

현장데이터를 이용한 신뢰성 분석의 연구 동향

An Investigative Study on the Reliability Analysis using Field Data.

김 종 걸¹⁾, 박 창 규²⁾

Kim Jong-Gurl, Park Chang-Kyu
School of System Management Engineering

< Abstract >

제품의 사용환경에서의 고장데이터는 일반적으로 보증기간 동안 발생하는 클레임을 처리하거나 A/S를 하는 과정에서 얻어지게 되는데 이러한 현장데이터를 적절한 방법으로 분석하여 제품의 신뢰도에 대한 올바른 정보를 얻고 품질개선에 활용하는 것은 기업에 있어서 매우 중요하다.

본 연구에서는 사용현장 데이터의 수집방법 연구와 현장 데이터의 신뢰성 분석 방법의 연구 흐름을 살펴보고 이후 사용현장데이터의 신뢰성 분석 방법연구에 도움을 주는 것이 목적이다.

1. 서론

고객에게 제품의 품질을 보증하는데는 기능적인 측면보다는 제품의 모든 성질을 만족해야 한다. 가격 경쟁력에서 품질 경쟁력으로 이전해 나가는 양상도 이를 반영하고 있는 것이다. 예전의 품질특성이 흠, 결점수라면 지금의 품질특성은 제품의 수명과 관련된 즉, 제품의 평균고장시간, 보증기간 동안 고장날 제품의 비율, 품질보증비용 등이다.

제품의 품질특성치들을 구하기 위한 데이터들은 실험실에서 행해지는 수명시험(life testing), 환경 시험(environment testing) 등에 의해 얻어지는 실험실 데이터를 많이 사용한다. 가장 이상적인 고장데이터는 제품의 실제 사용현장에서 얻어져야 한다. 실험실에서의 수명시험은 시험환경이 실제환경과 다를 수밖에 없고, 얻어진 데이터들은 사용현장에서의 제품 수명에 대한 정보를 왜곡하여 나타낼 위험이 있다. 따라서 이러한 위험을 줄이고 제품의 수명에 대한 합리적인 정보를 얻어내기 위하여 사용현장 데이터를 수집하고 이를 분석하는 방법을 연구하는 것은 중요한 일이다.

본 연구에서는 사용현장데이터를 이용한 신뢰성 분석방법의 연구활동들에 대해 살펴본다. 2절에서 기존연구의 이론적 내용을 보여주고 3절에서는 결론으로 기존 연구내용의 흐름을 파악하여 적용분야의 확대 가능성을 알아보고자 한다.

2. 연구 내용

사용하는 기호는 다음과 같다.

N	총 시험 제품 수
θ	제품의 수명 모수(벡터)
T_i	제품 i 의 고장시간
x_i	제품 i 에 대한 설명변수(열벡터)

1) 성균관대학교 시스템 경영공학부
2) 성균관대학교 시스템 경영공학부

T^0	운영시간 기준으로 정해진 제품의 보증기간
D_1	보증기간 동안 고장난 제품의 집합
D_2	보증기간 동안 고장나지 않고 추적 조사된 제품의 집합
n_u	보증기간 동안 고장난 제품의 수
n_c	추적 조사된 제품의 수
n_l	$= N - n_u n_c$
p^*	추적 조사 비율
$f(t x;\theta), F(t x;\theta), \bar{F}(t x;\theta)$	설명변수 x 와 수명모수 θ 가 주어진 경우 수명시간의 확률밀도 함수, 분포함수, 신뢰도 함수

1) 설명변수가 있는 경우

설명 변수가 있을 때 사용된 우도함수를 총 시험제품수와 설명변수에 대한 정보의 양에 따라 세 가지 경우로 나누어 살펴보면 다음과 같다.

- i) 총 시험 제품수와 보증기간 동안 고장나지 않은 제품에 대한 설명변수를 모를 경우
보증기간 동안에 고장난 제품의 고장시간과 설명변수를 이용하면 우도 함수는

$$L_T(\theta) = \prod_{i: t_i \leq T^0} \frac{f(t_i | x_i; \theta)}{F(T^0 | x_i; \theta)} \quad (1)$$

이 되고, 이를 최대화하는 θ 값을 구함으로써 최우추정량 $\hat{\theta}$ 를 구할 수 있다.

- ii) 총 시험 제품수 N 과 모든 시험제품에 대해 설명변수 x_i 를 알고 있을 경우
이 경우의 우도함수는

$$L_F(\theta) = \prod_{i: t_i \leq T^0} f(t_i | x_i; \theta) \prod_{i: t_i > T^0} \bar{F}(T^0 | x_i; \theta) \quad (2)$$

과 같으므로 θ 에 대한 추정량을 구할 수 있다.

- iii) 총 시험 제품수는 알고 있고, 보증기간 동안 고장나지 않은 제품에 대해서는 일정비율 p^* 만큼 추적 조사하여 설명변수의 값을 알 수 있을 경우

Kalbfleisch와 Lawless(1988)이 제안한 의사우도함수

$$L_p(\theta) = \prod_{i \in D_1} f(t_i | x_i; \theta) \prod_{i \in D_2} [\bar{F}(T^0 | x_i; \theta)]^{1/p^*} \quad (3)$$

를 이용하여 θ 에 대한 추정량을 구한다.

- iv) 다수의 고장원인이 있는 경우

배 도선(1995a,b)는 다수의 고장원인이 있는 제품의 현장데이터에 대한 분석방법을 제안한 의사우도함수

$$L_p(\theta) = \prod_{i \in D_1} \prod_{m=1}^k [f_m(t_i | x_i; \theta) \prod_{l \neq m} \bar{F}_l(t_i | x_i; \theta)]^{[r_i = m]} \cdot \prod_{i \in D_2} \left[\prod_{m=1}^k \bar{F}_m(T^0 | x_i; \theta) \right]^{1/p^*} \quad (4)$$

를 이용하여 θ 에 대한 추정량을 구한다.

2) 설명변수가 없는 경우

고장시간과 보증기간의 시간단위가 동일하지 않을 경우에서 보증기간이 달력시간으로 정해져 있을 경우는 Suzuki(1995a,b)가 제안한 의사우도함수

$$L^*(\theta) = \prod_{i=1}^{n_i} f(\tilde{z}_i; \theta) \left[\prod_{j=1}^{n_c} \overline{F}(\overline{z}_j; \theta) \right]^{1 + \frac{n_i}{n_c}} \quad (5)$$

를 이용하여 θ 에 대한 추정량을 구할 수 있다. 단, $z_i = \min(T_i, Y_i)$ 이고, \overline{z}_i 는 $T_i > Y_i$ 이고 추적 조사된 z_i 의 관측값, \tilde{z}_i 는 $T_i \leq Y_i$ 일 때의 z_i 의 관측값을 나타낸다. 보증기간이 달력시간과 운영시간의 혼합형으로 정해져 있을 경우의 의사우도함수는

$$L^*(\theta) = \prod_{i=1}^{n_{110} + n_{100}} f(\tilde{z}_i; \theta) \prod_{i=1}^{n_{111} + n_{101}} [f(\tilde{z}_i; \theta)]^{1 + \frac{\widehat{N}_{1010}}{n_{111} + n_{101}}} \prod_{k=1}^{n_c} [\overline{F}(\overline{z}_k; \theta)]^{1 + \frac{\pi}{n_c}} \quad (6)$$

와 같이된다. 단, n_{111} 은 추적 조사 대상 제품 중 운영시간의 보증기간 이후 달력시간의 보증기간 이내에 고장나는 제품수, n_{110} 은 추적 조사 대상 제품 중 운영시간 및 달력시간의 보증기간 이내에 고장나는 제품수, n_{1011} 은 추적 조사 대상이 아닌 제품 중 운영시간의 보증기간 이후 달력시간의 보증기간 이전에 고장이 발생하여 서비스센터를 찾는 제품수, n_{1010} 은 추적 조사 대상이 아닌 제품 중 운영시간의 보증기간 이후 달력시간의 보증기간 이전에 고장이 발생하였으나 서비스센터를 찾지 않는 제품수, n_{100} 은 추적 조사 대상이 아닌 제품 중 운영시간과 달력시간의 보증기간 이내에 고장이 발생한 제품수를 나타내며,

$$\widehat{n}_i = \frac{N - (n_{111} + n_{110} + n_c)}{n_{111} + n_{110} + n_c} \cdot n_c,$$

$$\widehat{n}_{1010} = N - n_c - \widehat{n}_i - n_{111} - n_{110} - n_{1011} - n_{100}$$

이다.

3) 추적조사를 하는 경우

서비스센터에 들어오지 않은 제품에 관한 정보를 추정하는 다른 방법으로는 전화, 우편엽서 등을 이용하여 고장 시간이나 제조 특성 및 사용 환경 등의 고장 정보를 얻는 추적조사 방법이 있다. 추적조사를 할 경우 추가적인 정보 취득으로 인한 비용의 증가와 분석의 어려움은 따르지만, 데이터를 얻지 못한 제품으로부터 정보를 얻음으로써 보다 정확한 분석이 가능해진다. 이에 관한 연구들은 다음과 같다.

Suzuki(1985 a, b)는 제품의 보증기간은 달력시간이고 고장은 운영시간을 따를 때, 사용현장 데이터와 총 판매 제품 중 일정 비율을 랜덤샘플 하여 고장이 나지 않은 제품을 추적 조사하여 얻은 관측 중단 시간데이터를 이용하여 제품의 신뢰도를 추정하는 문제를 다루었다. Kalbfleisch 와 Lawless(1988)는 제품의 수명에 영향을 주는 제품의 제조특성, 환경 특성 등과 같은 설명변수가 존재할 때, 수명 분포의 모수와 설명변수가 대수선형 관계임을 가정하고 사용현장 데이터와 총 판매 제품 중 보증기간 내에 고장이 발생하지 않은 제품의 일정 비율을 추적조사 하여 얻은 설명변수에 대한 데이터를 이용하여 수명 분포의 모수를 추정하였다. 배도선 등(1995 a)과 배도선 등(1995 b)은 각각 Suzuki(1985 a), Kalbfleisch 와 Lawless(1988)의 연구를 다수 고장원인 모형과 다수고장원인이 있고 수리 가능한 모형으로 확장하였다.

4) 추적조사를 하지 않는 경우

추적조사를 하지 않는 경우 분석방법은 서비스센터를 들어오는 제품과 그렇지 않은 제품과의 수학적 모형을 가정한 후, 서비스 센터를 통하여 얻어진 정보를 이용하여 제품의 수명에 관련된 품질특성들을 구하는 것이다 이에 대한 기존 연구들은 다음과 같다.

Kalbfleisch등(1991)은 고장난 제품이 서비스센터에 들어오기까지 시간지연이 있는 경우에 사용현장 데이터를 이용하여 시간지연에 관한 정보를 추정한 후 보증기간동안의 기대고장횟수와 기대보증비용을 구했고, Lawless(1994)는 이들의 연구를 시간지연이 제품의 고장시간에 따라 변하는 변량모형인 경우로 확장하였다. Lawless등(1995)은 보증기간이 달력시간이고 제품의 고장은 운영시간을 따를 때 달력시간과 운영시간 사이에 선형관계가 존재한다고 가정한 후 달력시간의 사용현장데이터를 운영시간의

데이터로 변환하여 수명 분포의 모수를 추정하였다. 또한 Hu와 Lawless(1996)는 이들의 연구를 수리 가능한 제품에 대한 경우로 확장하여 비모수적 추정방법을 제안하였고 Lawless와 Thiagarajah(1996)는 수리 가능한 제품의 수명 분포의 모수가 운영시간 및 과거의 고장횟수와 선형적인 관계가 있는 경우를 다루었다.

< 표 1 > 사용 현장데이터를 이용한 신뢰성 분석에 관한 연구

기존 연구	수명 분포	설명변수 유·무	추적 조사 유·무	고장 원인	내 용
Hahn & Meeker(1982a)	일반적인 경우 (와이블, 대수정규)				고장데이터 분석
Miyamura(1982)	일반적인 경우				보증제도와 고장데이터를 이용한 부품에 대한 수명시간 추정
Nelson(1982)	일반적인 경우			다수 고장 원인	여러 가지 고장데이터의 체계적인 분석
Nelson(1990)	지수, 대수정규, 와이블			단일 및 다수 고장 원인	가속수명시험 하에서의 수명분포 모수 추정
Lawless(1983)	지수, 대수정규		○		사용환경에서의 데이터 수집방법
Kalbfleisch & Lawless(1988)	지수, 와이블	○	○	단일 고장 원인	설명변수 데이터를 이용한 수명분포의 모수 추정량과 점근분산
Suzuki(1985a,b)	지수, 와이블	×	○	단일 고장 원인	시험 제품 일부를 추적조사, 보증기간은 달력시간인 경우
Herman & Patell(1971)	지수, 와이블			다수 고장 원인	수명분포의 모수 추정
Miyakawa(1984)	일반적인 경우			다수 고장 원인	일부 고장원인이 알 수 없는 경우 수명시간 추정
Bai et al.(1995a,b)		○,×	○	다수 고장 원인	다수 고장원인 모형과 다수 고장인 경우 수리 가능한 경우의 신뢰성 분석
Kalbfleisch, J.D., Lawless, J.F. and Robinson, J.A.(1991)		○	×		시간지연 정보를 추정 후 기대고장횟수와 기대보증비용 추정
Lawless(1994)		○	×		Kalbfleisch et al.(1991)의 연구를 제품 고장 시간에 따라 변하는 변량모형으로 확장
Lawless, J.F., Hu, J. & Cao, J.(1995)		○	×		달력시간 사용현장데이터를 운영시간으로 변환하여 수명모수 추정
Hu, X. J. & Lawless, J.F.(1996)		○	×		수리 가능한 제품에 대해 비모수적 추정방법 제안
Lawless, J.F. & Thiagarajah, K.(1996)		○	×		수리 가능한 제품에 대해 수명분포 모수가 운영시간 및 과거의 고장횟수와 선형적인 관계 적용

3. 결론

사용현장데이터의 신뢰성 분석 문헌들은 대개 보증기간과 운영시간, 설명변수와 추적 조사, 분포의 모수 추정, 그리고 제품의 고장원인이 하나인 경우에서 다수 고장원인으로 확장 연구되어 왔다. 조사 문헌들에서 사용현장데이터의 형태와 수집방법에 대한 가정은 다음과 같다.

1) 설명변수가 없는 경우

- ① 제품은 k 개의 고장원인이 있으며, 이들 고장원인별 수명은 서로 독립이다.
- ② 모든 제품에 대하여, 제품의 수명 X 와 관측중단시간 Y 는 서로 독립이며 시간 척도는 운영시간(주행거리, 빈도 등)인 반면에 연구의 관측기간(제품의 보증기간)은 달력시간(달, 년 등)으로 측정된다.
- ③ 관측기간 동안 고장난 제품의 고장시간과 고장원인은 정확히 기록된다.
- ④ 추적 조사되는 제품의 비율 p^* 는 0이 아니며, $n_c \neq 0$, $n_c^m \neq 0$, $m=1, \dots, k$ 이다. 이는 일치 추정량을 얻기 위함이다.
- ⑤ 추적 조사되어지는 제품은 총 시험제품 N 중 랜덤하게 선택되며, 선택된 제품 중 보증기간 내에 고장나지 않은 제품의 관측중단시간은 정확하게 얻을 수 있다.

2) 설명변수가 있는 경우

- ① 제품은 k 개의 고장원인이 있으며, 이들 고장원인별 수명은 서로 독립이다.
- ② 제품의 고장시간과 보증시간의 시간 척도는 같고, 보증기간 동안 고장이 발생한 모든 제품의 고장

시간과 고장원인, 설명변수 값은 기록된다.

- ③ 추적 조사비율은 0이 아니고, 관측 중단된 제품 중 추적 조사되는 제품은 랜덤하게 선택되며, 선택된 제품의 설명변수에 대한 정보는 정확하게 얻을 수 있다.

위 가정에서 보증기간내의 데이터를 모두 얻지 못하고 일부만을 얻을 경우의 분석방법에 대한 연구는 권 혁무(1995)가 다루었고, 보증기간 후의 사용현장데이터를 추가한 경우의 신뢰성 분석방법에는 배 도선(1997)에 의해 다루어졌다.

사용현장데이터의 추가적인 분석방법에는 설명변수의 종류에 따른 효과분석, 추적 조사에서의 샘플링 방법, 고장모드별 신뢰성 분석방법, 적절한 모형의 선택방법, 의사우도함수의 추정량성질, 사용현장데이터의 수집방법, 그리고 실험실 데이터와 사용현장데이터를 통합한 Bayesian 신뢰성 분석 등이 계속해서 연구되어야 할 분야이다.

< 참고문헌 >

- [1] 권혁무(1995), "현장데이터를 이용한 수명모수의 추정", 제 5회 신뢰성·보전성 심포지엄, 59-65
- [1] 배도선, 최인수, 황용근(1995a), "고장 원인이 여럿인 제품의 사용현장데이터 분석," 응용통계연구, 제8권 1호, 89-104.
- [2] 배도선, 윤형제, 최인수(1995b), "수리가능한 제품의 사용현장데이터 분석," 응용통계연구, 제8권 2호, 133-145.
- [4] 배도선, 오영석(1997), "보증기간 후의 사용현장데이터를 추가한 신뢰성 분석", KAIST 석사논문.
- [3] Amster. S. J., Brush, G. G., and Saperstein. B.(1982), "Planning and Conducting Field-Tracking Studies." *Bell System Technical Journal*, 61, 2333-2364.
- [4] Godambe. V. P., and Thompson, M. E.(1986), "Parameters of Superpopulation and Survey Population: Their Relationships and Estimation." *International Statistical Review*, 54, 127-138.
- [5] Hahn, G. J. and Meeker, W. Q.(1982a), "Pitfalls and Practical Considerations in Product Life Analysis, Part I ; Basic Concepts and Dangers of Extrapolation," *Journal of Quality Technology*, 14, 144-152.
- [6] Hahn, G. J. and Meeker, W. Q.(1982a), "Pitfalls and Practical Considerations in Product Life Analysis, Part II; Mixtures of Product Populations and More General Models," *Journal of Quality Technology*, 14, 177-185.
- [7] Holt, D., and Scott, A. J.(1981), "Regression Analysis Using Survey Data," *The Statistician*, 30, 169-178.
- [8] Holt. D., Smith, T. M. F., and Winter, P. D.(1980), "Regression Analysis of Data From Complex Surveys," *Journal of the Royal Statistical Society, Sep. A*, 143, 474-487.
- [9] Hu, X. J. and Lawless, J. F.(1996), "Estimation of Rate and Mean Functions From Truncated Recurrent Event Data," *Journal of American Statistical Association*, 91, 300-310.
- [10] Jewell, N. P.(1985), "Least squares Regression With Data Arising From Stratified Samples of the Dependent Variable," *Biometrika*, 73, 11-21.
- [11] Kalbfleisch, J. D., and Lawless, J. F.(1988), "Likelihood Analysis of Multi-state Models for Disease Incidence and Mortality," *Statistics in Medicine*, 7, 149-169.
- [12] Kalbfleisch, J. D. and Lawless, J. F.(1988), "Estimation of Reliability in Field Performance Studies," *Technometrics*, 30, 365-388.
- [13] Kalbfleisch, J. D., Lawless, J. F. and Robinson, J. A.(1991), "Methods for the Analysis and Prediction of Warranty Claims," *Technometrics*, 33, 273-285.
- [14] Kalbfleisch, J. D. and Prentice, R. L.(1980), *The Statistical Analysis of Failure Time Data*, New York ; John Wiley.

- [15] Lawless, J. F.(1982), *Statistical Models and Methods for Lifetime Data*, New York ; *John Wiley*.
- [16] Lawless, J. F.(1983), "Statistical Methods in Reliability," *Technometrics*, 25, 305-335.
- [17] Lawless,J.F.(1994), "Adjustments for reporting delays and the prediction of occurred but not reported events," *The Canadian Journal of Statistics*, 22, 15-31.
- [18] 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*, 1, 227-240.
- [19] Lawless,J.F. and Thiagarajah,K.(1996),"A Point-Process Model Incorporating Renewals and Time Trends, With Application to Repairable System," *Technometrics*, 38, 131-138.
- [20] Miller, R. G.(1981), *Survival Analysis*, New York ; *John Wiley*.
- [21] Miyamura, T.(1982), "Estimating Component Failure Rates From Combined Component and Systems Data : Exponentially Distributed Component Lifetimes," *Technometrics*, 24, 313-318.
- [22] Nelson, W.(1982), *Applied Life Data Analysis*, New York ; *John Wiley*.
- [23] Nelson, W.(1990), *Accelerated Testing - Statistical Models, Test Plans, and Data Analyses*, New York ; *John Wiley*.
- [24] Scott, A. J., and Wild, C. J.(1986), "Fitting Logistic Models Under Case Control of Choice-Based Sampling," *Journal of the Royal Statistical Society*, Sep. B, 48, 170-182.
- [25] Suzuki,K.(1985a), "Estimation of Lifetime parameters From Incomplete Field Data," *Technometrics*, 27, 263-272.
- [26] Suzuki,K.(1985b), "Nonparametric Estimation of Lifetime Distributions From a Record of Failures and Follow - Ups," *Journal of the American Statistical Association*, 80, 68-72.