시간과 사용량의 속성을 지닌 데이터의 분석방안

조진남*† · 백재욱**

* 동덕여자대학교 정보과학대학 데이터정보전공
** 한국방송통신대학교 정보통계학과

Analysis Approaches to Data of Both Age and Usage Attributes

Jinnam Jo*† · Jaiwook Baik**

* Department of Data & Information Science, Dongduk Women's University
** Department of Information Statistics, Korea National Open University

Key Words : Two-dimensional data; accelerated life test model; maximum likelihood estimator

Abstract

For many products failures depend on age and usage and, in this case, failures are random points in a two-dimensional plane with the two axes representing age and usage. Models play an important role in decision-making. In this research, an accelerate failure test (AFT) model is proposed to deal with the two-dimensional data. The parameters are proposed to be estimated through maximum likelihood estimators.

1. 서 론

품질보증 데이터는 현장에서 아이템의 신뢰성을 평가하는데 유용한 정보를 제공해주며, 추후의 품질보증 비용을 평가하고 차세대 아이템의 품질보증비용이나 고객의 불만을 줄여 신뢰성을 높이는데 사용된다. 하지만 품질특성 중 한 가지만 보증(예를 들어 자동차 출하 후 5년 품질보증)하는 1차원 품질보증 데이터의 경우와 달리 품질특성 중 두 가지를 보증(예를 들어 자동차 출하 후 5년 또는 10만km 품질보증)하는 2차원 데이터의 경우 아이템의 신뢰성을 평가하는 데에는 여러 가지 어려운 점들이 대두된다(Baik et al., 2004). 따라서 본 연구에서는 자동차 부품과 같은 아이템의 클레임 데이터에 대해 부품의 신뢰성을 평가하기 위해 어떻게 모형화 하는지 알아본다.

부품의 신뢰성은 사용률(예를 들어 1년당 주행 마일리지)에 따라 다르며, 이는 고객에 따라 다르다. 따라서 부품의 신뢰성을 올바로 평가하기 위해서는 사람들간의 차이를 고려하기 위해 확률적인(stochastic) 모형을 상정하고 사용률이 높으면 아이템의 신뢰성이 떨어지는 가속수명시험 모형을 고안하여 이용할 수 있을 것이다.

그러므로 본 연구에서는 부품의 신뢰성에 미치는 고장율의 영향을 고려하는 모형을 고안하고 평가하면서, 실제 2차원 품질보증 데이터의 분석에서 나오게 되는 여러 가지 애매한 문제들을 어떻게 풀어갈 수 있는지 살펴보고자 한다. 본 연구는 다음과 같은 순서로 짜여져 있다. 우선 제 2장에서는 본 연구에 대한 기존의 연구들을 섭렵하고, 제 3장에서는 새로운 모형화 과정에 대해 기술하며, 제 4장에서는 현장에서 실제로 나오는 2차원 품질보증 데이터가 어떤 형태를 띠는지 살펴보면서 분석의 기틀을 마련한다. 제 5장에서는 제 4장에서 나온 2차원 품질보증 데이터에 대해 모형화를 어떻게 하고 모수를 어떻게 추정할 것인지 살펴본다. 마지막으로 제 6장에서는 본 연구에 대해 잠정적인 결론을 내리고, 추후 연구과제가 무엇인

† 교신저자 jinnam@dongduk.ac.kr
※ 이 논문은 2004년도 동덕여자대학교 학술연구비 지원에 의하여 수행된 것임.

지 살펴본다.

2. 과거문헌 검토

모형은 의사결정에 있어서 결정적인 역할을 한다. 고장, 품질보증 비용 및 신뢰성향상 등에 사용되는 모형은 지금까지 여러 가지가 제안되었고, 이들은 Lawless (1982), Blischke and Murthy(1994 ; 2000) 및 Meeker and Escobar(1998)에 설명되어 있다. 하지만 이들 모형은 대부분 1차원 고장모형에 속하며, 2차원 고장모형은 비교적 적은 편이다.

2차원 고장모형은 크게 두 가지가 있다. 그 중 첫 번째 방법은 시간(age)과 사용량(usage)간의 관계를 이용하여, 2차원의 문제를 1차원으로 효율적으로 축소하는 것이다. 지금까지 연구된 대부분의 모형에서 시간(age)과 사용량(usage)간에는 선형관계(linear relationship)를 가정했다. 예를 들어 Blischke and Murthy (1994)와 Gertsbakh and Kordonsky(1998)를 보기 바란다. Iskandar and Blischke(2003)는 이 방법에 기초하여 오토바이 품질보증 데이터에 대해 분석을 했고, Lawless et al.(1995)와 Yang and Zaghati(2002)는 자동차 품질보증 데이터에 대해 분석을 했다. 또 다른 연구로 Rai and Singh(2005)는 사용률이 주어진 상태에서 생존함수에 대해 도출하였으며, Kalbfleisch and Lawless(1988), Lawless(1998) 및 Lawless et al.(1995)는 이런 상황에서 모수의 추정문제를 살펴보았다.

2차원 고장모형 중 두 번째는 시간(age)과 사용량(usage)을 각각 비음(non-negative) 확률변수로 보는 이변량분포(bivariate distribution)를 사용하는 것이다. 이에 대한 연구로는 Blischke and Murthy (1994 ; 2000)와 Murthy and Djamaludin(2003)를 보기 바란다. 여러 종류의 2변량 와이블분포 모형에 대해서는 Murthy et al.(2003)에 나와 있다.

3. 모형화

여기서 관심의 대상이 되는 아이템은 2차원 품질보증을 받는 자동차 부품이다. 예를 들어 자동차 타이어가 2년 4만km의 품질보증을 받는다면 2년내 또는 4만km이내 타이어에 문제가 생긴다면 무상교체를 해준다. 이 경우 품질보증영역은 품질보증시간 W(=2년) 및 품질보증마일리지 U(=4만km)로 2차원 그림에서 직사각형이 된다.

아이템은 시간(age)이 지나고 사용량(usage)이 많아지면서 그 성능이 열화되고, 궁극적으로 고장난다. 따라서 아이템이 고장 났을 때 시간(age)은 사용률(usage rate 예를 들어, distance traveled per year)의 함수로 나타낼 필요가 있다.

Y를 사용률이라고 하고, 이것의 분포함수를 $G(y\,;\theta_1)$라고 하자. 여기서 θ_1은 분포의 모수(들)를 나타낸다. 사용률 Y의 분포함수로는 일양분포, 와이블분포 등 여러 가지가 있을 수 있다.

사용률이 구체적으로 y인 경우 고장시간 T는 분포함수가 $F(t\,;\theta_2,\theta_3)$인 분포를 따른다고 하자. 여기서 θ_2는 척도모수이고 θ_3는 기타모수(예를 들어 형상, 위치모수 등)를 나타낸다. 보통 사용률은 아이템에 주어지는 스트레스로 생각할 수 있으므로 사용률이 각각 (y_1, y_2)로 다른 경우 두 수명 (T_1, T_2)간에는 가속수명시험 모형(Lawless, 1982 : Rai and Singh, 2005 참조)을 적용하면 다음과 같이 쓸 수 있다.

$$T_1 / T_2 = (y_1 / y_2)^{\gamma}$$

따라서 사용률이 어떻든 고장시간은 척도모수가 사용률의 함수로 표현되는 동일한 분포를 따르게 되며, 사용률이 (y_1, y_2)인 경우의 각 분포의 척도모수간에는 다음과 같은 관계가 성립한다.

$$\theta_2(y_2)/\theta_2(y_1) = (y_1/y_2)^{\gamma}.$$

여기서 $\theta_2^0 \equiv \theta_2(U/W)$로 두는 경우($\theta_2^0$는 사용률 $y=U/W$인 경우의 척도모수로, 예를 들어 5년 10만km를 보증하는 경우 y=10/5임) 일반적인 사용률 y의 경우 척도모수는 다음과 같이 표현할 수 있다.

$$(\frac{U/W}{y})^{\gamma}\,\theta_2^0.$$

모형설정은 다음 두 단계를 거쳐서 이루어진다.

두 분포함수 $G(y\,;\theta_1)$와 $F(t\,;\theta_2(y),\theta_3)$의 형태를 결정한다.

모수 $\Theta \equiv \{\theta_1, \theta_2^0, \gamma, \theta_3\}$의 값을 그림 또는 통계적인 방법으로 결정한다.

4. 2차원 품질보증 데이터

고객이 겪게 되는 고장의 수는 한개 이상일 수 있

다. 하지만 고장이 하나도 없을 수도 있다. 후자는 보증기간 또는 보증 시간(age)내 고장이 하나도 없어서 그럴 수 있고 또는 보증기간이 지나지 않았더라도 사용량(usage)이 보증사용량을 초과해서 그럴 수 있다.

이제 아이템의 신뢰성을 추정하기 위한 모형설정에 이용되는 품질보증 데이터는 다음과 같다고 하자.

n : 고객의 수

n_j : $j(1 \le j \le n)$번째 고객이 겪게 되는 고장수

$t_{i,j}$, $u_{i,j}$: $j(1 \le j \le n)$번째 고객이 겪게 되는 $j(1 \le i \le n_j)$번째 고장시 시간(age)과 사용량(usage)

이때 품질보증영역 (W, U) 내에서 고장이 없는 경우 사용률에 대한 정보가 전혀 없다. 하지만 품질보증영역 (W, U) 내에서 한 번 고장난 고객 j의 경우 사용률은 $(u_{j,1}/t_{j,1})$으로 표현할 수 있고, 두 번 고장 난 고객 j의 경우 두 번의 고장을 겪었으므로 사용률은 $(u_{j,2}-u_{j,1})/(t_{j,2}-t_{j,1})$과 $(u_{j,1}/t_{j,1})$로 나타낼 수 있다. 후자의 경우 편의상 처음부터 두 번째까지의 고장시간과 사용량의 비율인$(u_{j,2}/t_{j,2})$로 사용률을 나타낼 수 있다.

4.1 중도중단 시간(censored age)

아이템의 제조업자는 아직 작동중인 아이템은 클레임이 들어오지 않으므로 품질보증 만료시 시간(age)과 사용량(usage)을 알 수가 없다. 이런 데이터는 중도중단(censored) 데이터로 볼 수 있다. 하지만 품질보증영역 (W, U)내에서 한 번이라도 고장난 경우 고장 데이터에 기반하여 사용률을 추정할 수 있으며, 이 정보를 이용하여 품질보증이 끝날 때 작동중인 아이템의 시간(age)과 사용량(usage)을 결정할 수 있다. 품질보증 영역 중 고장이 1번밖에 일어나지 않는 고객 j의 경우 사용률은 (u_{j1}/t_{j1})이며, 고장이 두 번 일어난 고객 j의 사용률은 평균사용률인 $(u_{j,2}/t_{j,2})$로 본다. 따라서 고장이 일어난 고객 j의 경우 중도중단 시간(censored age)은 $\min(W-t_{n_j j},\ (U-u_{n_j j})/y_{n_j j})$로 볼 수 있다. 여기서 $y_{n_j j}=u_{n_j j}/t_{n_j j}$이다.

4.2 잠정적인 데이터 분석

n_A를 품질보증 영역내 고장이 한번 이상 일어난 고객의 수라고 하자. 이들 각각의 고객에 대하여 앞에서 기술한 바와 같이 사용률을 구할 수 있다. 이때 사용률은 0부터 무한대까지의 값을 취할 것이며, 이들 값을 K개의 서로 다른 구간(또는 그룹)으로 나눌 수 있다. 이제 m_k를 $k(k=1,\ 2,\ \cdots,\ K)$번째 구간에 속한 사용률의 고객수라고 하고, $n_{j,k}$를 k구간에 속한 고객 j $(j=1,\ 2,\ \cdots,\ n_k)$가 겪게 되는 고장수라고 하자. 그러면 k구간에 속한 고객 j가 품질보증이 끝날 때의 중도중단 시간(censored age)은 다음과 같다.

$$\min(W-t_{n_{jk}j,k},(U-u_{n_{jk}j,k})/y_{n_{jk}j,k}).$$

여기서 $y_{n_{jk}j,k}=u_{n_{jk}j,k}/t_{n_{jk}j,k}$.이다.

이제 $n_B(=n-n_A)$를 사각형의 품질보증 영역내에서 고장이 한번도 일어나지 않은 고객수라고 하자. 그러면 이들의 경우 고장이 한번도 일어나지 않았기 때문에 품질보증이 끝날 때 아이템의 중도중단 시간(censored age)을 추정할 수가 없다. 사실 사용률이 U/W보다 작은 고객의 경우에는 중도중단 시간(censored age)은 W일 것이며, 그렇지 않은 고객의 경우 W보다 작을 것이다.

5. 모형선택 및 모수 추정

5.1 모형선택 과정

이제 일반적인 2차원 품질보증 데이터가 주어진 경우 앞 4절의 '잠정적 데이터분석'에서 전개한 내용을 바탕으로 모형선택 및 모수추정은 다음과 같은 과정을 밟아 할 수 있다. 여기에서 데이터는 사각형의 품질보증 영역내에서 고장이 한번이라도 일어난 고객으로부터 나온 것을 이용한다.

1단계 : $G(y;\theta_1)$의 형태 결정 $y_{i,j}$를 이용하여 사용률에 대한 히스토그램을 그린다. 히스토그램의 모양은 $G(y;\theta_1)$의 분포함수를 적절히 결정하는 근거를 제공한다. 모형의 모수 추정치를 구하는데 히스토그램이나 도수다각형 같은 그림을 이용할 수 있고 또는 최대우도추정치(maximum likelihood estimation)와 같은 분석적 방법을 이용할 수도 있다. 2차원 품질보증을 실시하는 경우 현실에 맞는 최대우도추정치를 구하는 방법은 다음 절에서 설명한다.

2단계 : 사용률에 근거하여 고객을 $k(k=1,\ 2,\ \cdots,\ K)$개의 상이한 그룹으로 나눈다. 각 그룹에 속하

는 고객들에 대해서 사용률은 모두 같다고 본다(이를 $\bar{y}_k$라고 표시하자). 따라서 이들 고객의 고장시 시간(age)은 $F(t\,;\theta_{2k},\,\theta_3)$인 분포를 따른다고 본다. 여기서 θ_{2k}는 척도모수를 나타낸다. 이는 곧 k번째 그룹에 속한 사람들이 겪게 되는 아이템 고장시간은 분포함수가 $F(t\,;\theta_{2k},\,\theta_3)$인 iid(independently and identically distributed) 확률변수를 따른다는 것을 나타낸다. n_A개의 데이터에 대해 고장시간 $t_{i,j,k}$는 알려져 있고, 품질보증 만료시 아직 작동중인 $n_{j,k}$ 아이템의 중도중단 시간(censored age)은 앞의 4장에서 논의한대로 $\bar{y}_k$를 이용하여 구한다. 그러면 K개의 그룹 각각에 대해 경험적 분포함수(empirical distribution function)나 와이블 plot같은 그림을 그려 $F(t\,;\theta_{2k},\,\theta_3)$를 적절히 결정할 수 있을 것이다. 앞의 1단계에서와 마찬가지로 모수추정을 위해 그림 또는 분석적인 방법을 사용할 수 있다.

3단계 : 앞의 2단계에서 구한 $\hat{\theta}_{2k}(k=1,\,2,\,\cdots,\,K)$를 $\bar{y}_k$에 대비하여 2차원상에 그리고, 각 점을 잇는 부드러운 곡선을 그린다. 이 그림으로부터 θ_2^0를 추정한다{$\theta_2^0=\theta(U/W)$임}. 이 곡선으로부터 γ에 대한 가장 적절한 값을 추정할 수 있을 것이다.

5.2 최대우도법에 의한 모수의 추정

데이터가 어느 정도 모아진다면 최대우도(maximum likelihood)법은 여러 가지 통계적인 좋은 성질{예를 들어 불편성(unbiasedness)이나 최소분산(minimum variance) 등}들로 인해 모수추정에 바람직한 방법이 된다.

$\Psi\equiv\theta_2^0,\,\gamma,\,\theta_3$라고 하고 우도함수 $L(\Psi)$가 다음 두 가지 경우에 어떻게 전개되는지 살펴본다.

첫 번째 경우 : $G(y\,;\theta_1)$의 형태는 알려져 있지 않고, 따라서 추정해야 할 모수는 Ψ로 주어진 경우이다. 이때에는 고장을 한번이라도 겪은 n_A고객의 경우

고장데이터로부터 $t_{i,j}-t_{i-1,j}$ (고장시간임)
중도중단 데이터로부터 $\min(W-t_{n_j,j},\,(U-u_{n_j,j})/y_{n_j,j})$ (중도중단 시간임)

를 얻게 된다. 이제 $S(t\,;\theta_2(y),\,\theta_3)=1-F(t\,;\theta_2(y),\,\theta_3)$라고 하면 n_A고객으로 나오는 데이터의 우도함수는

$$L_A(\Psi)=\prod_{j=1}^{n}\prod_{i=1}^{n_j}f(t_{i,j}-t_{i-1,j}\,;\theta_3)$$
$$S(\min((U-u_{n_j,j})/y_{n_j,j},W-t_{n_j,j})\,;\,\theta_2(y_{n_j,j}),\,\theta_3)$$

과 같다. 여기서 $t_{0,j}=0,\ y_{i,j}=u_{i,j}/t_{i,j}$.이다.

다음으로 고장을 한번도 겪지 않은 n_B고객을 각 그룹의 크기에 따라 비례적으로 할당할 수 있을 것이다. 그러면 n_B고객 중 k번째 그룹에 속하는 고객 $n_{Bk}=n_B{}^*(m_k/n_A)$가 되며, 따라서 $n_{B1}+n_{B2}+\ldots+n_{BK}=n_B$가 된다. 각 그룹의 경우 중도중단 시간(censored age)은 해당 그룹의 사용률을 이용하여 구할 수 있으며(여기서 사용률은 n_A고객의 데이터로부터 나온 것임), 구체적으로 $\min(W,\,U/\bar{y}_k)$라고 할 수 있다. 여기서 $\bar{y}_k$는 k번째 그룹의 평균사용률을 나타낸다. 그러면 n_B의 고객들로부터 나온 데이터에 대한 우도함수는 다음과 같다고 할 수 있다.

$$L_B(\Psi)=\prod_{k=1}^{k}\prod_{j=1}^{n_{Bk}}S(\min(U/\bar{y}_k,\,W)\,;\theta_2(\bar{y}_k),\,\theta_3).$$

따라서 전반적인 우도함수는 $L_{A\cdot B}(\Psi)=L_A(\Psi)L_B(\Psi)$와 같다.

두 번째 경우 : $G(y\,;\theta_1)$의 형태는 알려져 있고, 추정해야 할 모수는 $\Theta\equiv\{\theta_1,\,\theta_2,\,\gamma,\,\theta_3\}$로 주어진 경우이다. 이때 n_A고객들로부터 나오는 데이터에 대한 우도함수는 다음과 같다.

$$L_A(\Theta)=\prod_{j=1}^{n}\left\{\prod_{i=1}^{n_j}\int_0^\infty f(t_{i,j}-t_{i-1,j}\,;\,\theta_2(y),\,\theta_3)dG(y)\right\}$$
$$S(\min((U-u_{n_j,j})/y,\,W-t_{n_j,j})\,;\,\theta_2(y),\,\theta_3)dG(y\,;\theta_1).$$

이때 n_B 고객들로부터 나오는 데이터에 대한 우도함수는 다음과 같다.

$$L_B(\Theta)=[\int_0^\infty S(\min(U/y,W)\,;\,\theta_2(y),\,\theta_3)dG(y)]^{n_B}.$$

따라서 전반적인 우도함수는 $L_{A\cdot B}(\Theta)=L_A(\Theta)L_B(\Theta)$와 같다.

각각의 경우 모수는 앞의 우도함수(또는 로그우도함수)를 극대화하는 값으로 결정되며, 이때 수리적인 방법이 종종 사용된다.

6. 잠정적인 결론 및 추후 연구과제

본 연구에서는 2차원 품질보증하에서 클레임이 제기된 아이템의 신뢰성평가에 대해 살펴보았다. 여기에서는 비수리계 부품이 클레임을 받은 경우 얻어지는 데이터를 모형화하는 새로운 방법을 제안한다. 사용률이 크면 클수록 아이템의 수명이 짧아지므로 사용률을 스트레스로 보았으며, 따라서 가속수명시험 모형을 이용한다. 모형에서 모수의 추정을 위해서는 중도중단 데이터(censored data)에 대한 적절한 가정이 필요했으며, 이런 가정하에 최대우도법으로 모수를 추정할 수 있다. 현재 이 연구와 관련된 주제들로는 다음과 같은 것이 있다.

(i) 본 연구에서 제시된 모형과 모수추정방법을 현장 데이터에 적용시키는 사례분석
(ii) 본 연구에서는 가속수명시험 모형 중 누승(power law) 모형을 적용했다. 하지만 경우에 따라서는 지수모형이 더 적합할 수 있다. 따라서 다른 모형을 적용하는 경우 구체적으로 모형화는 어떻게 하며, 모수추정은 어떻게 하는지 살펴볼 필요가 있다.
(iii) 본 연구에서 살펴본 아이템은 비수리계 아이템 즉, 부품이다. 하지만 이런 부품들이 모여서 시스템을 형성하며, 현실적으로 자동차와 같은 아이템은 시스템 수준에서 클레임이 들어오게 된다. 시스템에 클레임이 들어올 때마다 최소수리(minimal repair)가 이루어진다고 가정할 수 있는 경우(자동차의 경우 각 부품에 대한 수리 또는 교체가 이루어지면 그 전후 자동차의 상태가 똑같다고 볼 수 있으므로 최소수리가 이루어진다고 볼 수 있음) 시스템은 비동질적 포아송과정{non-homogeneous Poisson process(NHPP)}을 따른다고 볼 수 있다{NHPP의 강도함수(intensity function)에서 척도모수가 사용률에 의해 영향을 받는다고 가정할 수 있음}.

이런 연구주제들은 본 연구에서 제시된 연구주제와 관련된 것으로 2차원 품질보증이 이루어지는 아이템의 신뢰성평가 및 신뢰성향상에 필요하다고 판단된다.

참 고 문 헌

[1] Baik, J., Murthy, D. N. P., and Jack, N.(2004), "Two-dimensional failure modelling with minimal repair", *Naval Research Logistics*, Vol. 51, pp. 345-362.
[2] Blischke, W. R. and Murthy, D. N. P.(1994), *Warranty cost analysis*, Marcel Dekker, Inc., New York.
[3] Blischke, W. R. and Murthy, D. N. P.(2000), *Reliability*, Wiley, New York.
[4] Gertsbakh, I. B. and Kordonsky, H. B.(1998), "Parallel time scales and two-dimensional manufacturer and individual customer warranties", *IEE Transactions*, Vol. 30, pp. 1181-1189.
[5] Iskandar, B. P. and Blischke, W. R.(2003), "Reliability and warranty analysis of a motorcycle based on claims data", in *Case Studies in Reliability and Maintenance* (W.R. Blischke and D.N.P. Murthy eds.), Wiley, pp. 623-656.
[6] Kalbfleisch, J. D. and Lawless, J. F.(1988), "Estimation of reliability in field-performance Studies", *Technometrics*, Vol. 30, pp. 365-388.
[7] Lawless, J. F.(1982), *Statistical models and methods for lifetime data*, John Wiley & Sons, Inc., New York.
[8] Lawless, J. F.(1998), "Statistical analysis of product warranty data", *International Statistical Review*, Vol. 66, pp. 41-60.
[9] Lawless, J. F., Hu, J., and Cao, J.(1995), "Methods for the estimation of failure distributions and rates from automobile warranty data", *Lifetime Data Analysis*, Vol. 1, pp. 227-240.
[10] Meeker, W. Q. and Escobar, L. A.(1998), *Statistical methods for reliability data*, John Wiley & Sons, Inc., New York.
[11] Murthy, D. N. P. and Djamaludin, I.(2003), "Product warranty A review", *International Journal of Production Economics*, Vol. 79, pp. 231-260.
[12] Murthy, D. N. P., Xie, M., and Jiang, R. (2003), *Weibull models*, Wiley & Sons, New York.

[13] Rai, B. and Singh, N.(2005), "A modeling framework for assessing the impact of new time/mileage warranty limits on the number and cost of automotive warranty claims", *Reliability Engineering and System Safety*, Vol. 88, pp. 157-169.

[14] Yang, G and Zaghati, Z.(2002), "Two-dimensional reliability modeling from warranty data", *Annual Reliability and Maintainability Symposium*, pp. 272-278.