응용통계연구 제 8권 1호
1995년 3월 pp. 89-104.

고장원인이 여럿인 제품의 사용현장 데이타 분석[1]

배 도 선[2], 최 인 수[2], 황 용 근[3]

요 약

사용현장에서의 고장데이타는 미리정해진 보증기간동안 고장이 발생한 제품으로 부터 얻어지는 고장시간, 고장원인, 설명변수값과 보증기간동안 고장나지 않는 제품 중 일정비율을 추적조사하여 얻은 설명변수 값들로 구성된다. 사용현장에서 얻어지는 이와같은 데이타를 이용하여 제품수명분포의 모수가 설명변수와 대수선형관계일 때, 수명분포의 모수에 대한 의사(pseudo) 최우추정량을 구하고 그 점근성질을 규명하였으며, 고장원인별 제품수명이 와이블분포를 따를 때의 의사최우추정량과 점근분산을 구하였다. 제품의 보증기간이 달력시간이고 제품의 고장이 운영시간에 의존하는 경우와 제품의 보증이 달력시간과 운영시간의 혼합인 경우의 분석방법도 제시하였다. 또한 모의실험을 통하여 추적조사비율에 따른 효과를 알아보았다.

1. 서 론

고장데이타는 수명분포를 추정하고, 신뢰도, 고장률, 백분위수, 평균 고장시간 등과 같은 제품의 수명과 관련된 여러 정보를 얻는데 사용된다. 가장 이상적인 고장데이타는 사용환경에서 얻어져야 한다. 그러나 고장데이타들은 시간과 비용, 고장데이타 취득의 어려움, 그리고 사용환경에서 얻어지는 데이타의 중요성에 대한 인식 부족 등으로 인하여 사용환경에서 얻기보다는 실험실에서 행해지는 수명시험을 통해 얻어지는 것이 보통이다. 실험실에서 수행하는 수명시험을 통하여 얻어지는 데이타들은 실제 사용환경에서의 제품수명에 대한 정보를 왜곡하여 나타낼 수 있는 위험이 있다. 따라서 이러한 위험을 줄이고 제품의 수명에 대한 올바른 정보를 얻어내기 위하여, 사용환경에서 고장데이타를 얻고 이를 분석하는 방법들을 연구하는 것은 매우 중요한 일이다.

사용환경에서 얻어지는 고장데이타는 보통 제품에 고장이 발생했을 때 보증제도를 통해 얻어지는 고장시간에 관한 데이타이다. 즉, 보증기간 내에 제품이 고장나면 소비자는 아프터서비스를 받기 위해 서비스센터나 대리점을 찾게되고, 이 때 제품의 고장시간과 고장원인, 그리고 제품모델, 제조시간 등의 제조특성(manufacturing characteristic) 및 사용자의 성격, 기후조건 등과 같은 사용특성(environmental characteristic)에 관한 데이타들을 얻게 된다. 그러나, 보증기간 내에 고장이 발생하지 않은 제품의 경우에는 제품의 수명에 대한 이러한 정보들을 얻을 수 없다. 따라서 보증기간 동안에 얻어지는 고장데이타만을 이용하여 제품의 수명을 추정한다면 추정치는 왜곡된 결과를 보일 것이다. Suzuki(1985a)에 의하면 이러한 보증기간 동안의 고장데

1) 본 연구는 1993년 한국과학재단 연구과제 KOSEF 931-1000-004-2의 지원에 의하여 수행되었음.
2) (305-701) 대전시 유성구 구성동 373-1, 한국과학기술원 산업공학과.
3) (150-042) 서울시 영등포구 당산동 2가 11-5, 주식회사 큐빅테크.

이타만을 이용하여 수명분포의 모수를 추정할 경우 일치추정량(consistent estimator)을 얻을 수 없게 된다. 사용환경에서 얻어진 고장데이타에 관한 이와 같은 문제점을 해결하기 위해, 보증기간 동안의 고장데이타 뿐만 아니라 고장이 발생하지 않은 제품에 대한 정보를 소비자로 부터 설문조사 형식으로 얻어내어 제품의 수명을 추정하는데 사용한다면 추정치의 정확도가 훨씬 증가하게 될 것이다.

사용환경에서 얻어진 고장데이타의 분석에 관한 기존연구로는 Hahn 과 Meeker(1982), Lawless(1983), Suzuki(1985a,b), Kalbfleish 와 Lawless(1988)등이 있다. 이중에서 특히, Suzuki(1985a,b)는 보증기간이 달력시간(calendar time)인 경우 사용환경에서 얻어지는 제품의 고장시간과 총 시험제품 중 일정비율을 추적조사(follow-up)하여 얻어진 보증기간동안 제품의 운영시간(operating time)을 이용하여 제품의 수명을 추정하는 문제를 다루었고, Kalbfleish와 Lawless(1988)는 제품의 수명에 영향을 주는 제조특성, 환경특성 등과 같은 설명변수(covariate)가 존재할 때 사용환경에서 보증기간동안 얻어지는 고장데이타와 사용환경에 있는 총 시험제품 중 고장이 발생하지 않은 제품의 일정비율을 추적조사하여 얻어진 설명변수에 대한 데이타를 이용하여 수명분포를 추정하였다.

사용환경에서 얻어진 고장데이타의 분석에 관한 이러한 기존연구들은 시험제품에 고장을 일으키는 원인이 한가지라고 가정하였으나, 실제로 제품의 고장은 여러가지 원인으로 인하여 발생하게 된다. 예를 들어 자동차가 정상적인 운행을 하지 못하는 상황을 고장이라고 하면, 고장은 엔진, 밧데리, 브레이크, 트랜스밋션 등과 같은 여러가지 원인들 중 하나 이상의 원인들의 결합에 의하여 발생된다. 이와 같이 여러가지 원인에 의하여 고장이 발생하는 제품의 각 고장원인별 수명을 추정하기 위해서는 다수 고장원인(multiple modes of failure)모형이 필요하게 된다. 다수 고장원인모형에 대한 기존연구로는 Herman 과 Patell(1971), David 와 Moeschberger(1978), Nelson(1982), Miyakawa(1984), Bai 와 Chun(1991) 등이 있다.

이 논문에서는 고장원인이 여럿있는 제품에 대하여 사용환경에서 얻어진 고장데이타와 추적조사에 의해 얻어진 설명변수에 관한 데이타를 이용하여 제품의 고장원인별 수명분포의 모수를 추정한다. 수명분포의 모수와 설명변수가 대수선형관계일 때, 의사 우도함수(pseudo likelihood function)를 세워 의사 최우추정량(pseudo maximum likelihood estimator)을 구하는 식을 유도하고, 특히 각 고장원인별 수명분포가 와이블분포일 때, 의사 최우추정량과 이들의 점근분산을 구한다. 또한 이러한 추정량의 점근성질을 규명하고, 모의실험을 통하여 추적조사 비율에 따른 효과에 대하여 조사한다. 이 논문에서 사용되는 기호는 다음과 같다.

N 총 시험제품의 수

k 제품의 고장원인 수

n 설명변수의 수

θ 제품 수명분포의 모수벡터

T_{im} 고장원인 m에 의하여 고장난 제품 i의 수명시간

$$i=1,2,\cdot\cdot\cdot,N \quad ; \quad m=1,2,\cdot\cdot\cdot,k$$

T_i 제품 i의 수명시간 ; $\min\{T_{i1},\cdot\cdot\cdot,T_{ik}\}$

x_i 제품 i에 대한 설명변수 열벡터 ; $(1, x_{i1}, x_{i2}, \cdots, x_{in})'$

r_i 제품 i의 고장원인

T^0 제품의 보증기간

D_1 보증기간동안 고장난 제품의 집합

D_1^c 보증기간동안 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 ; $p_2 = n_2/N_2$

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

2. 데이타 형태 및 추적조사 방법

수명분포의 모수를 추정하는데 사용되는 사용환경에서의 수명데이타들은 보증기간내에 고장이 발생한 제품의 수리를 위해 소비자가 서비스센터를 찾았을 때 얻어지며, 제품의 고장원인과 고장시간, 그리고 사용중인 제품의 제조특성(모형, 제조시간 또는 장소등), 환경특성(사용자의 특징, 기후조건등) 등과 같은 설명변수에 관한 데이타를 의미한다. 그러나 이 기간중 고장이 발생하지 않은 제품에 대하여는 수명시간이 보증기간보다 크다는 것 이외에는 설명변수에 대한 아무런 정보도 얻을 수 없게 된다. 이런 경우 보증기간 동안 고장이 발생하지 않은 제품에 대하여 추적조사를 통해 설명변수들을 알아냄으로써 수명분포의 모수를 추정하는데 이용할 수 있다. 이 논문에서는 추적조사를 위한 샘플링방법으로 총 N개의 시험제품에서 보증기간 동안 고장나지 않은 제품 N_2개 중 p_2 비율의 제품을 단순 랜덤샘플링에 의하여 추적조사하는 Kalbfleish 와 Lawless(1988)의 접근방법을 사용한다. 즉, 보증기간 T^0까지 고장난 제품으로부터 고장시간, 고장원인 및 설명변수에 관한 데이타를 얻고, 보증기간내에 고장이 발생하지 않은 제품의 일정 비율을 추적조사하여 제품의 설명변수에 관한 데이타를 얻는다. 이 논문에서는 우선 제품의 보증기간과 고장시간의 시간척도가 동일한 경우, 즉 제품의 보증기간과 고장시간이 모두 운영시간이나 달력시간에 의존하는 경우에 대하여 다루고, 보증기간과 고장시간의 시간 척도가 다른 경우에 대해서는 제 5 절에서 다룬다. 데이타의 형태와 추적조사 방법에 따른 가정은 다음과 같다.

① 제품에는 k개의 고장원인이 있으며, 고장원인별 수명은 서로 독립이다.

② 보증기간동안 고장이 발생한 모든 제품의 고장시간과 고장원인, 설명변수의 값은 기록

된다.

③ 보증기간동안 고장이 발생하지 않은 제품 중 추적조사되는 제품은 랜덤하게 선택되며, 선택된 제품의 설명변수에 대한 정보는 정확하게 얻을 수 있다. 또한 추적조사 비율 p_2는 0이 아니다.

3. 사용현장 데이타의 분석방법

이 절에서는 사용현장에서 얻어진 데이타를 이용하여 제품의 고장원인별 수명분포의 모수를 추정하기 위한 의사(pseudo) 우도함수를 세우고, 이 함수를 이용하여 의사 최우추정량을 구하며, 의사 최우추정량의 점근성질을 규명한다. 또한 고장원인별 수명분포가 와이블분포인 경우의 의사 최우추정량과 점근분산을 구한다.

3.1 의사 최우추정량 및 추정량의 점근성질

N개의 제품 모두에 대하여 설명변수 x_i들을 안다면, 우도함수 $L_F(\theta)$는 고장원인이 하나인 경우

$$L_F(\theta) = \prod_{i:t_i \le T^0} f(t_i \mid x_i;\theta) \prod_{i:t_i > T^0} \overline{F}(T^0 \mid x_i;\theta) \tag{1}$$

이고, 고장원인이 $k(>1)$개인 경우 $I[r_i=m]$을 제품 i에 고장이 발생했을 때 고장원인이 m인가를 나타내는 지시함수(indicator function) 라 하면

$$\begin{aligned} L_F(\theta) = & \prod_{i:t_i \le T^0} \prod_{m=1}^{k} \left[f_m(t_i \mid x_i;\theta) \prod_{l \ne m}^{k} \overline{F}_l(t_i \mid x_i;\theta) \right]^{I[r_i=m]} \\ & \cdot \prod_{i:t_i > T^0} \prod_{m=1}^{k} \overline{F}_m(T^0 \mid x_i;\theta) \end{aligned} \tag{2}$$

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 하면, 고장원인이 하나인 경우 Kalbfleisch 와 Lawless(1988)가 제안한 모수 θ에 대한 의사 우도함수 $L_p(\theta)$는 다음과 같다.

$$L_p(\theta) = \prod_{i \in D_1} f(t_i \mid x_i;\theta) \cdot \prod_{i \in D_2} [\overline{F}(T^0 \mid x_i;\theta)]^{1/p_2} \tag{3}$$

식 (3)에서 k개의 고장원인이 존재하면, 제품 i에 대한 데이타는 보증기간 T^0동안 고장이 발생하였을 경우에는 고장시간 t_i, 고장원인 r_i, 설명변수 x_i가 얻어지고, 보증기간 T^0동안 고장이 발생하지 않은 제품들 중에서 추적조사 비율 p_2로 샘플링된 경우에는, 설명변수 x_i가

얻어진다. 따라서 고장원인이 여럿인 경우 모수 θ에 대한 의사 우도함수 $L_p(\theta)$는

$$L_p(\theta) = \prod_{i \in D_1} \prod_{m=1}^{k} \left[f_m(t_i \mid x_i;\theta) \prod_{l \neq m}^{k} \overline{F}_l(t_i \mid x_i;\theta) \right]^{I[r_i=m]} \cdot \prod_{i \in D_2} \left[\prod_{m=1}^{k} \overline{F}_m(T^0 \mid x_i;\theta) \right]^{1/p_2} \tag{4}$$

이 되고, 식 (4)의 양변에 로그를 취하면, 다음과 같은 의사 대수우도함수를 얻는다.

$$\log L_p(\theta) = \sum_{i \in D_1} \sum_{m=1}^{k} I[r_i=m] \left[\log f_m(t_i \mid x_i;\theta) + \sum_{l \neq m}^{k} \log \overline{F}_l(t_i \mid x_i;\theta) \right] + \frac{1}{p_2} \sum_{i \in D_2} \sum_{m=1}^{k} \log \overline{F}_m(T^0 \mid x_i;\theta) \tag{5}$$

식 (5)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 다음과 같은 성질들을 갖는다.

정리 1 설명변수 x 와 모수 θ가 주어졌을 때 고장원인 m에 의한 제품 수명시간의 조건부 확률밀도함수 $f_m(t \mid x;\theta)$ 가 정칙조건(regularity conditions, Cramér, 1946, 33.3절 참조)을 따른다면, 의사 최우추정량 θ^*는 N이 커짐에 따라 다음과 같은 성질들을 갖는다.

(i) θ^*는 모수 θ의 일치추정량(consistent estimator)이다.

(ii) $\sqrt{N}(\theta^*-\theta)$ 는 N이 커짐에 따라 평균이 0 이고 분산-공분산 행렬이

$$V(\theta) = A(\theta)^{-1} + A(\theta)^{-1} C(\theta) A(\theta)^{-1} \tag{6}$$

인 다변량 정규분포를 따른다. 여기서 행렬 $A(\theta)$와 $C(\theta)$의 원소 $A(\theta)_{mr,ms}$ 와 $C(\theta)_{mr,ms}$ 는

$$A(\theta)_{mr,ms} = \lim_{N \to \infty} \frac{1}{N} E\left\{ -\frac{\partial^2 \log L_p}{\partial \theta_{mr} \partial \theta_{ms}} \right\} \tag{7}$$

$$C(\theta)_{mr,ms} = \frac{1-p_2}{p_2} \lim_{N \to \infty} \frac{1}{N} E\left\{ \frac{N_2}{n_2-1} \sum_{i \in D_2} (U_{imr} - \overline{U}_{mr})(U_{ims} - \overline{U}_{ms}) \right\} \tag{8}$$

이다. 단, 여기서 $U_{imr} = \partial \log \overline{F}_m(T^0 \mid x_i;\theta) / \partial \theta_{mr}$, $\overline{U}_{mr} = \sum_{i \in D_2} U_{imr}/n_2$, $r, s = 1,\cdots,n$; $m = 1,\cdots,k$ 이다.■

정리 1에 대한 증명은 부록 1에서 다룬다. 식 (7)과 (8)의 $A(\theta)_{mr,ms}$와 $C(\theta)_{mr,ms}$ 대신 다음과 같이 정의되는 $A_N(\theta^*)_{mr,ms}$와 $C_N(\theta^*)_{mr,ms}$ 를 대입하면 $V(\theta)$의 일치추정량 $V_N(\theta^*)$를 얻을 수 있다.

$$A_N(\theta^*)_{mr,ms} = -\frac{1}{N}\frac{\partial^2 \log L_p}{\partial\theta_{mr}\partial\theta_{ms}}$$
$$= -\frac{1}{N}\sum_{i\in D_1}\left\{\frac{\partial^2 \log f_m(t_i \mid x_i;\theta)}{\partial\theta_{mr}\partial\theta_{ms}} + \sum_{l\neq m}^{k}\frac{\partial^2 \log \overline{F}_l(t_i \mid x_i;\theta)}{\partial\theta_{mr}\partial\theta_{ms}}\right\}, \quad (9)$$
$$-\frac{1}{Np_2}\sum_{i\in D_2}\frac{\partial^2 \log \overline{F}_m(T^0 \mid x_i;\theta)}{\partial\theta_{mr}\partial\theta_{ms}}$$

$$C_N(\theta^*)_{mr,ms} = \frac{N_2(1-p_2)}{Np_2(n_2-1)}\sum_{i\in D_2}(U_{imr}-\overline{U}_{mr})(U_{ims}-\overline{U}_{ms}) \quad . \quad (10)$$

3.2 고장 원인별 수명분포가 와이블인 경우

각 고장원인별 제품 수명이 와이블분포를 따르고 그 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.

$$\overline{F}_m(t \mid x_i) = \exp(-t^{\delta_m}\cdot\exp(x_i'\beta_m)), \qquad t \geq 0 \quad , \quad (11)$$

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(11)을 이용하여 의사 대수우도함수를 구하면

$$\log L_p(\theta) = \sum_{m=1}^{k}\left\{\sum_{i\in D_1}I[r_i=m](\log\delta_m+(\delta_m-1)\log t_i+x_i'\beta_m) - \sum_{i\in D_1}t_i^{\delta_m}e^{x_i'\beta_m} - \frac{1}{p_2}\sum_{i\in D_2}((T^0)^{\delta_m}e^{x_i'\beta_m})\right\} \quad (12)$$

이고, 식(12)를 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

$$\sum_{i\in D_1}I[r_i=m]\ x_{ir} = \sum_{i\in D_1}t_i^{\delta^*_m}\ x_{ir}\ e^{x_i'\beta^*_m} + \frac{1}{p_2}\sum_{i\in D_2}(T^0)^{\delta^*_m}x_{ir}\ e^{x_i'\beta^*_m} \quad (13)$$

$$\sum_{i\in D_1}I[r_i=m]\left(\frac{1}{\delta^*_m}+\log t_i\right) = \sum_{i\in D_1}t_i^{\delta^*_m}\ (\log t_i)\ e^{x_i'\beta^*_m} + \frac{1}{p_2}\sum_{i\in D_2}(T^0)^{\delta^*_m}\ (\log T^0)\ e^{x_i'\beta^*_m} \quad (14)$$

을 만족하는 값이되며, 이는 수치적 방법에 의하여 구할 수 있다.

의사 최우추정량 $\beta^*_{mr}, \delta^*_m,\ m=1,2,\cdot\cdot\cdot,k\ ;\ r=0,1,2,\cdot\cdot\cdot,n$ 의 점근분산을 구하기 위해 식(9)와 (10)으로 부터 행렬 $A_N(\beta,\delta)$ 와 $C_N(\beta,\delta)$ 의 원소를 구하면

$$A_N(\beta,\delta)_{mr,ms} = \frac{-\partial^2 \log L_p}{\partial \beta_{mr} \partial \beta_{ms}} = \frac{1}{N}\sum_{i\in D_1} x_{ir}x_{is}W_{im0} - \frac{1}{Np_2}\sum_{i\in D_2} x_{ir}x_{is}H_{im} \tag{15}$$

$$A_N(\beta,\delta)_{mr,m(n+1)} = \frac{-\partial^2 \log L_p}{\partial \beta_{mr} \partial \delta_m} = \frac{1}{N}\sum_{i\in D_1} x_{ir}W_{im1} - \frac{1}{Np_2}\sum_{i\in D_2} x_{ir}(\log T^0)H_{im} \tag{16}$$

$$A_N(\beta,\delta)_{m(n+1),m(n+1)} = \frac{-\partial^2 \log L_p}{\partial \delta_m^2} = \frac{1}{N}\sum_{i\in D_1}(I[r_i=m]\delta_m^{-2}+W_{im2}) - \frac{1}{Np_2}\sum_{i\in D_2}(\log T^0)^2 H_{im} \tag{17}$$

$$C_N(\beta,\delta)_{mr,ms} = \frac{N_2(1-p_2)}{Np_2(n_2-1)}\sum_{i\in D_2}(x_{ir}H_{im}-\overline{H}_{mr})(x_{is}H_{im}-\overline{H}_{ms}) \tag{18}$$

$$C_N(\beta,\delta)_{mr,m(n+1)} = \frac{N_2(1-p_2)}{Np_2(n_2-1)}\sum_{i\in D_2}(\log T^0)(x_{ir}H_{im}-\overline{H}_{mr})(H_{im}-\overline{H}_m) \tag{19}$$

$$C_N(\beta,\delta)_{m(n+1),m(n+1)} = \frac{N_2(1-p_2)}{Np_2(n_2-1)}\sum_{i\in D_2}(\log T^0)^2(H_{im}-\overline{H}_m)^2 \tag{20}$$

이고, 여기서 $W_{imj} = (\log t_i)^j t_i^{\delta_m} e^{x_i'\beta_m}$, $H_{im} = -(T^0)^{\delta_m} e^{x_i'\beta_m}$, $\overline{H}_m = \sum_{i\in D_2} H_{im}/n_2$,

$\overline{H}_{mr} = \sum_{i\in D_2} x_{ir}H_{im}/n_2$, $j=0,1,2$; $r,s=1,2,\cdots,n$; $m=1,2,\cdots,k$ 이다.

식 (15)-(20)을 이용하면 분산-공분산행렬의 일치추정량을 구할 수 있고, 이로 부터 고장원인별 수명분포의 모수에 대한 의사 최우추정량의 점근분산을 구할 수 있다. 각 고장원인별 수명이 지수분포를 따르고, 그 고장률이 설명변수와 대수선형관계를 갖는 경우에는 식(11)-(20)에서 $\delta_m = 1$ 로 하여 구할 수 있다.

예제 설명변수가 있는 다수 고장원인모형의 설명을 위해서 어떤 전자제품의 중요한 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

$$\overline{F}_m(t \mid x_{i1}) = \exp(-t^{\delta_m}\cdot \exp(\beta_{m0}+\beta_{m1}x_{i1})), \quad t \geq 0, \quad m=1,2, \tag{21}$$

인 와이블분포를 따른다고 하자. Kalbfleish와 Lawless(1988)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, IMSL(1987) 서브루틴을 이용하여 난수를 발생한 결과 270개의 고장데이타를 얻었다. 270개의

고장데이타중 $(x=0,\ m=1)$, $(x=1,\ m=1)$, $(x=0,\ m=2)$, $(x=1,\ m=2)$ 일 때의 고장 갯수는 각각 25, 60, 40, 145이었고, 또한 38개월 동안 고장이 발생하지 않은 N_2 = 5,100 개의 제품중 $p_2=0.1$ 의 비율로 추적조사하여, 510개의 제품중 262개의 $x_{i1}=0$ 과 248개의 $x_{i1}=1$ 값을 얻었다. 이와 같은 데이타를 이용하여 식 (13)과 (14)의 의사 최우추정치를 IMSL(1987) 서브루틴을 이용하여 구하면

$$\beta_{10}^*=-12.0097,\ \beta_{11}^*=0.9003,\ \delta_1^*=1.9872$$

$$\beta_{20}^*=-15.0841,\ \beta_{21}^*=1.2794,\ \delta_2^*=3.0126$$

이며, 각 추정치들의 오차를 구해보면 다음과 같다.

$$error(\beta_{10}^*)=-0.0297,\ error(\beta_{11}^*)=0.0003,\ error(\delta_1^*)=0.0128$$

$$error(\beta_{20}^*)=0.0159,\ error(\beta_{21}^*)=-0.0206,\ error(\delta_2^*)=0.0126$$

식 (15)-(20)을 이용하여 분산-공분산 행렬 $V_N(\theta)$의 추정치 $V_N(\theta^*)$를 구하면

$$V_N(\theta^*)=\begin{bmatrix} 3038.88 & -542.37 & 8.61 & 0.0 & 0.0 & 0.0 \\ -542.37 & 158.42 & 39.47 & 0.0 & 0.0 & 0.0 \\ 8.61 & 39.47 & 327.03 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 3260.66 & -1264.24 & -5.16 \\ 0.0 & 0.0 & 0.0 & -2164.24 & 221.78 & 203.65 \\ 0.0 & 0.0 & 0.0 & -5.16 & 203.65 & 345.29 \end{bmatrix}$$

이며, N=5,370이므로 각 의사최우추정량들의 점근분산은 다음과 같다.

$$asvar(\beta_{10}^*)=0.5659,\ asvar(\beta_{11}^*)=0.0295,\ asvar(\delta_1^*)=0.0609$$

$$asvar(\beta_{20}^*)=0.6072,\ asvar(\beta_{21}^*)=0.0413,\ asvar(\delta_2^*)=0.0643$$

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	135.79	70.81
1	86.32	46.31

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.

제품에 고장원인이 2가지인 경우 Kalbfleish와 Lawless(1988)의 단일 고장원인모형을 이용하여 고장원인을 구별하지 않고 제품의 10백분위수를 구해 다수 고장원인모형과 비교해보면 다음과 같다.

x_{i1}	제품의 10백분위수의 참값	다수 고장원인모형에 의한 제품의 10백분위수의 추정치	단일 고장원인모형에 의한 제품의 10백분위수의 추정치
0	65.66	66.51	67.72
1	42.43	42.84	42.90

이로부터 고장원인이 2가지인 경우 고장원인을 구별하지 않고 단일 고장원인모형으로 구한 제품의 10백분위수의 추정치와 다수 고장원인모형을 이용하여 구한 추정치는 큰 차이가 없음을 알 수 있다. 따라서 제품에 고장원인이 여럿인 경우 고장원인별로 제품의 수명을 추정하기 위해서는 다수 고장원인모형을 이용하는 것이 바람직함을 알 수 있다.■

4. 추적조사 비율에 따른 의사최우추정량의 성질

추적조사 비율 p_2에 따른 의사 최우추정량 θ^*의 성질을 알아보기 위해 예제와 동일한 상황과 모수를 이용하여 $p_2 = 0.01$, 0.05, 0.1(0.1)1.0일 때의 10백분위수의 효율(efficiency)을 구하는 모의실험을 다음과 같이 수행하였다. 단, 총 시험제품수 N=1,000 으로 하였다.

<**모의실험절차**>

① N=1,000 중 500개를 랜덤하게 추출하여 x_{i1}=1로 하고 나머지 500개를 x_{i1}=0로 한다.

② 고장원인 1에 의한 수명시간에 대한 난수를 얻기 위해 x_{i1}=1 인 제품 i에 대하여는 $x_i'\beta_1 = -11.98 + 0.90$, $\delta_1 = 2.0$, 그리고 x_{i1}=0인 제품 i에 대하여는 $x_i'\beta_1 = -11.98$, $\delta_1 = 2.0$인 와이블분포에서 난수를 발생시킨다. 또한 고장원인 2에 의한 수명시간에 대한 난수를 얻기 위해 x_{i1}=1 인 제품 i에 대하여는 $x_i'\beta_2 = -15.10 + 1.30$, $\delta_2 = 3.0$, 그리고 x_{i1}=0인 제품 i에 대하여는 $x_i'\beta_2 = -15.10$, $\delta_2 = 3.0$인 와이블분포에서 난수를 발생시킨다.

③ 절차 ②에서 얻어진 제품 1,000개의 고장원인별 수명시간으로 부터 제품의 수명시간 $T_i = \min\{T_{i1}, T_{i2}\}$를 구한 후 $T_i \le T^0$인 제품 i를 D_1집합에 포함시킨다. 그리고 D_1^c

에 속한 제품들에 대하여 p_2의 비율로 샘플링한 제품들을 D_2집합에 포함시킨다.

④ 절차 ③에서 얻어진 정보를 식(13)과 (14)에 대입하여 β_{10}, β_{11}, δ_1, β_{20}, β_{21}, δ_2 의 의사 최우추정치를 구하고, 이를 이용하여 추적조사비율이 p_2일 때의 고장원인과 설명변수의 값에 따른 수명분포의 10백분위수 $t^m_{p_2}(x)$의 추정치 $t^m_{p_2}(x)^*$, $m=1,2$, $x=0,1$ 를 다음식을 이용하여 구한다.

$$t^m_{p_2}(x)^* = [-\log 0.9 \cdot e^{-x\beta^*_m}]^{\frac{1}{\delta^*_m}}$$

단, 여기서 $m=1,2$, $x=0,1$ 이고, $p_2=0.01$, 0.05, 0.1(0.1)1.0 이다.

⑤ ①에서 ④까지의 절차를 1,000번 반복하여 설명변수값과 고장원인별로 10백분위수 $t^m_{p_2}(x)^*$에 대한 효율을

$$eff[\,t^m_{p_2}(x)^*\,] = \frac{p_2=1.0\text{일 때의 } t^m_{1.0}(x)^*\text{의 평균제곱오차}}{t^m_{p_2}(x)^*\text{의 평균제곱오차}}$$

으로 구한다.

표 1. 의사최우추정량을 이용하여 구한 10백분위수의 p_2값에 따른 효율

p_2	고장원인1 (m = 1)		고장원인2 (m = 2)	
	설명변수 $x_{i1}=0$	설명변수 $x_{i1}=1$	설명변수 $x_{i1}=0$	설명변수 $x_{i1}=1$
0.01	0.746	0.701	0.715	0.699
0.05	0.790	0.761	0.762	0.783
0.1	0.833	0.818	0.813	0.846
0.2	0.889	0.874	0.871	0.883
0.3	0.932	0.911	0.917	0.907
0.4	0.954	0.945	0.938	0.940
0.5	0.976	0.969	0.967	0.962
0.6	0.985	0.980	0.978	0.976
0.7	0.990	0.991	0.989	0.984
0.8	0.998	0.997	0.997	0.996
0.9	0.999	0.999	0.999	0.999
1.0	1.000	1.000	1.000	1.000

표 1은 이와 같은 모의실험의 결과로서 고장원인별 수명분포가 와이블인 경우 추적조사 비율 p_2에 따른 10백분위수에 대한 의사 최우추정치의 효율을 나타낸 것이다. 표 1에 의하면 추적조

사 비율 p_2가 0.1 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.

5. 보증제도에 따른 분석방법

이 절에서는 보증제도의 시간척도와 고장시간의 시간척도가 다른 경우, 즉 제품의 보증기간은 달력시간이고 고장시간이 운영시간에 의존하는 경우와 보증기간이 달력시간과 운영시간의 혼합이고 고장시간이 운영시간에 의존하는 경우에 대하여, 의사 우도함수를 유도하고 이의 분석방법을 제시한다.

5.1 보증기간이 달력시간이고 고장시간이 운영시간에 의존하는 경우

각각의 제품에 대하여 $Y_i(t)$를 t시점에서 제품이 사용되고 있으면 1, 그렇지 않으면 0으로 정의되는 사용과정(usage process)이라 하면, t시간까지의 제품 i의 운영시간은 $\int_0^t Y_i(u)du$ 이 된다. 따라서 보증기간 T^0까지 고장이 발생한 제품은 설명변수 x_i, 고장원인 r_i, 고장시간 t_i 그리고 고장시간까지의 운영시간 $s_i = \int_0^{t_i} Y_i(u)du$ 가 관측된다. 또한 보증기간 T^0까지 고장이 발생하지 않은 제품에 대해서는 p_2의 비율로 추적조사 하게되며, 추적조사된 제품으로 부터 설명변수 x_i와 보증기간동안의 운영시간 $s_i' = \int_0^{T^0} Y_i(u)du$ 를 얻을 수 있게 된다. $f_m{}^0(s|x_i\ ;\theta)$, $\overline{F^0}_m(s|x_i\ ;\theta) = 1-\int_0^s f_m{}^0(u|x_i\ ;\theta)du$ 를 각각 설명변수 x_i와 모수벡터 θ가 주어졌을 때 m번째 고장원인에 의한 운영시간의 확률밀도함수와 신뢰도함수라 하고, 이를 이용하여 식 (5)와 유사한 방법으로 의사 대수우도함수를 구하면

$$\log L_p = \sum_{i\in D_1}\sum_{m=1}^{k} I[r_i=m]\left[\log f_m{}^0(s_i|x_i;\theta)+\sum_{l\neq m}^{k}\log \overline{F^0}_l(s_i|x_i;\theta)\right] + \frac{1}{p_2}\sum_{l\in D_2}\sum_{m=1}^{k}\log \overline{F^0}_m(s_i'|x_i;\theta) \tag{23}$$

이 된다. 따라서 이 경우 식 (5)와 확률밀도함수만 다르므로 $f_m{}^0(s|x_i\ ;\theta)$의 형태를 알면 제 3절에서 다룬 고장시간과 보증기간의 시간척도가 같은 경우의 방법과 동일하게 분석할 수 있다. 즉, 식 (23)을 최대로하는 θ^*를 구하면, 이때의 θ^*가 의사 최우추정량이 되며 θ^*는 N이 커짐에 따라 정리1과 같은 성질들을 갖는다.

5.2 보증기간이 달력시간과 운영시간의 혼합이고 고장시간이 운영시간에 의존하는 경우

판매된 제품을 달력시간이나 운영시간 중 하나 만을 보증해주는 경우도 있지만, 보증기간이 달력시간과 운영시간의 혼합된 형태인 경우도 있다. 예를 들어 우리나라 자동차의 보증은 자동차 구입 후 1년이내와 20,000 km의 운행 중 먼저 도달하는 것에 의한다. 이와 같이 보증기간이 달력시간과 운영시간의 혼합인 경우는 5.1절의 보증기간이 달력시간이고 고장시간이 운영시간에 의존하는 경우와 유사하게 분석할 수 있다.

T^0와 S^0를 각각 달력시간과 운영시간에 의한 보증기간이라하고, $Y_i(t)$를 5.1절에서 정의된 사용과정이라하자. 만약 i번째 제품의 고장이 발생하는 시점 t_i가 $t_i \le T^0$이고 t_i까지의 운영시간 $s_i = \int_0^{t_i} Y_i(u)du$ 가 $s_i \le S^0$이면, 이 제품은 보증기간 동안 고장나는 제품이 되고, 이 제품으로 부터 고장 데이타 (t_i, s_i, x_i, r_i) 가 관측된다. 또한 고장시간 t_i가 달력시간에 의한 보증기간 T^0보다 큰 경우와 $t_i \le T^0$이지만 고장시간 까지의 운영시간 s_i가 운영시간에 대한 보증기간 S^0보다 큰 경우는 보증기간 동안 고장이 발생하지 않은 제품이 된다. 이러한 제품들은 p_2의 확률로 추적조사되며, 설명변수 x_i와 $s_i^* = \min\left(S^0, \int_0^{T^0} Y_i(u)du\right)$ 가 관측된다.

따라서 $f_m^0(s|x_i;\theta)$, $\overline{F^0}_m(s|x_i;\theta)$를 각각 설명변수 x_i와 모수벡터 θ가 주어졌을 때 m번째 고장원인에 의한 운영시간의 확률밀도함수와 신뢰도함수라 하고, 이를 이용하여 식 (5)와 유사한 방법으로 의사 대수우도함수를 구하면

$$\log L_p = \sum_{i\in D_1}\sum_{m=1}^{k} I[r_i=m]\left[\log f_m^0(s_i|x_i;\theta)+\sum_{l\ne m}^{k}\log \overline{F^0}_l(s_i|x_i;\theta)\right] + \frac{1}{p_2}\sum_{i\in D_2}\sum_{m=1}^{k}\log \overline{F^0}_m(s_i^*|x_i;\theta) \tag{24}$$

이 된다. 따라서 이 경우도 식 (23)과 유사하게 식 (5)에 비해 확률밀도함수 및 데이타의 형태만 다르므로 $f_m^0(s|x_i;\theta)$의 형태를 알면 제 3절에서 다룬 고장시간과 보증기간의 시간척도가 같은 경우의 방법과 동일하게 분석할 수 있다. 즉, 식 (24)를 최대로하는 θ^*를 구하면, 이때의 θ^*가 의사 최우추정량이 되며 θ^*는 N이 커짐에 따라 정리1과 같은 성질들을 갖는다.

6. 결 론

이 논문에서는 k(>1)개의 고장원인이 있는 제품에 대하여 사용환경에서 얻어진 고장데이타와 추적조사된 정보를 이용하여 고장원인별 수명분포를 추정하는 문제를 다루었다. 제품수명분포의 모수가 설명변수와 대수선형 관계일 때 의사 우도함수를 세워 의사 최우추정량을 구하였

으며, 의사 최우추정량의 점근성질을 규명하였다. 이러한 결과들은 여러가지 보증제도에 대해서도 쉽게 확장될 수 있음을 알 수 있다. 또한 모의실험을 통하여 추적조사 비율에 따른 의사 최우추정량의 점근효율을 구한 결과 추적조사비율이 그리 크지 않더라도 의사 최우추정량을 이용하여 구한 10백분위수의 추정량의 효율이 크게 떨어지지 않음을 알 수 있다.

이 논문에서는 보증기간동안 제품의 첫번째 고장만을 관측하는 경우에 대하여 다루었지만, 실제로 사용현장에서 사용되는 제품은 보증기간동안 고장이 발생하면 수리하거나 교체를 하여 다시 사용되며, 이러한 제품에 다시 고장이 발생할 수 있다. 따라서 수리가능한 제품에 대한 사용현장 데이타의 분석에 관한 연구가 이루어져야 할 것이다.

부록 : 정리 1의 증명

Inagaki(1973)와 Crowder(1986)는 일반적인 우도함수 $L(\theta)$에 포함된 확률밀도함수 $f(t)$가 정칙조건(Cramér, 1946, 33.3절 참조)을 따르고 $S(\theta)=\partial \log L(\theta)/\partial\theta$라 할 때, ① $E[S(\theta)]=0$이면 $S(\theta)=0$을 만족하는 추정량 $\hat{\theta}$는 모수 θ의 일치추정량이고, ② $\sqrt{N}(\hat{\theta}-\theta)$는 점근적으로 평균 0이고 분산-공분산 행렬이 $A(\theta)^{-1}B(\theta)A(\theta)^{-1}$인 다변량 정규분포를 따름을 보였다. 단, 이때의 $A(\theta)$와 $B(\theta)$는 각각 다음과 같다.

$$A(\theta)=\lim_{N\to\infty}\frac{1}{N}E\left(-\frac{\partial S(\theta)}{\partial\theta}\right) \tag{A1}$$

$$B(\theta)=\lim_{N\to\infty}\frac{1}{N}E(S(\theta)S(\theta)') \tag{A2}$$

여기서는 이들의 결과를 이용하여 정리 1을 증명한다.

(i) θ^*가 모수 θ의 일치추정량임을 증명

N개의 모든 제품에 대하여 보증기간 동안의 모든 정보를 아는 경우 식(2)에 대수를 취하면

$$\begin{aligned}\log L_F(\theta) \;=\; &\sum_{i\in D_1}\sum_{m=1}^{k} I[r_i=m]\Big[\log f_m(t_i\mid x_i;\theta)+\sum_{l\neq m}^{k}\log \overline{F}_l(t_i\mid x_i;\theta)\Big]\\ &+\sum_{i\in D_1^c}\sum_{m=1}^{k}\log \overline{F}_m(T^0\mid x_i;\theta)\end{aligned} \tag{A3}$$

이 되고, 식(5)의 $\log L_p(\theta)$를 식 (A3)를 이용하여 다시 정리하면

$$\log L_p(\theta) \;=\; \log L_F(\theta)+\sum_{i\in D_1^c}\sum_{m=1}^{k}\left(\frac{R_{2i}}{p_2}-1\right)\log \overline{F}_m(T^0\mid x_i;\theta) \tag{A4}$$

이 된다. 여기서, R_{2i}는 제품 i의 추적조사 여부에 대한 지시함수로 $R_{2i}=I[i\in D_2]$이고, $p_2=P_r(R_{2i}=1\mid i\in D_1^c)$ 이다.

$S_p(\theta)=\partial \log L_p(\theta)/\partial\theta,\quad S_F(\theta)=\partial \log L_F(\theta)/\partial\theta,\quad U_{im}(\theta)=\partial \log \overline{F}_m(T^0\mid x_i;\theta)/\partial\theta$

라 하고, 식(A3)의 양변을 모수벡터 θ로 미분 하면,

$$S_p(\theta) = S_F(\theta) + \sum_{i\in D_1^c}\sum_{m=1}^{k}\left(\frac{R_{2i}}{p_2}-1\right)U_{im}(\theta) \tag{A5}$$

이 된다. $i\in D_1^c$에 대하여 $E[R_{2i} \mid (t_i,x_i),\ i=1,\cdots,N] = p_2$이므로 식(A5)의 양변에 기대값을 취하면

$$E[S_p(\theta) \mid (t_i,x_i),\ i=1,\cdots,N] = S_F(\theta) \tag{A6}$$

가 되고, 또한

$$E[E[S_p(\theta) \mid (t_i,x_i),\ i=1,\cdots,N]] = E[S_F(\theta)] \tag{A7}$$

이므로 $E[S_p(\theta)]=0$ 이 된다. 따라서 Inagaki(1973)와 Crowder(1986)의 결과를 이용하면 $S_p(\theta)=0$ 을 만족하는 의사 최우추정량 θ^*는 모수 θ의 일치추정량이다.

(ii) 의사 최우추정량의 분산-공분산행렬 유도

$S_p(\theta)=0$ 는 불편추정식이므로 $E(S_p(\theta)S_p(\theta)')=Var(S_p(\theta))$ 이고, 식(A4)의 우변의 두 항 $S_F(\theta)$와 $\sum_{i\in D_1^c}\sum_{m=1}^{k}(R_{2i}/p_2-1)U_{im}(\theta)$는 서로 독립이므로

$$Cov\left(S(\theta),\ \sum_{i\in D_1^c}\sum_{m=1}^{k}\left(\frac{R_{2i}}{p_2}-1\right)U_{im}(\theta)\right) = 0 \tag{A8}$$

이다. 식(A4)에서 $S_F(\theta)$는 $p_2=1$ 인 일반적인 대수 우도함수를 모수 θ에 대하여 편미분한 함수이므로 $S_F(\theta)$의 분산-공분산행렬 $V(S_F(\theta))$는

$$\begin{aligned} V(S_F(\theta)) = -E\Big[&\sum_{i\in D_1}\sum_{m=1}^{k}I[r_i=m]\Big[\frac{\partial^2}{\partial\theta\,\partial\theta'}\log f_m(t_i \mid x_i;\theta) \\ &+ \sum_{l\neq m}^{k}\frac{\partial^2}{\partial\theta\,\partial\theta'}\log\overline{F}_l(t_i \mid x_i;\theta)\Big] \\ &+ \sum_{i\in D_1^c}\sum_{m=1}^{k}\frac{\partial^2}{\partial\theta\,\partial\theta'}\log\overline{F}_m(T^0 \mid x_i;\theta)\Big] \end{aligned} \tag{A9}$$

이고, k개의 $\sum_{i\in D_1^c}(R_{2i}/p_2-1)U_{im}(\theta)$들은 서로 독립이므로

$$V\left(\sum_{i\in D_1^c}\sum_{m=1}^{k}\left(\frac{R_{2i}}{p_2}-1\right)U_{im}(\theta)\right) = \sum_{m=1}^{k}V\left(\sum_{i\in D_1^c}\left(\frac{R_{2i}}{p_2}-1\right)U_{im}(\theta)\right) \tag{A10}$$

이 된다. Kalbfleisch 와 Lawless(1988)의 결과를 이용하면

$$\begin{aligned} &V\left(\sum_{i\in D_1^c}\left(R_{2i}\frac{1}{p_2}-1\right)U_{im}(\theta)\right) \\ &= \frac{1-p_2}{p_2}E\left\{\frac{N_2}{n_2-1}\sum_{i\in D_2}(U_{imr}-\overline{U}_{mr})(U_{ims}-\overline{U}_{ms})\right\} \end{aligned} \tag{A11}$$

가 되며, $E(S_p(\theta)S_p(\theta)')$는 식(A9)와 식(A11)을 더한 값이된다. 따라서,

$$B(\theta) = \lim_{N\to\infty} \frac{1}{N} E(S_p(\theta)S_p(\theta)') = A(\theta) + C(\theta) \tag{A12}$$

이고, 의사 최우추정량의 분산-공분산행렬은

$$A(\theta)^{-1}B(\theta)A(\theta)^{-1} = A(\theta)^{-1} + A(\theta)^{-1}C(\theta)A(\theta)^{-1} \tag{A13}$$

이 된다.

참고문헌

[1] Bai, D. S. and Chun, Y. R. (1991). Optimal simple step-stress accelerated life tests with competing causes of failure, *IEEE Transaction on Reliability,* R-40, 622-627.

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

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

[4] David, H. A. and Moeschberger, M. L. (1978). *The theory of competing risks,* Griffin, London.

[5] 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.

[6] Herman, R. J. and Patell K. N. (1971). Maximum likelihood for multi-risk model, *Technometrics,* Vol. 13, 385-396.

[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] International Mathematical and Statistical Libraries, Inc. (1987). *IMSL library: reference manual,* Houston.

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

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

[11] Miyakawa, M. (1984). Analysis of incomplete data in competing risks model, *IEEE Transaction on Reliability,* R-33, 293-318.

[12] Nelson, W. (1982). *Applied Life Data Analysis,* John Wiley, New York.

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

[14] 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 Products with Multiple-Modes of Failure[1)]

Do Sun Bai, In Su Choi[2)], Yong Keun Hwang[3)]

Abstract

This paper is concerned with the method of estimating lifetime distribution from field data for products with multiple modes of failure. When product failures occur within warranty period, a manufacturer can obtain failure-record data; failure times, causes of failure, and covariates. Since these data are seriously incomplete for satisfactory inference, that is, only failures occured during warranty period may be recorded, it is usually necessary to incoporate the failure-record data by taking a supplementary sample of items obtained following up a portion of products that survive warranty time. The log linear function is considered as a model for describing the relation between failure time of a product and covariates. General methods for obtaining pseudo maximum likelihood estimators(PMLEs) for the parameters are outlined and their asymptotic properties are studied, and specific formulas for exponential or Weibull distribution are obtained. Effects of follow-up percentage on the PMLEs are investigated. Extensions to calendar time warranty or calendar and operating time warranty are also considered.

---

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

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

3) CUBIC TEK Co., LTD., Seoul 105-402, KOREA.