보증기간을 고려한 대수정규분포를 따르는 시장자료의 신뢰성 분석

Reliability Analysis for Field Data following Lognormal Distribution after Warranty Period

김종걸* · 최영진** · 정연승*
성균관대학교 시스템경영공학부*
삼성전자**

ABSTRACT

This paper is concerned with the method of estimating lifetime distribution for field data in warranty period and for a situation where some additional field data can be gathered after the warranty period. Implementing the proposed methods in this paper will result in obtaining the more precise product life time estimation and product improvement.

1. 서 론

오늘날 첨단기술산업에서 국제경쟁력을 확보하기 위해서는 신제품 개발과 장기간 수명을 보장하는 고 신뢰성 제품의 개발이 최우선 과제이다. 그러나 제품의 복잡화에 따른 고장빈도의 증가로 제품의 고장형태 분석기법의 개발이 시급히 요구되고 있다. 급속한 기술발전으로 인해 신기술, 신 재료 등에 대한 성능평가가 미흡해 짐에 따라 불신뢰·불안전이 확산되면서 신뢰성 평가기간의 단축이 요구되는 실정이다. 또한 사용자 요구가 다양화하고 극단적 환경에서 사용되는 제품이 증가함으로써 사용환경에 대해 강건성을 갖는 제품개발 및 평가기술이 절실히 필요하게 되었다.

그러나 대다수의 기업에서는 수집된 현장 데이터를 분석하여 제품의 수명모수 추정 및 제품의 신뢰도 측정을 할 때에 데이터가 미비하고 혹은 현장자료 수집에 있어서 많은 시간과 비용이 요구될 뿐만 아니라 랜덤 샘플링을 통해서 얻기 어렵다는 문제점으로 인하여 과거데이터의 활용 효율성이 매우 낮은 실정이다. 즉 제품에 고장이 발생하면 소비자는 A/S를 받기위하여 서비스센터를 찾게 되고, 따라서 서비스센터에 들어온 제품으로부터 고장 시간, 제조 특성, 사용 환경 등과 같은 고장에 관련된 정보가 얻어지게 된다. 그러나 고장이 발생하지 않았거나 고장이 발생한 제품이라 할지라도 서비스센터에 들어오지 않으면 제품수명에 대한 정보를 얻을 수 없게 된다. 특히 보증기간 안에 고장이 발생한 제품은 보증제도의 적용을 받으므로 대부분의 제품이 서비스센터에 들어오지만, 보증기간 이후에 고장이 발생한 제품에 대해서는 보증제도의 적용을 받지 못하므로 소비자의 성향에 따라 고장이 발생한 제품중의 일부분만이 서비스센터에 들어오게 된다.

본 연구에서는 보증기간 동안의 사용현장 데이터와 보증시점 이후 분석시점까지의 추가적으로 얻어진 사용현장 데이타를 이용하여, 제품의 수명모수가 대수정규분포를 따르는 경우, 모델의 모수를 추정하는 최우추정방법을 통해서 우도함수와 최우추정량 및 추정량의 점근분산을 구한다.

이 연구에서 사용되는 기호는 다음과 같이 정의한다.

N	총 판매 제품수(부품수)
θ	수명분포의 모수벡터
$f(t_i\,;\,\theta)$	i번째 제품의 수명 t_i 에서의 확률밀도 함수
$R(t_i\,;\,\theta)$	i번째 제품의 수명 t_i 에서의 신뢰도 함수
T_1	제품(부품)의 보증시점
T_2	제품 수명의 분석시점 (이 경우 $T_2 > T_1$)
p_1	T_1과 T_2사이에서 고장이 발생한 제품이 서비스센터에 들어올 확률
S_1	N개의 제품(부품) 중 T_1 이전에 서비스센터에 들어온 제품의 집합
S_2	T_1과 T_2 사이에서 서비스센터에 들어온 제품의 집합
S_3	T_2까지 서비스센터에 들어오지 않은 제품의 집합
K_{ji}	$t_i \in S_j$ 이면 1, $t_i \notin S_j$ 이면 0의 값을 갖는 지시변수 $j=1,2,3$
n_j	집합 S_j 에 속한 제품(부품)의 수, $j=1,2,3$

2. 수명분포 기본모형

2.1 이론적 모형

현재 많은 기계작업에서는 밑이 e인 대수정규분포를 많이 사용하고 있으며, 이장에서 추가적으로 사용되는 기호는 다음과 같다.

μ	대수정규분포의 평균
σ	대수정규분포의 표준편차
$F(t_i)$	t_i시점까지의 누적분포함수
T	제품의 수명
$\Phi(\frac{t_i-\mu}{\sigma})$	표준화 정규누적 분포값

2.2 보존기간과 자료의 형태

보증기간 내에 제품에 고장이 발생하면 대부분의 소비자는 서비스센터를 찾아 제품이나 부품을 수리 또는 교체하기 때문에 제품의 고장시간이 모니터 제도등에 의해 기록이 된다. 따라서 제품의 수명에 관한 대부분의 사용현장자료를 얻을수가 있지만 보증기간 이후에 고장이 발생하는 경우 소비자의 성향에 따라 서비스센터를 찾아 고장수리를 할수도 있고, 하지 않을 수도 있기 때문에 제품의 수명에 관한 데이터를 얻지 못할 수도 있다. 따라서 분석시점에서는 사용현장 데이터의 취득과정(그림1)과 같이 서비스 센타에 들어온 제품과 그렇지 못한 제품으로 분류된다.

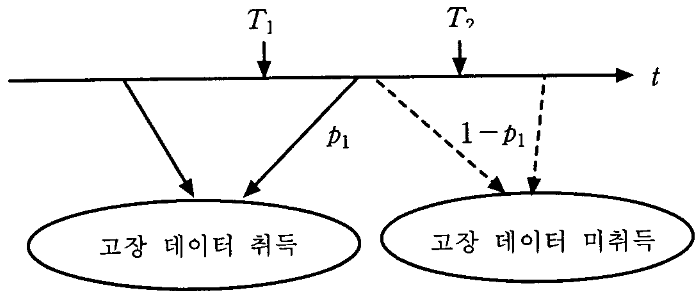


그림 1. 사용현장 데이터의 취득 과정

2.3 우도함수 및 최우추정량

대수정규 분포의 우도함수 및 최우 추정량을 구하기 위해 먼저 제품의 수명 모수가 μ , σ 를 따르는 대수정규분포의 경우를 가정하면, 확률밀도 함수와 신뢰도 함수는 각각

$$f(t_i;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma t_i}\exp[-(\log t_i-\mu)^2/2\sigma^2],\ t_i>0 \tag{1}$$

$$\begin{aligned}R(t_i;\mu,\sigma) &= \Pr(T\geq t_i)\\ &=\int_{t_i}^{\infty}\frac{1}{\sqrt{2\pi}\sigma t_i}\exp[-(\log t_i-\mu)^2/2\sigma^2]dt, \qquad t_i>0\\ &=1-\Phi(\frac{\log t_i-\mu}{\sigma})\end{aligned} \tag{2}$$

이다[18].

대수우도 함수를 구하면

$$\begin{aligned}\log L(\mu,\sigma) &= \sum_{i\in S_1}[\frac{1}{\sqrt{2\pi}\sigma t_i}-\frac{(\log t_i-\mu)^2}{2\sigma^2}]\\ &+\sum_{i\in S_2}[\log p_1+\frac{1}{\sqrt{2\pi}\sigma t_i}-\frac{(\log t_i-\mu)^2}{2\sigma^2}]\\ &+\sum_{i\in S_3}\log[p_1(1-\int_{T_1}^{T_2}\frac{1}{\sqrt{2\pi}\sigma t_i}\exp[-(\log t_i-\mu)^2/2\sigma^2]dt)\\ &+(1-p_1)(1-\int_0^{T_1}\frac{1}{\sqrt{2\pi}\sigma t_i}\exp[-(\log t_i-\mu)^2/2\sigma^2]dt)]\end{aligned} \tag{3}$$

로 되며, 식(3)을 μ 와 σ에 대하여 각각 일차, 이차 편미분 하고 이를 바탕으로 관측치로 부터 얻은 총 제품 수 N에 따른 피셔정보 행렬을 구하면

$$I_N^0 = \frac{1}{N}\begin{bmatrix} -\frac{\partial^2 \log L}{\partial \mu^2} & -\frac{\partial^2 \log L}{\partial \mu \partial \sigma} \\ -\frac{\partial^2 \log L}{\partial \mu \partial \sigma} & -\frac{\partial^2 \log L}{\partial \sigma^2} \end{bmatrix}, \mu = \mu^*, \sigma = \sigma^* \qquad (4)$$

최우추정량 μ^* 와 σ^*의 총 제품 수 N에 따른 관측된 점근 분산, 공분산 행렬은 피셔정보행렬의 역행렬의 전체 제품수에 대한 합 $\sum \frac{1}{N}\,(I_N^0)^{-1}$으로 정의된다.

여기서 최우추정량 μ^* 와 σ^*를 이용하여 제품수명의 $100\times p$ 백분위수 t_p의 최우정량 t_p를 구하면

$$t_p^*(\mu^*,\ \sigma^*) = \exp(\mu^* + Z_{p_i}\sigma^*) \qquad (5)$$

이고 총 제품수 N이 커질 때 t_p^*가 근사적으로 정규분포를 따른다.

$$\sqrt{N}\,(t_p^*(\mu^*, \sigma^*) - t_p(\mu, \sigma)) \xrightarrow{L} \mathrm{N}(0\,, J(\mu, \sigma)' I_0^{-1} J(\mu, \sigma)) \qquad (6)$$

여기서,

$$J(\mu, \sigma)' = \left[\frac{\partial}{\partial \mu} t_p(\mu, \sigma) \quad \frac{\partial}{\partial \sigma} t_p(\mu, \sigma)\right]$$

$$= \left[e^{\mu + Z_p \sigma}\ ,\ Z_p e^{\mu + Z_p \sigma}\right]$$

로 표현된다.

3. 예제 및 비교분석

이 절에서는 앞에서 유도한 대수정규분포의 우도함수 및 점근분산을 구하는 식을 이용하여 예제를 통해 보증기간 이후의 자료를 추가한 제품의 수명모수를 구하고 p_1값에 대한 편의와 평균제곱오차로서 최우추정량의 효과를 분석해 본다.

3.1 고장데이타 및 결과

데이터의 수집 및 정리는 특성치의 분포를 결정하는 기초적인 단계로서 제품 특성치의 신뢰성을 결정하는데 가장 중요한 자료로 사용된다. 여기서 사용한 데이터는 고장시간을 랜덤하게 발생시켜 얻은 데이터이다.

다음과 같은 데이터를 가정하여 각 수명분포의 모수를 추정하고 t_p의 수명 백분위수 및 점근분산을 추정한다.

예제1) 유형 1 : 보증기간 (T_1) 이내에 고장난 제품

유형 2 : 보증기간 (T_1) 과 분석시점 (T_2) 이내에 고장이 발생하여 서비스센터에 들어온 제품

어느 기계제품의 수명은 모수가 $\mu=200$ 이고 $\sigma=10$을 따르는 정규분포를 가정하고, 보증기간 내에 고장이 발생한 제품은 보증제도의 적용을 받아 무상으로 서비스를 받으므로 모두 서비스 센터에 들어오게 되고, 보증기간 이후의 제품은 모두 들어오지는 않지만 비교적 저렴한 가격으로 수리를 받을 수 있으므로, 대부분이 서비스 센터에 들어와서 수리를 받게 된다고 가정한다. 총 제품의 수는 50대, 보증기간은 200(일) 분석시점은 400(일) 이라고 가정하고, 보증기간이후에서 분석시간 까지 고장난 부품은 $p_1=0.5$ 의 확률로 들어온다고 하여 난수를 발생시킨 결과 <표1> 의 고장시간에 관한 데이터를 얻었다. 보증기간 이내에 들어온 제품의 수 n_1은 13대, 보증기간에서 분석기간 사이에 들어온 제품의 수 n_2 는 7대, 보증시점과 분석시점 사이에서 고장이 발생하였지만 서비스센터에 들어오지 않은 제품의 수 n_3는 30 대이다. <표1>을 대수정규확률밀도 함수에 대한 그림으로 표현하면 다음과 같다.

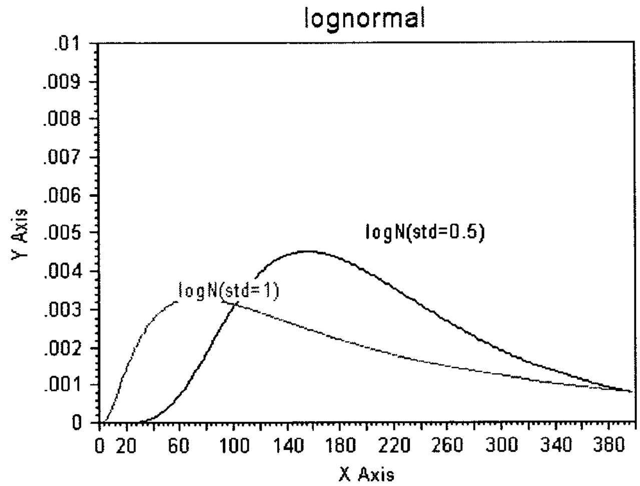


그림 2. 예제1) 데이터에 의한 대수정규분포의 확률밀도 함수

표 1. 고장 데이타 예

No.	고 장 시 간	유 형	No.	고 장 시 간	유 형
1	197.13170	1	11	197.05935	1
2	197.19529	1	12	207.00314	2
3	197.99360	1	13	183.88609	1
4	200.51390	2	14	189.80297	1
5	193.72990	1	15	199.64157	1
6	211.02496	2	16	195.22088	1
7	211.87282	2	17	200.14863	2
8	198.81257	1	18	199.44871	1
9	203.22637	2	19	192.87049	1
10	210.17682	2	20	198.54373	1

<표1>을 이용하여 μ^* 와 σ^*를 구하면,

$$\mu^*=5.293$$

$$\sigma^*=0.224$$

가 된다.

식(4)을 이용하여 점근 분산을 구하면,

$$As\,Var(\mu^*) = 0.0236$$

$$As\,Var(\sigma^*) = 0.000002706$$

또한 식(5) 와 식(6)를 이용하여 10백분위수에 대한 최우추정치 $t_p^*(\mu^*, \sigma^*)$와 점근분산을 구해보면

$$t_{0.1}^* = 149.282$$

$$As\,Var(t_{0.1}^*) = 4.53816$$

이다. 이로서 이 기계제품의 10 백분위수는 149.282일로 볼수가 있다.

최우 추정량 $\mu^* = 5.293$ $\sigma^* = 0.224$ 와 $t_{0.1}^* = 149.282$은 점근적으로 정규분포를 따르므로, 수명분포 모수 μ , σ 와 10 백분위수 $t_{0.1}$의 95% 신뢰구간은 다음과 같다.

$$4.99189 \le \mu^* \le 5.5941$$

$$0.22077 \le \sigma^* \le 0.2272$$

$$145.1066 \le t_{0.1} \le 153.4573$$

가 된다.

예제 1)에 대한 p_1값에 따른 수명모수 μ^* 와 $t_{0.1}^*$ 값을 나타내면 <표2>와 같다.

표 2. 예제1)에 대한 p_1값에 따른 수명모수 및 백분위수

p_1	μ^*	$t_{0.1}^*$
0.1	5.0985	132.8916
0.2	5.1250	127.9877
0.3	5.1760	130.3532
0.4	5.2370	139.3532
0.5	5.2930	149.2812
0.6	5.3435	159.6522
0.7	5.3905	170.5840
0.8	5.4355	182.3672
0.9	5.4835	196.8075

<표2>에서 보면, p_1값에 따른 수명모수 μ^*와 백분위수 t_p^* 는 점진적으로 증가하고 있음을 알 수 가있다.

3.2 비교 및 분석

이 절에서는 예제를 통하여 p_1의 변화가 백분위수들의 최우추정량들에 미치는 효과를 편의(bias)와 평균제곱오차(mean squared error :MSE)로서 분석을 실시한다.

3.2.1 p_1값에 따른 최우추정량의 효과분석

보증시점과 분석시점 사이에서 고장이 발생한 제품중 일부분이 p_1의 확률로 들어오므로 p_1값에 따라서 최우추정량의 정확도 및 신뢰성 분석에 있어서 영향을 미치게 된다. 따라서 이 절에서는 p_1값에 따른 최우추정량의 효과를 알아보기 위해 p_1값을 0.1에서 0.9 까지 변화시켜 가면서 수명모수들의 최우추정치의 편의와 평균제곱오차를 알아본다.
예제1)과 같이 제품수명분포가 대수정규를 따르는 경우의 평균과 표준편차로부터 각 고장시간에 대수를 취한 값의 평균 $\mu_0 = 5.29405$ 을 구한 후, p_1값에 따른 수명모수 값에 대한 성질을 알아보고자 각 p_1값에 대하여 제품 수명모수의 최우추정치를 30번 이상 구하여 계산한 후 수명모수의 편의와 평균제곱오차를 계산하는 방식을 취한다.

각 p_1값에 따른 편의와 평균제곱오차의 값은 다음과 같이 구한다.

편의 : $Bias(\mu^*) = \sum_{i=1}^{30} \frac{\mu_i^*}{30} - \mu_0$

평균제곱 오차 : $MSE = \sum_{i=0}^{30} \frac{(\mu_i^* - \mu_0)^2}{30}$

표 3 p_1값에 따른 μ_i^*의 편의와 평균제곱 오차

p_1	편 의	평균제곱 오차
0.1	-0.5116	0.26269850
0.2	-0.4837	0.23544879
0.3	-0.3736	0.13963529
0.4	-0.2908	0.08462638
0.5	-0.2260	0.05115696
0.6	-0.1725	0.02979821
0.7	-0.1262	0.01592978
0.8	-0.0845	0.00726988
0.9	-0.0426	0.00240489

표3은 대수정규 분포의 경우 p_1값에 따른 수명 모수의 최우추정량에 대한 편의와 평균제곱오차를 구한 표이다.

이를 요약하면 다음과 같다.

· p_1 값이 증가함에 따라 제품의 수명모수 μ값의 최우추정치의 평균제곱오차가 감소한다.

· p_1 값이 증가함에 따라 제품의 수명모수 μ값의 편의가 줄어든다.

· p_1 값이 0.6 이상이 되면 제품의 수명모수 μ값의 편의의 변화가 작다.

· p_1 값이 0.5 이상이 되면 제품의 수명모수 μ값의 평균제곱오차의 변화가 작다.

p_1값이 0.1에서 0.9까지로 증가함에 따라 서비스 센타에 들어오는 사용현장 데이터가 많아지므로 최우추정량의 정확도가 높아질 것이라는 가정을 할 수가 있다. 대수정규분포의 경우

에 제품 수명모수 μ의 편의는 p_1값이 커짐에 따라 거의 0에 가깝게 되고, 평균제곱오차 또한 p_1값이 커짐에 따라 최우추정치의 오차가 줄어듬을 알 수 있다. 따라서 p_1값이 커짐에 따라 보증기간 이후의 사용환경자료를 사용한 신뢰성 분석에 있어서 제품의 수명모수 μ는 p_1값이 1에 가까울수록 최우추정치의 정확도는 높아짐을 알 수가 있다.

3.3 p_1값을 모르는 경우

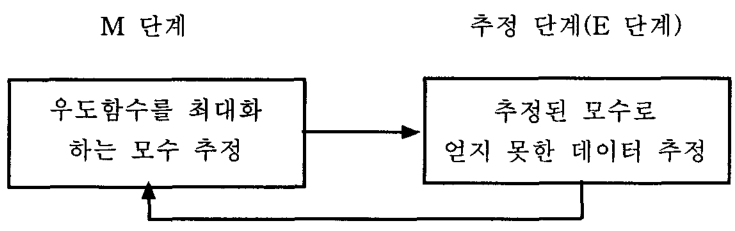
기업의 입장에서는 p_1값을 알고 있거나 알 수 있는 경우가 대부분이다. 그러나 신규사업 진출 등과 같이 새로운 제품을 생산할 경우에는 p_1값을 추정해야 하는 경우가 있을 수가 있다. 일반적으로 p_1값을 추정하기 위해서는 과거자료를 이용한다.

p_1값을 추정하기 위해서는 보증시점과 분석시점 사이에서 고장이 발생하여 서비스센터에 들어온 제품과 서비스센터에 들어오지 않은 제품의 수를 알아야 한다. 그러나 이때, 보증시점과 분석시점 사이에 고장이 발생하여 서비스센터에 들어온 제품의 수는 알 수가 있지만, 대부분의 경우 실제 고장이 발생했더라도 소비자가 제품 생산 회사의 서비스센터를 이용하지 않는 경우도 있으므로 보증시점과 분석시점 사이에는 모든 고장 데이터가 들어오지 않은 경우도 있다. 이렇게 얻어지지 않은 데이터는 실제 신뢰성 분석에 있어 사용하지 않거나, 얻어진 데이타의 평균으로 얻어지지 않은 데이터를 대체하거나, 또는 EM(expection and maximazition) 알고리즘과 같이 얻어진 데이터로 얻어지지 않은 데이터를 반복적으로 추정하는 방법이 있다.

여기서는 이러한 방법들 중에서 EM 알고리즘을 이용하여, 제품 수명분포의 모수와 p_1을 동시에 추정하는 방법을 살펴본 후 예제를 통하여 대수정규분포의 경우에 EM 알고리즘을 적용해 보기로 한다.

3.3.1 EM 알고리즘

EM 알고리즘은 얻지 못한 데이터가 있을 때 , 얻어진 데이터를 이용하여 얻지 못한 데이터의 기대치를 구하고 우도함수를 최대화하는 과정을 반복하여 모수들을 추정하는 방법이다. EM 알고리즘은 그림3 과 같이 크게 E-과정 (expectation step)과 M-과정(maximization)으로 나누어진다[10].



< 그림 3. EM 알고리즘 >

p_1을 모르는 경우 EM 알고리즘을 이용하여 p_1과 제품의 수명모수들을 추정하는 방법은 다음과 같이 진행된다.

① 임의로 p_1의 초기값을 설정한다.

② p_1을 참값으로 하여 T_2까지의 데이터를 가지고 식(3)의 우도함수를 최대로 하는 수명

모수를 추정한다.

③ 추정된 수명모수를 이용하여 보증시점과 분석시점 사이에 고장이 발생하였지만 서비스 센터에 들어오지 않은 제품 수의 기대치 m을 다음의 식으로 구한다.

$$m = n_3 \times \frac{(1-p_1)[R(T_1)-R(T_2)]}{p_1 R(T_2)+(1-p_1)R(T_1)} \tag{7}$$

여기서 $(1-p_1)[R(T_1)-R(T_2)]$는 T_1과 T_2사이에서 고장이 발생하였으나 서비스 센터에 들어오지 않은 경우의 확률을 나타내고, $p_1 R(T_2)+(1-p_1)R(T_1)$는 고장발생 여부에 관계없이 서비스센터에 들어오지 않는 모든 경우의 확률을 나타낸다. 따라서 식(7)은 T_1과 T_2사이에서 고장이 발생하였지만 서비스센터에 들어오지 않은 제품수의 기대값을 나타낸다.

④ 식(7)의 m을 이용하여 p_1의 새로운 기대값 p_1'을 구한다.

$$p_1' = \frac{n_2}{n_2+m} \tag{8}$$

⑤ p_1'을 p_1으로 하여 p_1 이 하나의 값으로 수렴할 때까지 ②~⑤ 까지의 과정을 반복한다.

예제 2) 대수정규분포의 가정(예제1) 에서 고장데이타를 얻은 후, 제품의 수명모수와 p_1 을 모를 때, 이들을 추정하는 방법을 EM 알고리즘에 적용하여 살펴보기로 한다.

① EM 알고리즘을 적용하기 위하여 p_1의 초기값을 0.7로 둔다.

② 0.7을 p_1의 참값으로 하여 T_2까지의 데이터를 사용하여 식(3)의 우도함수를 최대화하는 수명모수값을 구하면,

$\mu = 5.3905$, $\sigma = 0.192$, $p_1 = 0.7$

③ 식(7)을 이용하여 m을 구하면,

$m = 29.8414$

④ 식(8)을 이용하여 $p_1' = 0.1900405$

⑤ ④에서 구한 p_1'을 이용하여 p_1값이 하나의 값으로 근사할 때까지 ②~⑤의 과정을 반복한다.

예제(1)을 이용하여 계산하면 <표4>와 같이 p_1값은 0.189204 로 수렴함과 동시에 μ^*=5.1205 σ^*= 0.209 로 수렴함을 알 수 있다.

표 4 EM 알고리즘에 따른 수명모수와 p_1값들의 변화

p_1	μ^*	σ^*
0.700000	5.3905	0.196
0.190040	5.1210	0.208
0.189890	5.1210	0.209
0.189597	5.1215	0.208
0.189399	5.1215	0.208
0.189204	5.1205	0.209
0.189204	5.1205	0.209
0.189204	5.1205	0.209

<표4>에 의하면, 예제(1)의 경우에 추정된 p_1의 추정량 $p_1^* = 0.189204$, 이때의 $\mu^* = 5.1205$, $\sigma^* = 0.209$를 이용하여 t_p^*를 구하면

$t_p^* = 139.293$ 이다.

3.3.2 확률 질량함수를 이용한 p_1값 추정방법

분석시점이 $T_{1,2} = T_2 - T_1$로 만 주어진 경우 보증시점에서 분석시점 까지의 제품의 사용시간 t_i은 곧 $T_{1,2}$로서 상수가 된다. 따라서 보증시간에서 분석시간 동안 고장난 제품수 M은 $b(N, F(T_{1,2};\mu,\sigma))$를 따르게 된다. 만약 $M=m$일 때 p_1의 비율로 고장 데이터가 수집 된다고 하면 $Y \mid M=m$은 $b(m, p_1)$ 를 따르게 되므로 Y의 확률질량함수는

$$P(Y=y) = \sum_{m=y}^{N} \binom{N}{m} [F(T_{1,2};\mu,\sigma)]^m [1-F(T_{1,2};\mu,\sigma)]^{N-m} \times \binom{m}{y} p_1^y (1-p_1)^{m-y} \quad (9)$$

$$P(Y=y) = \binom{N}{y} [p_1 F(T_{1,2};\mu,\sigma)]^y [1-p_1 F(T_{1,2};\mu,\sigma)]^{N-y}$$

$$, (y=0,1,\ldots,N) \quad (10)$$

으로서 $b(N, p_1 F(T_{1,2};\mu,\sigma))$를 따르게 된다. 이 결과를 이용하여 $T_1 \sim T_2$사이에서 y의 고장 데이터가 수집 되었을 때 , 실제 고장난 제품의 갯수 M의 확률질량 함수는

$$P(M=m \mid Y=y) = \frac{\Pr(M=m, Y=y)}{\Pr(Y=y)} = \frac{\Pr(Y=y \mid M=m)\Pr(M=m)}{\Pr(Y=y)}$$

을 이용하여

$$P(M=m \mid Y=y) = \binom{N-y}{m-y}\left[\frac{(1-p_1)F(T_{1,2};\mu,\sigma)}{1-p_1F(T_{1,2};\mu,\sigma)}\right]^{m-y}$$

$$\times\left[\frac{1-F(T_{1,2};\mu,\sigma)}{1-p_1F(T_{1,2};\mu,\sigma)}\right]^{N-m} \quad (11)$$

$$,\ (m=y, y+1, \ldots, N)$$

가 된다.

식(11)를 이용하여 확률질량함수에 대한 기대치를 구하면 식(11)는 이항분포를 따르므로 기대치는

$$E(M=m \mid Y=y) = (N-y)\left[\frac{(1-p_1)F(T_{1,2};\mu,\sigma)}{1-p_1F(T_{1,2};\mu,\sigma)}\right] \quad (12)$$

이것은 T_1 와 T_2사이에서 y개의 고장 데이터가 수집 되었을 때, 실제 고장난 제품의 갯수 M의 확률질량 함수에 대한 기대치 이다.

이 기대치를 이용하여

$$p_1' = \frac{n_2}{E(M=m \mid Y=y)} \quad (13)$$

를 구하고 이 p_1'를 식 (12)에 다시 대입하여 기대치를 구하는 과정을 반복적으로 수행하여 수렴하는 p_1^*을 구한다.

예제(1)의 경우에 p_1의 초기치를 0.7로 가정하여 확률질량함수를 이용한 p_1^*값을 추정하면 표(5)과 같은 과정으로 수렴하게 된다 .

표(5)와 같이 $p_1^*=0.188521$, 로 수렴함과 동시에 $\mu^*=5.1115$, $\sigma^*=0.219$ 로 수렴함을 알 수 있다.

표 5 확률질량함수를 이용한 수명모수 p_1와 값들의 변화

p_1	μ^*	σ^*
0.700000	5.3910	0.196
0.149787	5.0990	0.200
0.194590	5.1145	0.221
0.187690	5.1110	0.219
0.188608	5.1125	0.218
0.188488	5.1115	0.219
0.188520	5.1115	0.219
0.188521	5.1115	0.219
0.188521	5.1115	0.219

예제1)의 경우에 추정된 p_1의 추정량 $p_1^*=0.188521$, 이때의 $\mu^*=5.1115$, $\sigma^*=0.219$를 이용하여 t_p^*를 구하면

$t_p^*=136.6858$이다.

EM 알고리즘과 확률질량함수의 경우에 p_1값과 을 비교해보면 다음과 같다.

표6. EM 알고리즘과 확률질량함수를 비교한 표

	p_1^*	μ^*	σ^*	t_p^*
EM 알고리즘	0.189204	5.1204	0.209	139.2390
확률질량함수	0.188521	5.1115	0.219	136.6858

4. 결 론

본 논문에서는 보증기간 동안의 사용현장 데이터와 보증시점이후 분석시점 까지의 사용현장 데이터를 추가하여, 대수 정규 분포의 경우를 가정하고 그에 따른 제품의 수명모수 및 백분위수를 구한다. 보증기간내의 제품은 모두 서비스 센타에 들어오지만 보증기간과 분석기간 사이의 데이터는 p_1의 확률로 서비스센타에 들어 온다고 가정하였으며, 사용현장 데이터와 보증기간 이후의 추가적으로 얻어진 자료를 이용하여 대수정규분포의 경우에 우도함수를 구하고, 그에따른 최우추정량을 구하였다. 이 결과를 제품의 수명모수가 대수정규분포를 따르는 경우에 적용하고 보증시점과 분석시점의 데이터가 서비스센타에 들어올 확률(p_1)의 변화에 따른 최우추정량의 효과를 편의와 평균오차제곱의 관점에서 분석하였으며, 보증시점과 분석시점의 데이터가 서비스센타에 들어올 확률(p_1)을 모른다는 경우를 가정하여 EM 알고리즘 및 확률질량함수에 의한 수명모수 및 보증시점과 분석시점의 데이터가 서비스센타에 들어올 확률(p_1) 값을 추정하였다.

최근 들어, 대부분의 회사가 고객만족 및 품질경영 을 내세우며 서비스센터의 규모를 확충하고 보증기간 이후에 고장이 발생한 제품에 대해서도 무상으로 제품이나 부품을 수리또는 교체하고 있어서, 보증기간 이후에 사용현장데이터도 쉽게 얻을 수 있게 되었다. 따라서 보증기간이후의 사용환경자료를 이용하면 분석의 어려움 등의 이유로 사장되어지거나 보증기간내의 사용환경자료로 간주되어 분석되는 문제점을 해결할 수 있을 것이며, 제품의 수명 모수 추정에 있어서 도움을 줄 수 있을 것이다. 추후 연구과제로는 사용현장자료를 추가한 경우에 지수 ,와이블, 대수정규가 아닌 또 다른 분포로의 확장도 생각해 볼 수가 있으며, A/S 자료수집 비용 등을 고려하여 총제품수에 대한 최적확률(p_1^*)의 값을 구하는 경우에도 확장을 할 수가 있겠다.

참고문헌

[1] Alexander,M.M and Franklin,A.G.(1974), *"Introduction to the theory of statistics"*, C. A., McGraw-hill.

[2] Drenick, R.F.(1960), *"The failure of Complex Equipment"*, Soc. indust. Appl. Math. ,Vol, 8, 680-689.

[3] Hahn, G.J. and Meeker, W.Q.(1982a), *"Pitfalls and Practical Considerations in Product Life Analysis, Part I ; Basic Concepts and Dangers of Extraction"*, Journal of Quality Technology,14,144-152.

[4] E.L.Lehmann, (1983), *"Theory of point Estimation"* . John Wiley.

[5] Hahn, G.J. and Meeker, W.Q.(1982b), *"Pitfalls and Practical Considerations in Product Life Analysis, Part II ; Mixtures of Product Populations and More Genral Models"*, Journal of Quality Technology,14,177-185.

[6] Kalbfleish, J.D. and Prentice,R.L.(1980), *"The statistical Analysis of Failure Time Data"*,

New York ; John Wiley.

[7] Kalbfleish.J.D and Laweless.J.F.(1988), *"Estimation of Relibility in Field-Preformance Studies"* ,Technonometrics, 30, 365-368.

[8] Lawless,J.(1983), *"Statistical Methods in Reliability",* Technometrics,25,305-335

[9] Lawless, J.(1982), *"Statistical Models and Method for Lifetime Data",* Wiley, New York

[10] Little,R.J.A and Rubin,D.B(1986), *"Statistical Analysis with Missing Data"*, New York,John Wiley & Sons.

[11] Miyamura, T.(1982),*"Estimating Component Failure Rates From Combined Component and System DATA : Exponentionally Distributed Component Lifetimes",* Technometrics, 25, 313-318

[12] Nelson, Wane, and Hahn ,G.J.(1972), *"Linear Estimation of a Regression Relationship from Censored DATA-PART1.Simple Method and Their Application",* Technometrics,15, 133-150

[13] Nelson, Wane.(1982), *"Applied Life Data analysis"*, New York;John wiley

[14] Nelson, Wane.(1990), *"Acclerated Testing - Statistical Testing, Test Plans, and Data analyses"*, New York ; John Wiley

[15] Serfling,R.J.(1980), *"Approximation Theorems of Mathmatical Statistics"*, New York, John Wiley & Sons.

[16] Suzuki, K (1985a), *"Estimation of Lifetime Distributions From Incomplet Field Data,"* Technometrics, 27, 263-272.

[17] Suzuki, K (1985b), *"Nonparametric Estimation of Lifetime Distributions From a Record of Failure and Follow-Ups"*, Journal of the American Statical Association,80,68-72.

[18] 오영석(1997), *"보증기간후의 사용현장 자료를 추가한 신뢰성 분석,"* 한국과학기술원 산업공학과.

[20] 황용근(1992), *"다수고장원인이 있는 제품에 대한 사용현장 데이터의 통계적 분석"*, 한국과학기술원 산업공학과.