 된다. 따라서 N개의 제품 모두에 대하여 설명변수 x_i를 안다면, 식 (2)와 (3)을 이용하여 우도함수 $L_F(\theta)$는 고장원인이 하나인 경우

$$L_F(\theta) = \prod_{i;t_{i1}\le T^0} \prod_{j=1}^{l_i} [\ f(t_{ij} \mid T_{ij} > t_{i,j-1},\ x_i;\theta)] \cdot \prod_{i;t_{i1}\le T^0} [\ \overline{F}(T^0 \mid T_{i,l_i+1} > t_{i,l_i},\ x_i;\theta)] \cdot \prod_{i;t_{i1}> T^0} \overline{F}(T^0 \mid x_i;\theta) \quad (4)$$

이 된다. 여기서

$$f(t_{ij} \mid T_{ij} > t_{i,j-1},\ x_i;\theta) = \frac{f(t_{ij} \mid x_i;\theta)}{\overline{F}(t_{i,j-1} \mid x_i;\theta)},$$

$$\overline{F}(T^0 \mid T_{ij} > t_{i,j-1},\ x_i;\theta) = \frac{\overline{F}(T^0 \mid x_i;\theta)}{\overline{F}(t_{i,j-1} \mid x_i;\theta)}, \quad t_{i0} = 0, \quad i=1,\cdots,N,$$

이다. 그러므로, 고장원인이 $k(>1)$개인 경우 $I[\ r_{ij}=m]$ 을 제품 i의 j번째 고장이 고장원인 m에 의하여 발생하였음을 나타내는 지시함수(indicator function)라 하고, 고장원인이 하나인 경우와 비슷하게 구하면 우도함수는

$$L_F(\theta) = \prod_{i;t_{i1}\le T^0} \prod_{j=1}^{l_i} \prod_{m=1}^{k} \left[\ f_m(t_{ij} \mid T_{ij} > t_{i,j-1},\ x_i;\theta) \prod_{l\ne m}^{k} \overline{F}_l(t_{ij} \mid T_{ij} > t_{i,j-1},\ x_i;\theta)\right]^{I[\ r_{ij}=m]} \cdot \prod_{i;t_{i1}\le T^0} \prod_{m=1}^{k} [\ \overline{F}_m(T^0 \mid T_{i,l_i+1} > t_{i,l_i},\ x_i;\theta)] \cdot \prod_{i;t_{i1}> T^0} \prod_{m=1}^{k} \overline{F}_m(T^0 \mid x_i;\theta) \quad (5)$$

가 된다.

그러나, 모든 제품에 대하여 설명변수 x_i를 안다는 것은 현실적으로 불가능하므로, 이 논문에서는 고장이 발생했거나 추적조사되는 제품에 대해서만 설명변수 x_i를 알 수 있다고 가정한다. D_1을 보증기간 동안 고장난 제품의 집합이라 하고, D_2를 D_1의 여집합 D_1^c에서 추적조사 비율 p_2로 샘플링된 제품의 집합이라고 정의하자. 보증기간 동안 고장이 한 번이라도 발생한 제품 i에 대한 데이타는 보증기간 T^0 동안 고장이 발생한 회수 $l_i(>0)$, 고장시간 t_{ij} 및 고장원인 r_{ij}, $j=1,\cdots,l_i$, 그리고 설명변수 x_i가 얻어지고, 보증기간 T^0 동안 고장이 발생하지 않은 제품들 중에서 추적조사 비율 p_2로 샘플링된 경우에는, 설명변수 x_i가 얻어진다. 따라서 고장원인이 여럿이고 수리가능한 제품의 경우 모수 θ에 대한 의사 우도함수 $L_p(\theta)$는

$$L_P(\theta) = \prod_{i\in D_1} \prod_{j=1}^{l_i} \prod_{m=1}^{k} \left[\ f_m(t_{ij} \mid T_{ij} > t_{i,j-1},\ x_i;\theta) \prod_{l\ne m}^{k} \overline{F}_l(t_{ij} \mid T_{ij} > t_{i,j-1},\ x_i;\theta)\right]^{I[\ r_{ij}=m]} \cdot \prod_{i\in D_1} \prod_{m=1}^{k} [\ \overline{F}_m(T^0 \mid T_{i,l_i+1} > t_{i,l_i},\ x_i;\theta)] \cdot \prod_{i\in D_2} \left[\ \prod_{m=1}^{k} \overline{F}_m(T^0 \mid x_i;\theta)\right]^{\frac{1}{p_2}} \quad (6)$$

이 되고, 식 (6)의 양변에 로그를 취하면 다음과 같은 의사 대수우도함수를 얻는다.

$$\log L_P(\theta) = \sum_{i\in D_1}\sum_{j=1}^{l_i}\sum_{m=1}^{k} I[\ r_{ij}=m]\ \log\left\{f_m(t_{ij} \mid T_{ij}> t_{i,j-1},\ x_i;\theta)\prod_{l\neq m}^{k}\overline{F}_l(t_{ij} \mid T_{ij}> t_{i,j-1},\ x_i;\theta)\right\}$$
$$+ \sum_{i\in D_1}\sum_{m=1}^{k}\log\ \overline{F}_m(T^0 \mid T_{i,l_i+1}> t_{i,l_i},\ x_i;\theta) \tag{7}$$
$$+ \frac{1}{p_2}\sum_{i\in D_2}\sum_{m=1}^{k}\log\ \overline{F}_m(T^0 \mid x_i;\theta)$$

식 (7)의 의사 대수우도함수는 보증기간에 고장이 발생하지 않는 제품 중 추적조사된 제품으로 부터 얻을 수 있는 정보에 $1/p_2$의 가중을 주는 것으로, 이를 최대로하는 θ^*를 구하면, 이때의 θ^*가 의사 최우추정량이 되며, θ^*는 N이 커짐에 따라 배도선 등(1995)의 정리1과 동일한 성질을 갖음을 보일 수 있다.

2.2 고장 원인별 수명분포가 와이블인 경우

각 고장원인별 제품 수명이 와이블분포를 따르고 척도모수(scale parameter)가 설명변수와 대수선형관계를 갖는 경우, 고장원인 m에 의한 제품수명분포의 모수가 $\beta_m=(\beta_{m0},\beta_{m1},\cdot\cdot\cdot,\beta_{mn})'$과 δ_m인 신뢰도함수는 다음과 같다.

$$\overline{F}_m(t \mid x_i) = \exp(-t^{\delta_m}\cdot \exp(x_i'\beta_m)),\qquad t \geq 0, \tag{8}$$

단, 여기서 $m=1,2,\cdot\cdot\cdot,k$ 이다. 이 경우 모수벡터 θ는

$$\theta = (\beta_{10},\beta_{11},\ldots,\beta_{1n},\delta_1,\beta_{20},\beta_{21},\ldots,\beta_{2n},\delta_2,\ldots,\beta_{k0},\beta_{k1},\ldots,\beta_{kn},\delta_k)$$

이고, $x_i'\beta_m = \beta_{m0}+\beta_{m1}x_{i1}+\cdots+\beta_{mn}x_{in}$이 된다. 식(9)를 이용하여 $j-1$번째 고장시간이 주어져 있을때 j번째 고장시간에 대한 확률밀도함수와 신뢰도함수를 구하면

$$f_m(t_{ij}|T_{ij}> t_{i,j-1},x_i;\theta) = \delta_m\cdot t_{ij}^{\delta_m-1}\exp(x_i'\beta_m)\cdot\exp[\ -(t_{ij}^{\delta_m}-t_{i,j-1}^{\delta_m})\exp(x_i'\beta_m)]\ , \tag{9}$$

$$\overline{F}_m(t_{ij}|T_{ij}> t_{i,j-1},x_i;\theta) = \exp[\ -(t_{ij}^{\delta_m}-t_{i,j-1}^{\delta_m})\exp(x_i'\beta_m)] \tag{10}$$

이 된다. 식(9)와 (10)을 식(7)의 의사 대수우도함수에 대입하면

$$\log L_P(\theta) = \sum_{m=1}^{k}\left\{\sum_{i\in D_1}\sum_{j=1}^{l_i} I[\ r_{ij}=m]\ \cdot[\ \log\delta_m+(\delta_m-1)\log t_{ij}+x_i'\beta_m]\right. \tag{11}$$
$$\left. - \sum_{i\in D_1}((T^0)^{\delta_m}\exp(x_i'\beta_m)) - \frac{1}{p_2}\sum_{i\in D_2}((T^0)^{\delta_m}\exp(x_i'\beta_m))\right\}$$

이 된다. 식(11)을 β_{mr}과 δ_m에 대하여 일차 편미분하여 0으로 놓고 정리하면 의사 최우추정치 $\beta^*_{mr},\delta^*_m,\ m=1,2,\cdot\cdot\cdot,k,\ r=1,2,\cdot\cdot\cdot,n$은

$$\sum_{i\in D_1}\sum_{j=1}^{l_i} I[\ r_{ij}=m]\ \cdot x_{ir} = \sum_{i\in D_1}A_{imr}+\frac{1}{p_2}\sum_{i\in D_2}A_{imr} \tag{12}$$

$$\sum_{i\in D_1}\sum_{j=1}^{l_i} I[\ r_{ij}=m]\ (\frac{1}{\delta_m}+\log t_{ij}) = \sum_{i\in D_1}B_{im}+\frac{1}{p_2}\sum_{i\in D_2}B_{im} \tag{13}$$

$$A_{imr} = (T^0)^{\delta_m} \exp(x_i'\beta_m) x_{ir}$$

$$B_{im} = (T^0)^{\delta_m} \exp(x_i'\beta_m)(\log T^0)$$

$$r = 1, 2, \cdots, n, \quad m = 1, 2, \cdots, k$$

을 만족하는 값이되며, 이는 수치적 방법에 의하여 구할 수 있다.

의사 최우추정량 β^*_{mr}, δ^*_m, $m=1,2,\cdots,k$, $r=1,2,\cdots,n$ 의 점근분산은 식 (9)-(13)과 배도선 등(1995)의 식 (9)-(10)을 이용하여 구할 수 있으며, 부록에서 자세히 다룬다. 부록의 식 (A.1) - (A.5)를 이용하면 분산-공분산 행렬의 일치추정량을 구할 수 있고 이로부터 고장원인별 수명분포의 모수에 대한 의사 최우추정량의 점근분산을 구할 수 있다.

2.3 Suzuki(1985a,b)의 추적조사 방법의 경우

이 절에서는 총 N개의 시험제품에서 p^*의 비율로 제품을 단순 랜덤샘플링하고 이들 중에서 보증기간 동안 고장이 한번도 나지 않은 제품에 대해서만 추적조사하는 Suzuki(1985a,b)의 추적조사방법에 대하여 의사 우도함수를 유도하고 이의 분석방법도 제시한다.

보증기간내에 한 번 이상 고장이 발생한 제품의 수를 n_u, 보증기간내에 한 번도 고장이 발생하지 않았으나 추적조사되는 제품과 추적조사되지 않는 제품의 수를 각각 n_c와 n_l 이라고 하자. 이 때 n_u, n_c와 n_l은 모두 확률변수가 된다. n_u개의 제품에 대해서는 그 제품의 설명변수의 값, 고장시간 그리고 고장원인에 대한 데이타를 얻을 수 있으며, Np^*개의 제품 중 보증기간 동안 고장이 발생하지 않아 추적조사되는 n_c개의 제품에 대해서는 설명변수의 값만을 얻을 수 있다. 이를 이용하여 의사 우도함수를 구하면

$$\begin{aligned}
\log L_P(\theta) \;=\; & \sum_{i=1}^{n_u} \sum_{j=1}^{l_i} \sum_{m=1}^{k} I[\; r_{ij}=m] \;\log\left\{ f_m(t_{ij} \mid T_{ij} > t_{i,j-1},\; x_i;\theta) \prod_{l\neq m}^{k} \overline{F}_l(t_{ij} \mid T_{ij} > t_{i,j-1},\; x_i;\theta) \right\} \\
& + \sum_{i=1}^{n_u} \sum_{m=1}^{k} [\; \log \;\overline{F}_m(T^0 \mid T_{i,l_i+1} > t_{i,l_i},\; x_i;\theta)] \qquad (14) \\
& + (1+n_l/n_c) \sum_{i=1}^{n_c} \sum_{m=1}^{k} \log \;\overline{F}_m(T^0 \mid x_i;\theta)
\end{aligned}$$

이 된다. 제2.1절에서와 마찬가지로 식 (14)를 최대로하는 θ^*를 구하면, 이때의 θ^*가 의사 최우추정량이 되며 θ^*는 N이 커짐에 따라 배도선 등(1995)의 정리 1과 같은 성질을 갖는다.

3. 예제 및 추적조사 비율에 따른 의사최우추정량의 성질

3.1. 예제

수리가 가능한 제품에 대한 사용현장 데이타의 분석방법의 설명을 위해서, 어떤 중장비의 고장원인은 2가지가 있고, 제품이 사용되는 환경을 두가지로 분류한 0, 1 값을 갖는 설명변수 x_{i1}이 있다고 하자. x_{i1}값의 반은 0 그리고 나머지 반은 1이며, x_{i1}이 주어졌을 때 고장시간 t_i는 고장원인별 신뢰도함수가

$$\overline{F}_m(t \mid x_{i1}) = \exp(-t^{\delta_m} \cdot \exp(\beta_{m0}+\beta_{m1}x_{i1})), \quad t \geq 0, \quad m=1,2, \qquad (15)$$

인 와이블분포를 따른다고 하자. Kalbfleish와 Lawless(1988)와 배도선 등(1995)의 예제와 유사하도록 총 시험제품수와 보증기간을 각각 N=5,370과 T^0= 38(개월)로 하고, 고장원인별 수명분포를 각각 모수 $(\beta_{10}, \beta_{11}, \delta_1)$=(-11.98, 0.90, 2.0)과 $(\beta_{20}, \beta_{21}, \delta_2)$=(-15.10, 1.30, 3.0)인 와이블분포로 하여, 난수를 발생한 결과 264개의 제품에 한 번 이상 고장이 발생하였고 총 고장회수는 279번이었다. 279번의 고장데이타중 $(x=0,\ m=1), (x=1,\ m=1), (x=0,\ m=2), (x=1,\ m=2)$일 때의 고장갯수는 각각 26, 60, 47, 146이었고, 또한 38개월 동안 고장이 발생하지 않은 N_2 = 5,106개의 제품 중 $p_2=0.1$의 비율로 추적조사하여, 511개의 제품중 263개의 $x_{i1}=0$과 248개의 $x_{i1}=1$ 값을 얻었다. 이와 같은 데이타와 식 (14)와 (15)를 이용하여 의사 최우추정치를 수치적인 방법으로 구하면

$$\beta^*_{10}=-11.65,\ \beta^*_{11}=0.88,\ \delta^*_1=1.92$$

$$\beta^*_{20}=-14.22,\ \beta^*_{21}=1.18,\ \delta^*_2=2.79$$

이며, 각 추정치들의 모수에 대한 오차를 구해보면 다음과 같다.

$$error(\beta^*_{10})=-0.33,\ error(\beta^*_{11})=0.02,\ error(\delta^*_1)=0.08$$

$$error(\beta^*_{20})=0.88,\ error(\beta^*_{21})=-0.12,\ error(\delta^*_2)=0.21$$

부록의 식 (A1)-(A6)를 이용하여 분산-공분산 행렬의 추정치를 구하고, 이를 이용하여 각 의사 최우추정량들의 점근분산을 구하면 다음과 같다.

$$asvar(\beta^*_{10})=0.6080,\ asvar(\beta^*_{11})=0.0615,\ asvar(\delta^*_1)=0.0429$$

$$asvar(\beta^*_{20})=0.5564,\ asvar(\beta^*_{21})=0.0345,\ asvar(\delta^*_2)=0.0403$$

제품이 사용되고 있는 환경, 즉 설명변수 x_{i1}에 따라 각 고장원인별로 10%의 제품이 고장나는 시점인 10백분위수의 추정치(단위:개월)를 구해보면 다음과 같다.

x_{i1}	고장원인 1에의한 10백분위수	고장원인 2에의한 10백분위수
0	133.22	72.95
1	84.14	47.76

이로부터 설명변수 x_{i1}이 1인 경우 더 많은 고장이 발생하고, 고장원인 2에 의한 고장이 고장원인 1에 의한 고장보다 자주 발생함을 알 수 있다. 만약 실제 상황에서 이런 결과가 얻어졌다면 설명변수 1이 의미하는 사용환경에서 제품이 사용될 경우 제품의 신뢰성을 높이기 위한 방법과 고장원인 2에 의한 고장을 줄이기 위한 방법이 강구되어야 할 것이다.

배도선 등(1995)의 수리불가능 제품에 대한 다수 고장원인모형을 이용하여 구한 제품의 10백분위수와 본 연구에서 제시된 모형에 의한 10백분위수를 비교해보면 다음과 같다.

x_{i1}	제품의 10백분위수의 참값	수리가능한 제품을 위한 다수 고장원인모형에 의한 10백분위수의 추정치	수리불가능한 제품을 위한 다수 고장원인모형에 의한 10백분위수의 추정치
0	65.66	65.60	66.51
1	42.43	42.64	42.84

이로부터 수리가 가능한 제품의 경우에는 본 연구에서 제시된 모형으로 사용현장 데이타를 분석하는 것이 더 정확한 추정치를 얻을 수 있다는 것을 알 수 있다. 이는 수리가능한 제품의 수명에 대한 분석을 수리불가능한 제품에 대한 모형을 이용하여 분석할 경우 많은 정보의 손실을 가져올 수 있기 때문이다.

3.2 추적조사 비율에 따른 의사최우추정량의 성질

추적조사 비율 p_2에 따른 의사 최우추정량 θ^*의 성질을 알아보고 수리불가능한 모형과 비교하기 위하여 예제와 동일한 상황과 모수를 이용하여 $p_2=0.01$, 0.05, 0.1, 0.2, 0.5, 1.0일 때의 10백분위수의 추정량의 효율(efficiency)을 구하는 모의실험을 배도선 등(1995)에서와 동일한 절차로 수행하였다. 단, 총 시험제품수 N=1,000으로 하였다.

<표 1> p_2값에 따른 10백분위수의 추정량의 효율

p_2	고장원인 1(m = 1)		고장원인 2(m = 2)	
	설명변수 $x_{i1}=0$	설명변수 $x_{i1}=1$	설명변수 $x_{i1}=0$	설명변수 $x_{i1}=1$
0.01	0.833	0.796	0.856	0.787
0.05	0.955	0.864	0.950	0.828
0.1	0.978	0.915	0.978	0.964
0.2	0.989	0.929	0.994	0.973
0.5	0.996	0.965	0.999	0.997
1.0	1.000	1.000	1.000	1.000

<표 1>은 이와 같은 모의실험의 결과로서 고장원인별 수명분포가 와이블인 경우 추적조사 비율 p_2에 따른 10백분위수에 대한 추정량의 효율을 나타낸 것이다. 표 1에 의하면 추적조사 비율 p_2가 0.05 이상이 되면 평균제곱오차의 효율이 0.80이상이 되며, 이는 추적조사 비율이 그리 크지 않더라도 의사 최우추정량을 이용하여 구한 10백분위수의 효율이 크게 떨어지지 않음을 알 수 있다.

4. 결 론

이 논문에서는 고장원인이 여럿인 제품의 사용현장 데이타 분석에 관한 배도선 등(1995)을 수리가능한 제품의 경우로 확장하여, 사용환경에서 얻어진 고장데이타와 추적조사된 정보를 이용하여 고장원인별 수명분포를 추정하는 문제를 다루었다. 제품수명분포의 모수가 설명변수와 대수선형관계일 때, NHPP를 이용하여 의사 우도함수를 유도 하였으며, 고장원인별 제품의 수명이 와이블분포를 따를 때의 의사 최우추정량과 점근분산을 구하였다. 이러한 결과들은 추적조사를 위한 샘플링방법이 다른 경우에 대해서도 쉽게 확장될 수 있음을 알 수 있다. 또한 모의실험을 통하여 추적조사 비율에 따른 의사 최우추정량의 점근효율을 구한 결과 추적조사 비율이 그리 크지 않더라도 의사 최우추정량을 이용하여 구한 10백분위수의 추정량의 효율이 크게 떨어지지 않음을 알 수 있었다.

부록 : 분산-공분산 행렬

제품의 수명분포가 와이블 분포인 경우 의사 최우추정량 β^*_{mr}, δ^*_m, $m=1,2,\cdots,k$, $r=1,2,\cdots,n$ 의 점근분산행렬 $A_N(\beta,\delta)$와 $C_N(\beta,\delta)$의 원소들을 식 (9)-(13)과 배도선 등 (1995)의 식 (9)-(10)을 이용하여 구하면

$$A_N(\beta,\delta)_{mr,ms} = -\frac{1}{N}\frac{\partial^2 \log L_p}{\partial\beta_{mr}\partial\beta_{ms}}$$

$$= \frac{1}{N}\sum_{i\in D_1} x_{ir}x_{is}W_{im} + \frac{1}{Np_2}\sum_{i\in D_2} x_{ir}x_{is}W_{im} \quad \text{(A1)}$$

$$A_N(\beta,\delta)_{mr,m(n+1)} = -\frac{1}{N}\frac{\partial^2 \log L_p}{\partial\beta_{mr}\partial\delta_m}$$

$$= \frac{1}{N}\sum_{i\in D_1}(\log T^0)(T^0)^{\delta_m}\exp(x_i'\beta_m)x_{ir}$$

$$+\frac{1}{Np_2}\sum_{i\in D_2}(\log T^0)(T^0)^{\delta_m}\exp(x_i'\beta_m)x_{ir}$$

$$= \frac{1}{N}\sum_{i\in D_1} x_{ir}(\log T^0)W_{im} + \frac{1}{Np_2}\sum_{i\in D_2} x_{ir}(\log T^0)W_{im} \quad \text{(A2)}$$

$$A_N(\beta,\delta)_{m(n+1),m(n+1)} = -\frac{1}{N}\frac{\partial^2 \log L_p}{\partial\delta_m^2}$$

$$= \frac{1}{N}\sum_{i\in D_1}\sum_{j=1}^{l_i} I[\ r_{ij}=m]\ (\delta_m)^{-2}$$

$$+\frac{1}{N}\sum_{i\in D_1}(\log T^0)^2 W_{im} + \frac{1}{Np_2}\sum_{i\in D_2}(\log T^0)^2 W_{im} \quad \text{(A3)}$$

$$C_N(\beta,\delta)_{mr,ms} = \frac{N_2(1-p_2)}{Np_2(n_2-1)}\sum_{i\in D_2}(x_{ir}H_{im}-\overline{H}_{mr})(x_{is}H_{im}-\overline{H}_{ms}) \quad \text{(A4)}$$

$$C_N(\beta,\delta)_{mr,m(n+1)} = \frac{N_2(1-p_2)}{Np_2(n_2-1)}\sum_{i\in D_2}(\log T^0)(x_{ir}H_{im}-\overline{H}_{mr})(H_{im}-\overline{H}_m) \quad \text{(A5)}$$

$$C_N(\beta,\delta)_{m(n+1),m(n+1)} = \frac{N_2(1-p_2)}{Np_2(n_2-1)}\sum_{i\in D_2}(\log T^0)^2(H_{im}-\overline{H}_m)^2 \quad \text{(A6)}$$

이고, 여기서 $W_{im}=(T^0)^{\delta_m}\exp(x_i'\beta_m)$, $H_{im}=-(T^0)^{\delta_m}\exp(x_i'\beta_m)$, $\overline{H}_m=\sum_{i\in D_2}\frac{H_{im}}{n_2}$, $\overline{H}_{mr}=\sum_{i\in D_2}\frac{x_{ir}H_{im}}{n_2}$, $r,s=1,2,\cdots,n$, $m=1,2,\cdots,k$ 이다.

참고 문헌

[1] 배도선, 최인수, 황용근 (1995). 고장원인이 여럿인 제품의 사용현장 데이타 분석, 응용통계연구, 제 8권 1호, 89-104.

[2] Asher, H. and Feingold, H. (1984). *Repairable Systems Reliability : Modeling, Inference, Misconceptions and Their Causes*, Marcel Dekker, Inc.

[3] Cramer, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J.

[4] Crow, L. H. (1974). Reliability Analysis for Complex Repairable Systems, in *Reliability and Biometry*, Proschan,F and Serfling,R.J, eds., SIAM, Philadelphia, 379-410.

[5] Crowder, M. (1986). On consistency and inconsistency of estimating equations, *Econometric Theory*, Vol. 2, 305-330.

[6] Hahn, G. J. and Meeker, W. Q. (1982). Pitfalls and practical considerations in product life analysis(part 1 and 2), *Journal of Quality Technology*, Vol. 14, 144-152, 177-185.

[7] Inagaki, N. (1973). Asymptotic relations between the likelihood estimating function and maximum likelihood estimator, *Annals of the Institute of Statistical Mathematics*, Vol. 25, 1-26.

[8] Kalbfleisch, J. D. and Lawless, J. F. (1988). Estimation of reliability in field-performance studies, *Technometrics*, Vol. 30, 365-388.

[9] Lawless, J. F. (1983). Statistical methods in reliability, *Technometrics*, Vol. 25, 305-335.

[10] Suzuki, K. (1985a). Estimation of lifetime parameters from incomplete field data, *Technometrics*, Vol. 27, 263-272.

[11] Suzuki, K. (1985b). Nonparametric estimation of lifetime distributions from a record of failures and follow-ups, *Journal of the American Statistical Association*, Vol. 80, 68-72.

Field Data Analyses for Repairable Products[3)]

Do Sun Bai, Hyung Je Yun, In Su Choi[4)]

Abstract

This paper is concerned with the method of estimating lifetime distribution from field data for repairable products with multiple modes of failure, and is an extension of Bai et al.(1995). The log linear function is considered as a model for describing the relation between failure time of a product and covariates. Using the nonhomogeneous poisson process, general methods for obtaining pseudo maximum likelihood estimators(PMLEs) for the parameters are outlined and specific formulas for Weibull distribution are obtained. Effects of follow-up percentage on the PMLEs are investigated. Extension to case-cohort design is also considered.

---

3) This research was supported by the Korean Science & Engineering Foundation Grant KOSEF 931-1000-004-2.

4) Department of Industrial Engineering, Korea Advanced Institute of Science and Technology, Taejon, 305-701, KOREA.