신뢰성응용연구 신뢰성기술
제10권, 제4호. pp.251-263, 2010

2차원 데이터의 여러 가지 분석방법

백재욱

한국방송통신대학교 정보통계학과

Various types of analyses for two-dimensional data

Jaiwook Baik

Department of Information Statistics
Korea National Open University
Seoul, South Korea

Abstract

Modelling for failures is important for reliability analysis since failures of products such as automobiles occur as both time and usage progress and the results from the proper analysis of the two-dimensional data can be used for establishing warranty assurance policy. Hence, in this paper general issues which concern modelling failures are discussed, and both one-dimensional approaches and two-dimensional approaches to two-dimensional data are investigated. Finally non-parametric approaches to two-dimensional data are presented as a means of exploratory data analyses.

Keywords : Two-dimensional data, Bivariate Weibull models, Minimal repair, Perfect repair.

이 논문은 2009년 한국방송통신대학교 학술연구비지원을 받아 작성된 것임
논문접수일 : 2010년 06월 03일 논문수정일 : 2010년 06월 16일 게재확정일 : 2010년 06월 25일

1. Introduction

The modelling of failures is an important element of reliability modelling. In one-dimensional failure modelling, failures are random points along a one-dimensional axis representing age or usage. For many products (for example automobiles), 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. Many different types of models are used and these can be found in Lawless (1982), Blischke and Murthy (2000 and 2006) and Meeker and Escobar (1998). One-dimensional modelling has received considerable attention and so there is a vast literature covering this area. In contrast, two-dimensional failure modelling has received relatively little attention.

The outline of the paper is as follows. In Section 2 we discuss various issues relating to the modelling of failures. Section 3 deals with one dimensional approach to two dimensional data while in Section 4 a two dimensional approach to two dimensional data is discussed. Section 5 deals with the exploratory approach for two dimensional data. Finally, we conclude with some comments and remarks in Section 6.

2. Modelling Failures

In this section we briefly discuss various issues of importance in the modelling of failures.

2.1 Product Failures

Products can vary from simple (such as an electric kettle) to complex (such as an aeroplane). A product can be viewed as a system consisting of several parts and can be decomposed into a hierarchy of levels with the system at the top level and components at the lowest level and several levels (such as sub-system, assembly, sub-assembly and so on) in between. The failure of a product is due to the failure of one or more of its components.

The occurrence of failure depends on several factors. These include decisions made during the design and manufacture of the product, usage intensity and operating environment, and the maintenance actions (corrective and preventive) carried out during the operating life. The modelling of failures can be done at any level ranging from system to component level.

2.2 Approaches to Modelling

The approach to modelling depends on the kind of information available and the goal of the modelling. There are two basic approaches to modelling failures as indicated below.

(i) Black-box Approach: Here the modelling is based solely on failure and censored data for similar items. This approach is used when there is very little understanding of the different mechanisms that lead to product failure or when the unit is too complex. This approach is also known as data based or empirical modelling.

(ii) White-box Approach: Here the failure modelling at the component level is based on the different mechanisms that lead to failure. At the system level, the failure is done in terms of the failures of the different components. This approach is also known as physics based modelling.

It should be noted that most statistical modelling incorporates features from both approaches and that the collaborative nature of such modelling is critical. The engineer or scientist needs to bring their expertise to develop the appropriate models in the white-box approach, whereas the statistician needs to be able to determine the appropriate data analysis for the given data set in its context.

2.3 First and Subsequent Failures

One needs to differentiate between the first failure and subsequent failures. The subsequent failures depend on the type of actions used to rectify the failures. In the case of a non-repairable item (component, system or something in between), the failed item needs to be replaced by either a new or used item to make the product functional. In the case of a repairable item, the product can be made operational through the repair of the failed item. Three types of repair are indicated below:

(i) Minimal repair, which restores the item to the condition just before failure;
(ii) Perfect repair (which makes the item as good as new); and
(iii) Imperfect repair that results in the item being better than what it was prior to failure but not as-good-as-new.

3. One dimensional approach to two-dimensional data

When failure depends on age and usage, one needs a two-dimensional failure model. Two different approaches (one-dimensional and two-dimensional approaches) have been proposed and we discuss both of these in this and next sections.
Here, the two-dimensional problem is effectively reduced to a one-dimensional problem by treating usage as a random function of age.

3.1 Modelling First Failure

Let $X(t)$ denote the usage of the item at age t. In the one-dimensional approach, $X(t)$ is modeled as a linear function of t and so given by

$$X(t) = \Gamma t$$

where Γ, $0 \leq \Gamma < \infty$, represents the usage rate and is a non-negative random variable with a distribution function $G(r)$ and density function $g(r)$.
The hazard function, conditional on $\Gamma = r$ is given by $h(t|r)$. Various forms of $h(t|r)$ have been proposed and one such is the following polynomial function:

$$h(t|r) = \theta_0 + \theta_1 r + \theta_2 t + \theta_3 X(t) + + \theta_4 t^2 + + \theta_5 t X(t)$$

The conditional distribution function for the time to first failure is given by

$$F(t|r) = 1 - \exp\left\{-\int_0^t h(u|r)du\right\}$$

On removing the conditioning, we have the distribution function for the time to first failure given by

$$F(t) = \int_0^\infty \left(1 - \exp\left\{-\int_0^t h(u|r)du\right\}\right) g(r)dr$$

3.2 Modelling Subsequent Failures

The modelling of subsequent failures, conditional on $\Gamma = r$, follows along the lines similar to that in Nakagawa and Kowada (1983) and Murthy (1991) for minimal repair

case and Cox (1967) for perfect repair case. As a result, under minimal repair, the failures over time occur according to a non-homogeneous Poisson process with intensity function $\lambda(t|r) = h(t|r)$ and, under perfect repair, the failures occur according to the renewal process associated with $F(t|r)$.
The bulk of the literature deals with a linear relationship between usage and age. See, for example, Lawless et al (1995) and Gertsbakh and Kordonsky (1998). Iskandar and Blischke (2003) deal with motorcycle data.

4. Two dimensional approach to two dimensional data

4.1 Modelling First Failure

Let T and X denote the system's age and usage at its first failure. In the two-dimensional approach to modelling, (T, X), is treated as a non-negative bivariate random variable and is modelled by a bivariate distribution function

$$F(t, x) = P\{T \leq t, X \leq x\}; t \geq 0, x \geq 0.$$

The survivor function is given by

$$\overline{F}(t, x) = \Pr\{T > t, X > x\} = \int_t^\infty \int_x^\infty f(u, \nu) d\nu du.$$

If $F(t, x)$ is differentiable, then the bivariate failure density function is given by

$$f(t, x) = \frac{\partial^2 F(t, x)}{\partial t \, \partial x}.$$

The hazard function is defined as

$$h(t, x) = f(t, x) / \overline{F}(t, x),$$

with $h(t, x)\delta t \delta x$ defining the probability that the first system failure will occur in the rectangle $[t, t+\delta t) \times [x, x+\delta x)$ given that $T > t$ and $X > x$. Note, however, that this is not the same as the probability that the first system failure will occur in the rectangle $[t, t+\delta t) \times [x, x+\delta x)$ given that it has not occurred beforetime t and usage x, which is given by $[f(t,x)/(1-F(t,x))]\delta t\,\delta x$.

4.2 Bivariate Weibull Models

A variety of bivariate Weibull models have been proposed in the literature. We indicate the forms of the models and interested readers can obtain more details from Murthy et al (2003).

Model 1 [Marshall and Olkin (1967)]

$$\overline{F}(t,x)=\exp\left\{-[\lambda_1 t^{\beta_1}+\lambda_2 x^{\beta_2}+\lambda_{12}\max(t^{\beta_1},x^{\beta_2})]\right\}$$

Model 2 [Lee (1979)]

$$\overline{F}(t,x)=\exp\left\{-[\lambda_1 c_1^{\beta} t^{\beta}+\lambda_2 c_2^{\beta} x^{\beta}+\lambda_{12}\max(\lambda_1 c_1^{\beta} t^{\beta},\lambda_2 c_2^{\beta} x^{\beta})]\right\}$$

Model 3 [Lee (1979)]

$$\overline{F}(t,x)=\exp\left\{-\lambda_1 t^{\beta_1}-\lambda_2 x^{\beta_2}-\lambda_0 \max(t,x)^{\beta_0}\right\}$$

Model 4 [Lu and Bhattacharyya (1990)]

$$\overline{F}(t,x)=\exp\left\{-\left[(t/\theta_1)^{\beta_1/\delta}+(x/\theta_2)^{\beta_2/\delta}\right]^{\delta}\right\}$$

$$\overline{F}(t,x)=\left[1+\left[\left\{\exp\left[(t/\theta_1)^{\beta_1}\right]-1\right\}^{1/\gamma}+\left\{\exp\left[(x/\theta_2)^{\beta_2}\right]-1\right\}^{1/\gamma}\right]^{\gamma}\right]^{-1}$$

$$\overline{F}(t,x)=\exp\left\{-(t/\alpha_1)^{\beta_1}-(x/\alpha_2)^{\beta_2}-\delta h(t,x)\right\} \tag{1}$$

Different forms for the function of $h(t,x)$ yield a family of models. One form for $h(t,x)$ is the following:

$$h(t,x)=[(t/\alpha_1)^{\beta_1/m}+(x/\alpha_2)^{\beta_2/m}]^m \tag{2}$$

which results in

$$\overline{F}(t,x) = \exp\left\{-(t/\alpha_1)^{\beta_1} - (t/\alpha_2)^{\beta_2} - \delta[(t/\alpha_1)^{\beta_1/m} + (x/\alpha_2)^{\beta_2/m}]^m\right\}$$

Two other variations are

$$\overline{F}(t_1, t_2) = \exp\left\{-(t_1/\alpha_1)^{\beta_1} - (t_2/\alpha_2)^{\beta_2} - \delta\left\{1 - \exp\left[-(t_1/\alpha_1)^{\beta_1}\right]\right\}\left\{1 - \exp\left[-(t_2/\alpha_2)^{\beta_2}\right]\right\}\right\}$$

$$\overline{F}(t_1, t_2) = \left[1 + \left[\left\{\exp\left[(t_1/\alpha_1)^{\beta_1}\right] - 1\right\}^{1/\gamma} + \left\{\exp\left[(t_2/\alpha_2)^{\beta_2}\right] - 1\right\}^{1/\gamma}\right]^{\gamma}\right]^{-1}$$

Model 5 [Sarkar (1987)]

$$\overline{F}(t_1, t_2) = \begin{cases} \exp\{-(\lambda_1 + \lambda_{12})t_1^{\beta_1}\}\{1 - [A(\lambda_2 t_1^{\beta_1})]^{-\gamma}[A(\lambda_2 t_2^{\beta_2})]^{1+\gamma}, \ t_1 \ge t_2 > 0; \\ \exp\{-(\lambda_2 + \lambda_{12})t_2^{\beta_2}\}\{1 - [A(\lambda_1 t_2^{\beta_2})]^{-\gamma}[A(\lambda_1 t_1^{\beta_1})]^{1+\gamma}, \ t_2 \ge t_1 > 0; \end{cases}$$

where $\gamma = \lambda_{12}/(\lambda_1 + \lambda_2)$ and $A(z) = 1 - e^{-z}$, $z > 0$.

Model 6 [Lee (1979)]

$$\overline{F}(t_1, t_2) = \exp\left\{-\left(\lambda_1 t_1^{\beta_1} + \lambda_2 t_2^{\beta_2}\right)^{\gamma}\right\}$$

Many other non-Weibull models can also be used for modelling. For more on this see Johnson and Kotz (1972) and Hutchinson and Lai (1990). Kim and Rao (2000), Murthy et al (1995), Singpurwalla and Wilson (1998), and Yang and Nachlas (2001) deal with two-dimensional warranty analysis.

4.3 Modelling Subsequent Failures

4.3.1 Minimal Repair

Let the system's age and usage at the j^{th} failure be given by t_j and x_j, respectively. Under minimal repair, we have that

$$h(t_j^+, x_j) = h(t_j^-, x_j),$$

as the hazard function after repair is the same as that just before failure. Note that there is no change in the usage when the failed system is undergoing minimal repair.

Let $\{N(t,x) : t \geq 0, x \geq 0\}$ denote the number of failures over the region $[0,t) \times [0,x)$. Unfortunately, as there is no complete ordering of points in two dimensions, there is no analogous result to that obtained for minimal repair in one dimension. In particular, the hazard rate does not provide an intensity rate at a point (t,x) as the failure after the last failure prior to (t,x) may be either prior to time t (though after usage x) or prior to usage x (though after time t), as well as possibly being after both time t and usage x. Hence, not only is it more difficult to obtain the distribution for $\{N(t,x) : t \geq 0, x \geq 0\}$, it is also more difficult to obtain even the mean function for this process.

4.3.2 Perfect Repair

In this case, we have a two-dimensional renewal process for system failures and the following results are from Hunter (1974):

$$p_n(t,x) = F^{(n)}(t,x) - F^{(n+1)}(t,x),\ n \geq 0,$$

where $F^{(n)}(t,x)$ is the n-fold bivariate convolution of $F(t,x)$ with itself. The expected number of failures over $[0,t) \times [0,x)$ is then given by the solution of the two-dimensional integral equation

$$M(t,x) = F(t,x) + \int_0^t \int_0^x M(t-u, x-\nu) f(u,\nu) d\nu du.$$

4.3.3 Imperfect Repair

This has not been studied and hence is a topic for future research.

4.4 A Numerical Example for perfect repair case

We confine our attention to a model proposed by Lu and Bhattacharyya (1990), where the survivor function is given by (1) with $h(t,x)$ given by (2) with $\alpha_1, \alpha_2, \beta_1, \beta_2 > 0$, $\delta \geq 0$ and $0 < m \leq 1$. If $m = 1$ then the hazard function is given by

$$h(t,x) = (1+\delta)^2 \frac{\beta_1}{\alpha_1}\left(\frac{t}{\alpha_1}\right)^{\beta_1 - 1} \frac{\beta_2}{\alpha_2}\left(\frac{t}{\alpha_2}\right)^{\beta_2 - 1}$$

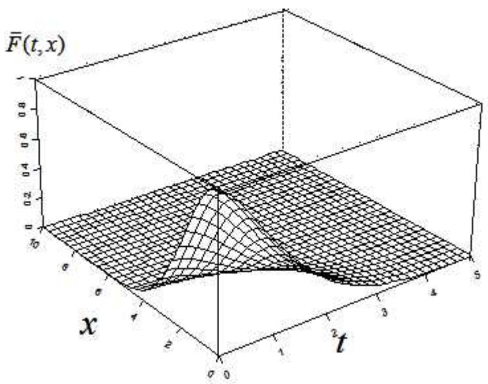
Let the model parameters be as follows:

$\alpha_1 = 2,\ \alpha_2 = 3,\ \beta_1 = 1.5,\ \beta_2 = 2.0,\ \delta = 0.5,\ m = 1$

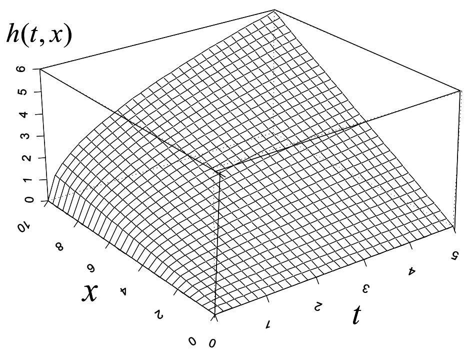
The units for age and usage are years and 10,000km respectively. The expected age and usage at first system failure are given by

$$E(T_1) = \theta_1 \Gamma(1/\beta_1 + 1) = 1.81\,(years) \text{ and } E(X_1) = \theta_2 \Gamma(1/\beta_2 + 1) = 2.66\,(10^3\,Km).$$

<Figure 1> is a plot of the survivor function $\overline{F}(t,x)$ and <Figure 2> is a plot of the hazard function $h(t,x)$. Note that $h(t,x)$ increases as t (age) and x (usage) increase since β_1 and β_2 are greater than 1.

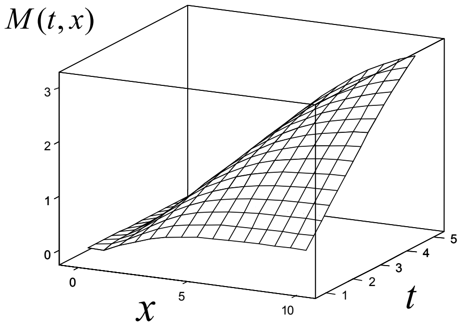


<Figure 1> Plot of survivor function $\overline{F}(t,x)$



<Figure 2> Plot of hazard function $h(t,x)$

The expected number of system failures in the rectangle $[0,t)\times[0,x)$ under replacement is given by the renewal function $M(t,x)$ in (40). <Figure 3> is a plot of $M(t,x)$, obtained using the two-dimensional renewal equation solver from Iskandar (1991).



<Figure 3> Plot of renewal function $M(t,x)$

5. Exploratory Data Analysis for two dimensional approach

We confine our discussion to the case of perfect repair.

5.1 Plot of Hazard Function [Nonparametric Approach]

We divide the region into rectangular cells. Cell (i,j) is given by $[i\delta_1,(i+1)\delta_1)\times[j\delta_2,(j+1)\delta_2)$, where δ_1 and δ_2 are the cells' width and height respectively. Let

N_{ij}^{f} : Number of items with failures times in cell $(i,j), i\geq 0, j\geq 0$

N_{ij}^{c} : Number of items with censoring times in cell $(i,j), i\geq 0, j\geq 0$

$N_{ij}^{f|sw}$: Number of failures in cells to the southwest of cell (i,j) $\left[=\sum_{i'=0}^{i-1}\sum_{j'=0}^{j-1}N_{i'\,j'}^{f}\right]$

$N_{ij}^{f|ne}$: Number of failures in cells to the northeast of cell $(i-1,j-1)$ $\left[=\sum_{i'=i}^{\infty}\sum_{j'=j}^{\infty}N_{i'\,j'}^{f}\right]$

Similarly define $N_{ij}^{c|ne}$ and $N_{ij}^{c|sw}$ for censored data.
A non-parametric estimator of the hazard function is

$$\widehat{h_{ij}}=\frac{N_{ij}^{f}}{N_{ij}^{f|ne}+N_{ij}^{c|ne}},\ i\geq 0,\ j\geq 0$$

5.2 Plot of Renewal Function [Nonparametric Approach]

A simple estimator of the renewal function in the case of complete data is given by the partial mean function over the cells; that is,

$$\widehat{M}(t_i,\ x_i)= \frac{N_{ij}^{f|sw}}{N},$$

where N is the total number of observations. A contour plot of this versus t and $x$can then be obtained.

6. Conclusions

In this paper we have looked at two dimensional failure modelling. We have discussed the two different approaches to two dimensional data. First is the one dimensional approach to two dimensional data where the two dimensional problem is effectively reduced to a one dimensional problem by treating usage as a random function of time. Second is the two dimensional approach to two dimensional data where a bivariate distributional function is used to explain the behaviour of the product. Next, nonparametric approaches to hazard function and renewal function are suggested for perfect repair case. However, there are some issues that need further study as follows.

1) Different empirical plotting of two-dimensional failure data.
2) Study on a new approach where usage may vary over time as opposed to linear trend in Section 3.
3) Modelling for subsequent failures in the case of imperfect repair
4) With the two dimensional warranty cases, the warranty can cease well before the time limit due to usage limit being exceeded. Appropriate models need to be developed for dealing with such.

References

[1] Blischke, W.R. and Murthy, D.N.P. (2000), Reliability, Wiley, New York

[2] Murthy, D.N.P. and Blischke, W.R. (2006), Warranty Management and Product Manufacture,

Springer, London

[3] Cox, D.R. (1967), RenewalTheory,Methuen,LondonXX

[4] Gertsbakh, I.B. and Kordonsky, H.B. (1998), Parallel time scales and two-dimensional manufacturer and individual customer warranties, IEE Transactions, 30, 1181-1189

[5] Hunter, J.J. (1974) Renewal theory in two dimensions: Basic results, Advances in Applied Probability, 6, 376-391

[6] Hutchinson, T.P. and Lai, C.D. (1990), Continuous Bivariate Distributions, Emphasising Applications, Rumsby Scientific, Adelaide, Australia

[7] Iskandar, B.P. (1991), Two-dimensional renewal function solver, Research Report No, 4/91, Dept. of Mechanical Engineering, The University of Queensland

[8] 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,623-656

[9] Johnson, N. L. and Kotz, S. (1972), Distributions in Statistics : Continuous Multivairate Distributions, Hohn Wiley & Sons, New York

[10] Kim, H.G. and Rao, B.M. (2000), Expected warranty cost of a two-attribute free-replacement warranties based on a bi-variate exponential distribution, Computers and Industrial Engineering, 38, 425-434.

[11] Lawless, J. F. (1982), Statistical Models and Methods for Lifetime Data, John Wiley & Sons, Inc., New York

[12] 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.

[13] Lee, L. (1979), Multivariate distributions having Weibull properties, Journal of Multivariate Analysis, 9, 267-277.

[14] Lu, J. C. and Bhattacharyya G. K. (1990), Some new constructions of bivariate Weibull Models, Ann. Inst. Statist Math., 42, 543-559.

[15] Marshall, A. W. and Olkin, I. (1967), A multivariate exponential distribution, Journal of American Statistical Association, 62, 30-44

[16] Meeker, W. Q. and Escobar, L. A. (1998), Statistical Methods for Reliability Data, John Wiley & Sons, Inc., New York

[17] Murthy, D. N. P. (1991), A note on minimal repair, IEEE Transactions on Reliability40:245-246

[18] Murthy, D.N.P., Iskandar, B.P and Wilson, R.J. (1995), Two-dimensional failure free warranties: Two-dimensional point process models, Operations Research, 43, 356-366.

[19] Murthy, D.N.P., Xie, M. and Jiang, R. (2003), Weibull Models,Wiley & Sons, New York

[20] Nakagawa, T. and Kowada, M. (1983), Analysis of a system with minimal repair and its

application to a replacement policy, European Journal of Operational Research 12: 176-182

[21] Sarkar, S.K. (1987), A continuous bivariate exponential distribution, journal of the American Statistical Association, 82, 667-675

[22] Singpurwalla, N.D. and Wilson, S.P. (1998), Failure models indexed by two scales, Advances in Applied Probability, 30, 1058-1072

[23] Yang, S-C and Nachlas, J.A. (2001), Bivariate reliability and availability modeling, IEEE Transactions on Reliability, 50, 26-35