대한산업공학회/한국경영과학회 2002 춘계공동학술대회
한국과학기술원(KAIST) 2002년 5월 3일-4일

이차원 보증 사용현장데이터의 분석

정 민 · 배도선
한국과학기술원 산업공학과

Field Data Analyses of Two-Dimensional Warranty Data

Min Jung · Do Sun Bai
Dept. of Industrial Engineering, Korea Advanced Institute of Science and Technology

Abstract

This paper proposes a method of estimating lifetime distribution for products under two-dimensional warranty in which age and usage are used simultaneously to determine the eligibility of a warranty claim. For such a case, existing methods reduce the two-dimensional time scale to a single scale assuming that the two variables have a functional relationship. This assumption is, however, not appropriate since the functional relationship is unknown in practice. In this paper, the field age and usage data are modeled with a bivariate lifetime distribution. Method of obtaining maximum likelihood estimators is outlined, their asymptotic properties are studied and specific formulas for a bivariate Weibull distribution are obtained. The proposed model is compared with the existing one which assumes a linear relationship between the two variables. Simulation studies are performed to investigate the effect of the degree of dependency between the two variables.

1. Introduction

Life data analyses are commonly used to estimate the lifetime distribution of a product and to obtain information on the life characteristics such as reliability, failure rate, percentile and mean time to failure, etc. This information is then used in developing new products or improving the reliability of existing products, in designing burn-in and warranty programs, and in planning the supply of replacement parts. Life data can be obtained from life testing at laboratory, warranty claims or failure-records in the field use. It is widely recognized that laboratory life test results frequently say little about the reliability of equipment in actual (field) use. When good field data are collected, more can be done to relate equipment or system reliability to the environment in which it operates, as well as to inherent properties of the equipment because it captures actual usage profiles and the combined environmental exposures that are difficult to simulate in the laboratory and it is more likely to observe longer time-to-failures. Thus it is important to develop procedures for the collection and analyses of the field data.

Several studies on analyses of field data appeared in the literature. Suzuki(1985a, 1985b) proposed parametric and nonparametric methods of estimating lifetime distribution from field failure data with supplementary information about censoring times obtained from following up a portion of the products that survive warranty time. Kalbfleisch and

Lawless (1988) suggested procedures for the collection of field data and used a regression model to estimate lifetime distribution from field failure data with supplementary information about covariates. Kalbfleisch et al.(1991) proposed methods of analysis and prediction of warranty claims with reporting delays. Hu and Lawless(1996) developed an estimation procedure with supplementary information about covariates and censoring times and Hu et al.(1998) proposed nonparametric procedures to analyze field data. David and Kieron(1999) presented methods of testing the suitability of the exponential distribution for grouped field data. Oh and Bai(2001) proposed methods of estimating the lifetime distribution for situations where additional field data can be gathered after the warranty expires.

Most of works on field data analyses have considered one-dimensional warranty data. However, some products, such as automobiles, power generators and factory equipments, are sold under two-dimensional warranty in which age and usage are used simultaneously to determine the eligibility of a warranty claim. For such a case, reliability models including both age and usage should be considered. Existing works on two-dimensional warranty data analyses reduce the two-dimensional time scale to a single scale assuming that two variables have a functional relationship. This assumption is, however, not appropriate since the functional relationship is unknown in practice. In this paper, we propose a two-dimensional reliability model in which a bivariate lifetime distribution is used to analyze the field age and usage data. Specific formulas for a Bivariate Weibull distribution is obtained and simulation studies are performed to investigate the effect of the degree of dependency between the two variables.

2 Estimation of Lifetime Distribution

Notations

N total number of items

T age to failure

U usage to failure

T^0 warranty limit for age

U^0 warranty limit for usage

$\boldsymbol{\theta}$ column vector of parameters of lifetime distribution

$f(t,u;\boldsymbol{\theta})$ probability density function(pdf) of lifetime distribution

$F(t,u;\boldsymbol{\theta})$, $S(t,u;\boldsymbol{\theta})$ cumulative distribution function(cdf) and survival function(sf) of lifetime distribution

D set of items failed in warranty region $(0,T^0]\times(0,U^0]$

$\bar{D}_1$ set of items not failed whose age does not exceed T^0

$\bar{D}_2$ set of items not failed up to T^0

$\bar{D}$ set of items not failed $\left(\bar{D}=\bar{D}_1\cup\bar{D}_2\right)$

2.1 Assumptions

The following assumptions are made:

1. Items are sold under two-dimensional warranty.
2. Failure times (t_i,u_i) are observed exactly.
3. A failed item after warranty expires is not reported.
4. Exact age of an item is known from sales records.

2.2 Likelihood

A failed item within warranty region is reported to the original manufacturer. However, after the warranty expires, a failed item is not reported. The likelihood function can be constructed as follows;

i) each failed item within warranty region $(0,T^0]\times(0,U^0]$ contributes a term $f(t_i,u_i;\boldsymbol{\theta})$ to the likelihood,

ii) since age of an item is known, each item not failed whose age does not exceed T^0 contributes a term $1-F(t_i,U^0;\boldsymbol{\theta})$ to likelihood,

iii) each item not failed up to T^0 contributes a term $1-F(T^0,U^0;\boldsymbol{\theta})$ to likelihood.

The log-likelihood function then becomes

$$
\begin{aligned}
&\log L(\boldsymbol{\theta}) \\
&= \sum_{i\in D}\log f(t_i,u_i;\boldsymbol{\theta}) + \sum_{i\in \bar{D}_1}\log\left[1-F(t_i,U^0;\boldsymbol{\theta})\right] \\
&\quad + \sum_{i\in \bar{D}_2}\log\left[1-F(T^0,U^0;\boldsymbol{\theta})\right] \\
&= \sum_{i\in D}\log f(t_i,u_i;\boldsymbol{\theta}) + \sum_{i\in \bar{D}}\log\left[1-F(\min(t_i,T^0),U^0;\boldsymbol{\theta})\right]
\end{aligned}
\tag{1}
$$

The maximum likelihood estimator(MLE) $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be obtained by maximizing Eq.(1), and under regularity conditions of Cramér(1946), $\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ has a limiting multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}^{-1}(\boldsymbol{\theta})$, where

$$
\mathbf{I}(\boldsymbol{\theta}) = \lim_{N\to\infty}\frac{1}{N}E\left[-\frac{\partial^2 \log L}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right], \tag{2}
$$

is the Fisher information matrix and $\boldsymbol{\theta}'$ is the transpose of vector $\boldsymbol{\theta}$. $\mathbf{I}(\boldsymbol{\theta})$ is consistently estimated by the observed Fisher information matrix;

$$
\mathbf{I}_N(\hat{\boldsymbol{\theta}}) = -\frac{1}{N}\left[\frac{\partial^2 \log L}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} . \tag{3}
$$

3. Formulas for a Bivariate Weibull Model

Hougaard(1986) proposed a bivariate Weibull distribution which is useful to describe dependent failure-times with positive correlation. If the lifetime (t_i,u_i) follows the Hougaard bivariate Weibull distribution proposed with scale parameter θ_1 and θ_2, and shape parameter β_1 and β_2, and dependency parameter δ then its cdf and pdf are

$$
\begin{aligned}
F(t,u;\boldsymbol{\theta}) &= 1-S(t,u;\boldsymbol{\theta}) \\
&= 1-\exp\left\{-\left[\left(\frac{t}{\theta_1}\right)^{\frac{\beta_1}{\delta}}+\left(\frac{u}{\theta_2}\right)^{\frac{\beta_2}{\delta}}\right]^{\delta}\right\}
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
f(t,u;\boldsymbol{\theta}) &= \frac{\beta_1}{\theta_1}\left(\frac{t}{\theta_1}\right)^{\frac{\beta_1}{\delta}-1}\frac{\beta_2}{\theta_2}\left(\frac{u}{\theta_2}\right)^{\frac{\beta_2}{\delta}-1}\left\{\left(\frac{t}{\theta_1}\right)^{\frac{\beta_1}{\delta}}+\left(\frac{u}{\theta_2}\right)^{\frac{\beta_2}{\delta}}\right\}^{\delta-2} \\
&\times\left\{\left[\left(\frac{t}{\theta_1}\right)^{\frac{\beta_1}{\delta}}+\left(\frac{u}{\theta_2}\right)^{\frac{\beta_2}{\delta}}\right]^{\delta}+\frac{1}{\delta}-1\right\}\exp\left\{-\left[\left(\frac{t}{\theta_1}\right)^{\frac{\beta_1}{\delta}}+\left(\frac{u}{\theta_2}\right)^{\frac{\beta_2}{\delta}}\right]^{\delta}\right\}
\end{aligned}
\tag{5}
$$

respectively. Its marginal survival functions are

$$
S_T(t) = \exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1}\right], \tag{6}
$$

$$
S_U(u) = \exp\left[-\left(\frac{u}{\theta_2}\right)^{\beta_2}\right]. \tag{7}
$$

One can see that marginal distributions of T and U are univariate Weibull distributions. The dependency parameter δ has a value in $(0,1]$. If δ is close to 0, the two variables have strong positive correlation. If $\delta=1$, the two variables are independent. If $\beta_1=\beta_2=\beta$, the log-likelihood function is

$$
\begin{aligned}
&\log L(\boldsymbol{\theta}) = \\
&\sum_{i\in D}\left[2\log\beta-\frac{\beta}{\delta}(\log\theta_1+\log\theta_2)+\left(\frac{\beta}{\delta}-1\right)(\log t_i+\log u_i)\right] \\
&+\sum_{i\in D}\left[(\delta-2)\log\left\{\left(\frac{t}{\theta_1}\right)^{\frac{\beta_1}{\delta}}+\left(\frac{u}{\theta_2}\right)^{\frac{\beta_2}{\delta}}\right\}\right]+ \\
&\sum_{i\in D}\left[\log\left\{\left(\left(\frac{t}{\theta_1}\right)^{\frac{\beta_1}{\delta}}+\left(\frac{u}{\theta_2}\right)^{\frac{\beta_2}{\delta}}\right)^{\delta}+\frac{1}{\delta}-1\right\}+\left(\frac{t}{\theta_1}\right)^{\frac{\beta_1}{\delta}}+\left(\frac{u}{\theta_2}\right)^{\frac{\beta_2}{\delta}}\right]^{\delta} \\
&+\sum_{i\in\bar{D}}P,
\end{aligned}
\tag{8}
$$

where

$$
\begin{aligned}
P = \log\Bigg[&\exp\left\{-\left(\frac{\min(t_i,T^0)}{\theta_1}\right)^{\beta}\right\}+\exp\left\{-\left(\frac{U^0}{\theta_2}\right)^{\beta}\right\} \\
&-\exp\left\{-\left(\left(\frac{\min(t_i,T^0)}{\theta_1}\right)^{\frac{\beta}{\delta}}+\left(\frac{U^0}{\theta_2}\right)^{\frac{\beta}{\delta}}\right)^{\delta}\right\}\Bigg]
\end{aligned}
\tag{9}
$$

<Table 1. Performance of the MLEs>

		θ_1		θ_2		β		δ	
N	δ	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
100	0.2	0.0137	0.0382	0.0246	0.1501	0.0201	0.1992	-0.0061	0.0030
	0.4	0.0225	0.0493	-0.0357	0.1834	0.0251	0.2003	-0.0082	0.0114
	0.6	-0.0308	0.0769	-0.0522	0.2208	0.0384	0.2155	-0.0145	0.0248
	0.8	-0.0360	0.1116	0.0717	0.2289	0.0430	0.2185	-0.0201	0.0334
500	0.2	0.0092	0.0221	0.0179	0.1056	0.0163	0.1239	-0.0025	0.0019
	0.4	0.0140	0.0301	0.0235	0.1389	0.0240	0.1276	-0.0045	0.0070
	0.6	0.0208	0.0417	0.0424	0.1807	0.0350	0.1341	-0.0087	0.0145
	0.8	0.0335	0.0749	0.0686	0.2170	0.0448	0.1517	-0.0187	0.0188

The MLEs, $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\beta}$ and $\hat{\delta}$ of θ_1, θ_2, β and δ, respectively, can be obtained by using a numerical method such as Newton-Raphson method. The observed Fisher information matrix $\mathbf{I}_N\left(\hat{\alpha},\hat{\beta}\right)$ is

$$\mathbf{I}_N\left(\hat{\theta}_1,\hat{\theta}_2,\hat{\beta},\hat{\delta}\right) = -\frac{1}{N}\begin{bmatrix} \frac{\partial^2 \log L}{\partial \theta_1^2} & \frac{\partial^2 \log L}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 \log L}{\partial \theta_1 \partial \beta} & \frac{\partial^2 \log L}{\partial \theta_1 \partial \delta} \\ & \frac{\partial^2 \log L}{\partial \theta_2^2} & \frac{\partial^2 \log L}{\partial \theta_2 \partial \beta} & \frac{\partial^2 \log L}{\partial \theta_2 \partial \delta} \\ & & \frac{\partial^2 \log L}{\partial \beta^2} & \frac{\partial^2 \log L}{\partial \beta \partial \delta} \\ Symm. & & & \frac{\partial^2 \log L}{\partial \delta^2} \end{bmatrix}_{\substack{\theta_1=\hat{\theta}_1 \\ \theta_2=\hat{\theta}_2 \\ \beta=\hat{\beta} \\ \delta=\hat{\delta}}} \quad (9)$$

4. Simulation Studies

We investigate, by Monte Carlo simulations, the properties of the estimators in terms of the bias and mean square error (MSE) for the Hougaard bivariate Weibull distribution. Bivariate Weibull random variates can be generated by using the results of Lu and Bhattacharyya(1990). One thousand estimates were computed, and deviations and squared deviations of the estimate from true value were averaged to obtain (estimated) bias and MSE for the following parameters: $N=100,500$, $T^0=2$, $U^0=3$ $\theta_1=2$, $\theta_2=4$, $\beta=3$, $\delta=0.2,0.4,0.6,0.8$.

The results of the simulations for various δ and N are shown in Table 1. One can see from the table that the (estimated) biases and MSEs of the MLEs increase as δ increases. It means that more accurate estimates can be obtained for the case where the two variables have strong positive correlation.

4.1 Comparisons with one-dimensional approach

Existing works on two-dimensional warranty data assume that the two variables have a functional relationship; see Lawless et al.(1996). If usage data are transformed to age data by means of a functional relationship between age and usage, general approaches dealing with univariate reliability data can be applied. If a univariate Weibull distribution is used

<Table 2. Comparison with one-dimensional approach>

	δ	1-dim. approach		2-dim. approach	
		Bias	MSE	Bias	MSE
θ_1	0.2	-0.0378	0.0228	0.0092	0.0221
	0.4	0.1628	0.0666	0.0140	0.0301
	0.6	0.4859	0.2182	0.0208	0.0417
	0.8	0.7449	0.5974	0.0335	0.0749
β	0.2	-0.0555	0.1286	0.0163	0.1239
	0.4	-0.4299	0.3366	0.0240	0.1276
	0.6	-0.6000	0.5688	0.0350	0.1341
	0.8	-0.6978	0.6752	0.0448	0.1517

to analyze one-dimensional data transformed with the functional relationship, estimates for parameters of the distribution can be compared with those of the bivariate Weibull distribution whose marginal distributions are univariate Weibull distributions.

Table 2 gives the results of the simulations with the same parameter set as Table 1 and $N = 500$. One can see from the table that (i) the (estimated) biases and MSEs of two-dimensional approach are always smaller than those of one-dimensional approach, (ii) the difference of (estimated) biases and MSEs between one-dimensional approach and two-dimensional approach increase as δ increases.

References

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

David W. C. and Kieron A. D. (1999), Analysis of Grouped Data from Field-Failure Reporting Systems, *Reliability Engineering and System Safety* **65**, 95-101.

Hougaard, P.(1986), A Class of Multivariate Failure Time Distributions, *Biometrica* **83**, 747-761.

Hu, J. X. and Lawless, J. F. (1996), Estimation from Truncated Lifetime Data with Supplementary Information on Covariates and Censoring Times, *Biometrika* **83**, 4, 747-761.

Hu, J. X., Lawless, J. F. and Suzuki, K. (1998), Nonparametric Estimation of a Lifetime Distribution When Censoring Times Are Missing, *Technometrics* **40**, 3-13.

Kalbfleisch, J. D. and Lawless, J. F. (1988), Estimation of Reliability in Field-Performance Studies, *Technometrics* **30**, 365-388.

Kalbfleisch, J. D., Lawless, J. F. and Robinson, J. A. (1991), Methods for the Analysis and Prediction of Warranty Claims, *Technometrics* **33**, 273-285..

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.

Lu, J.C. and Bhattacharyya, C.K. (1990), Some New Constructions of Bivariate Weibull Models, *Annals of Institute of Statistical Mathematics* **42**, 543-559.

Oh, Y.S. and Bai, D.S. (2001), Field Data Analysis with After-Warranty Failure Data, *Reliability Engineering and System Safety* **72**, 1-8.

Suzuki, K. (1985a), Estimation of Lifetime Parameters from Incomplete Field Data, *Technometrics* **27**, 263-272, 1985a.

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.