SCIENTIFIC PROBLEMS
OF MACHINES OPERATION
AND MAINTENANCE

2 (166) 2011

LESZEK KNOPIK*

Model for instantaneous failures

Key words

Weibull distribution, mixture of distribution, instantaneous failures, maximum likelihood estimates, confidence interval.

Słowa kluczowe

Rozkład Weibulla, mieszana rozkładów, uszkodzenia nagłe, estymacja maksymalnej wiarygodności, przedział ufności.

Summary

The lifetime distribution is important in reliability studies. There are many situations in lifetime testing, where an item (technical object) fails instantaneously; therefore, the observed lifetime is reported as zero. We suggest a mixture of a singular distribution and Weibull distribution. We apply maximum likelihood to estimate parameters of the mixture. The methods are illustrated by a numerical example of the time between the failures for bus engines.

* University of Technology and Life Science, Faculty of Management, Fordońska St. 430, 85-789 Bydgoszcz, Poland, phone 37 19 659, e-mail: knopikl@utp.edu.pl.

Introduction

An important topic in the field of lifetime analysis is to select and specify the most appropriate life distribution that describes the time to failure of a component, assembly or system (see [11]). Occurrence of instantaneous or early failures in lifetime testing is observed in sets of failures of machines. These occurrences may be due to faulty construction or inferior quality. Some failures result from natural damages of the machine, while other failures may be caused by inefficient repairs of previous failures resulting from incorrect organisation of those repairs. These situations can be modelled by modifying commonly used parametric models, such as exponential, gamma and Weibull distributions.

In papers [9] and [10], the set of failures of a machine is divided into two subsets, namely, into the set of primary failures and the set of secondary failures. This division suggests that the population of lifetime is heterogeneous. The population of time before failures can be described by using the statistical concept of mixture. This mixture, in a particular case, has the unimodal failure rate function. In paper [9], special attention has been paid to the determination of the shape of the failure rate function from the mixture of the exponential distribution and distribution with a linear increasing failure rate function. It is clear that instantaneous failures can be primary failures or secondary failures. In this paper, the mixture of a singular and Weibull distributions is considered.

In this paper, we will indicate that the mixture of a singular and Wiebull distributions is useful to describe the lifetime of machines. A numerical example is also provided to illustrate the practical impact of this approach. In this example, n = 1430 failures of a bus engine is studied. This example shows that, in this case, a mixture of singular distribution and exponential distribution is sufficient.

The model of lifetime distribution

We consider a family of continuous distribution functions $F(x;\Theta)$, where Θ is a set of the parameters, $F(0, \Theta) = 0$. To accommodate a real life situation, where instantaneous failures are observed at the origin, the model $F(x;\Theta)$ is modified to the model $G(x; \Theta, p)$ by using a mixture in the proportion $1-p$ and p with respect to the singular random variable Z at zero and with the random variable T with the distribution function $F(x; \Theta)$.

Thus, the modified distribution function of lifetime is given as

$$G(x;\Theta,p)=\begin{cases} 1-p & \text{for} \quad x=0 \\ 1-p+pF(x;\Theta) & \text{for} \quad x>0 \end{cases} \tag{1}$$

and the corresponding probability density function as

$$f(x;\Theta,p)=\begin{cases} 1-p & \text{for} \quad x=0 \\ 1-p+pf(x;\Theta) & \text{for} \quad x>0 \end{cases}$$

The problem of statistical inference about (Θ, p) has received considerable attention, particularly when T is exponential. Some of the early references are Aitchison [1], Kleyle and Dahiya [4], Jayade and Parasad [2], Muralidharan [5, 6], Kale and Muralidharan [3] and the references contained therein. Muralidharan and Kale [7] considered the case where F is a two parameter gamma distribution with shape parameter β and scale parameter α, and they obtained a confidence interval for $\delta = p\alpha\beta$, assuming α as being known and unknown respectively. The purpose of this paper is to consider the model G given by (1) when F(x; Θ) is two a parameter Weibull distribution with the parameters α and β and the distribution function

$$F(x;\alpha,\beta) = 1 - \exp(-(t^{\alpha}/\beta)) \quad \text{for} \ t \ge 0.$$

The maximum likelihood estimation

In paper [8], these distributions are considered, and the maximum likelihood estimates of the parameters p, α, and β are obtained.

Let $(X_1, X_2, \ldots, X_n)$ be a random sample size n. Then, the likelihood function is

$$L(x_1, x_2, \ldots, x_n; p, \alpha, \beta) = \prod_{i=1}^{n} g(x_i; p, \alpha, \beta) \tag{2}$$

where

$$g(x;p,\alpha,\beta) = \begin{cases} 1-p & \text{f} \quad x=0 \\ p\,x^{\alpha-1} e^{-x^{\alpha}/\beta} & \text{for} \quad x>0 \end{cases} \tag{3}$$

By (2) and (3) we obtain

$$L(x_1, x_2, \ldots, x_n; p, \alpha, \beta) = \prod_{i-1}^{n} (1-p)^{z(x_i)} \left(ax_i^{\alpha} e^{-x_i^{\alpha}/\beta}\right)^{1-z(x_i)} \tag{4}$$

where

$$z(x) = \begin{cases} 1 & \text{if} \quad x=0 \\ 0 & \text{if} \quad x>0. \end{cases}$$

After simple manipulation, we have

$$L(x_1, x_2, \ldots, x_n; p, \alpha, \beta) = (1-p)^{\sum z(x_i)} \left(\frac{p\alpha}{\beta}\right)^{n-\sum z(x_i)} \prod_{x_i>0} \left(x_i^{\alpha-1} e^{-x_i^{\alpha}/\beta} \right) \tag{5}$$

Taking the logarithm, we obtain

$$\ln(L) = \ln(1-p) \sum z(x_i) + (n - \sum z(x_i)) \ln \frac{p\alpha}{\beta} + (\alpha - 1) \sum_{x_i>0} \ln x_i - \frac{1}{\beta} \sum_{x_i>0} x_i^{\beta} \tag{6}$$

The likelihood equations are given by

$$\frac{\partial \ln L}{\partial p} = \frac{-n_0}{1-p} + \frac{n-n_0}{p} = 0, \quad \text{where } n_0 = \sum z(x_i) \tag{7}$$

$$\frac{\partial \ln L}{\partial \alpha} = \frac{n-n_0}{\alpha} + \sum_{x_i>0} \ln x_i - \frac{1}{\beta} \sum_{x_i>0} x_i^{\beta} \ln x_i = 0 \tag{8}$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{n_0 - n}{\beta} + \frac{1}{\beta^2} \sum_{x_i} x_i^{\alpha} = 0 \tag{9}$$

Then from (7), we have

$$\hat{\beta} = \frac{n-n_0}{n} \tag{10}$$

and from (9), we get

$$\hat{\beta} = \frac{1}{n-n_0} \sum_{x_i>0} x_i^{\alpha} \tag{11}$$

Using (9) and (10), we get

$$\frac{\sum_{x_i>0} x_i^{\alpha} \ln x_i}{\sum_{x_i>0} x_i^{\alpha}} - \frac{\sum_{x_i>0} \ln x_i}{n-n_0} = \frac{1}{\alpha} \tag{12}$$

Equation (12) is solved using iterative procedures to get α and then solved for (11) to get β.

Confidence interval for parameters p, α, and β

Note that

$$\frac{\partial \ln g}{\partial p} = \begin{cases} \dfrac{-1}{(1-p)} & \text{for} \quad x = 0 \\ \dfrac{1}{p} & \text{for} \quad x > 0 \end{cases}$$

$$\frac{\partial \ln g}{\partial \alpha} = \begin{cases} 0 & \text{for} \quad x = 0 \\ 1/\alpha + \ln x - (1/\beta) x^{\alpha} \ln x & \text{for} \quad x > 0 \end{cases}$$

$$\frac{\partial \ln g}{\partial \beta} = \begin{cases} 0 & \text{for} \quad x = 0 \\ x^{\alpha}/\beta - 1/\beta & \text{for} \quad x > 0 \end{cases}$$

[8] verifies that

$$E\left(\frac{\partial \ln g}{\partial p}\right) = 0, \quad E\left(\frac{\partial \ln g}{\partial \alpha}\right) = 0, \quad E\left(\frac{\partial \ln g}{\partial \beta}\right) = 0.$$

Now, we derive the second derivative

$$\frac{\partial^2 \ln g}{\partial p^2} = \begin{cases} 1/(1-p)^2 & \text{for} \quad x = 0 \\ -1/p^2 & \text{for} \quad x > 0 \end{cases}$$

$$\frac{\partial^2 \ln g}{\partial \alpha^2} = \begin{cases} 0 & \text{for} \quad x = 0 \\ -1/\alpha - (1/\beta) x^{\alpha} (\ln x)^2 & \text{for} \quad x > 0 \end{cases}$$

$$\frac{\partial^2 \ln g}{\partial \beta^2} = \begin{cases} 0 & \text{for} \quad x = 0 \\ 1/\beta^2 - (1/\beta^3) 2x^{\alpha} & \text{for} \quad x > 0 \end{cases}$$

The Fisher information matrix for (p, α, β) is

$$I(p, \alpha, \beta) = \begin{bmatrix} I_{pp} & I_{p\alpha} & I_{p\alpha} \\ I_{\alpha p} & I_{\alpha\alpha} & I_{\alpha\beta} \\ I_{\beta p} & I_{\beta\alpha} & I_{\beta\beta} \end{bmatrix}$$

Where by [8] we have

$$I_{pp} = E\left(-\frac{\partial^2 \ln g}{\partial p^2}\right) = \frac{1}{p(1-p)}$$

$$I_{p\alpha} = E\left(\frac{\partial \ln g}{\partial p}\frac{\partial \ln g}{\partial \alpha}\right) = 0$$

$$I_{p\beta} = E\left(\frac{\partial \ln g}{\partial p}\frac{\partial \ln g}{\partial \beta}\right) = 0$$

$$I_{\beta\beta} = E\left(-\frac{\partial^2 \ln g}{\partial \beta^2}\right) = \frac{p}{\beta^2}$$

$$I_{\alpha\beta} = E\left(\frac{\partial \ln g}{\partial \alpha}\frac{\partial \ln g}{\partial \beta}\right) = -\frac{p}{\alpha\beta}[1+\ln\beta - c]$$

$$I_{\alpha\alpha} = E\left(-\frac{\partial^2 \ln g}{\partial \alpha^2}\right) = \frac{p}{\alpha^2}[(c-\ln\beta)(c-\ln\beta-2)+\pi^2/6+1]$$

where c is Euler's constant, $c \approx 0.5772$.

The invert matrix $I^{-1}(p, \alpha, \beta)$ is

$$I^{-1}(p,\alpha,\beta) = \begin{bmatrix} p(1-p) & 0 & 0 \\ 0 & \frac{I_{\alpha\alpha}}{\Delta} & \frac{I_{\alpha\beta}}{\Delta} \\ 0 & \frac{I_{\beta\alpha}}{\Delta} & \frac{I_{\beta\beta}}{\Delta} \end{bmatrix}$$

where $\Delta = \dfrac{\pi p^2}{6\alpha^2\beta^2}$.

Using the estimate variances, we can propose the large sample confidence intervals for p, α, and β.

The approximate $(1-\delta)\%$ confidence intervals for p are given by

$$\left(\hat{p}-u_{\delta/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\;,\;\hat{p}+u_{\delta/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

for α

$$\left(\hat{\alpha}-u_{\delta/2}\sqrt{\frac{I_{\alpha\alpha}}{n\Delta}}\;,\;\hat{\alpha}+u_{\delta/2}\sqrt{\frac{I_{\alpha\alpha}}{n\Delta}}\right)$$

for β

$$\left(\hat{\beta}-u_{\delta/2}\sqrt{\frac{I_{\beta\beta}}{n\Delta}}\;,\;\hat{\beta}+u_{\delta/2}\sqrt{\frac{I_{\beta\beta}}{n\Delta}}\right)$$

where $u_{\delta/2}$ is $(1-\delta)\%$ a percentile of the standard Gaussian distribution. The length of confidence interval for p does not depend on α and β and it is well known. The length of confidence interval for α does not depend on α and for the length of confidence interval for β does not depend on β.

The numerical example

The object of investigation is a real municipal bus transport system within a large agglomeration. The analysed system operates and maintains 190 municipal buses of various makes and types. In this section, we consider a real lifetime data on failure of bus engines. The data set contains n = 1430 times between failures of a bus. This set contains $n - n_0 = 370$ times equal to zero.

We apply the maximum likelihood estimates of the parameters p, α, and β. As initial solution of the equation (12), we give α = 1. We then calculate the values of the parameters p = 0.23, α = 1.02, and β = 12.4, and the corresponding confidence interval, for p (0.21, 0.24), for α (0.94, 1.2) and β (10.2, 11.2). For these values of parameters, we prove the Pearson's test of fit and compute the associated p-value = 0.34. It shows a good conformity of the empirical data with the mixture distributions. Fig. 1 shows the probability density functions for the example.

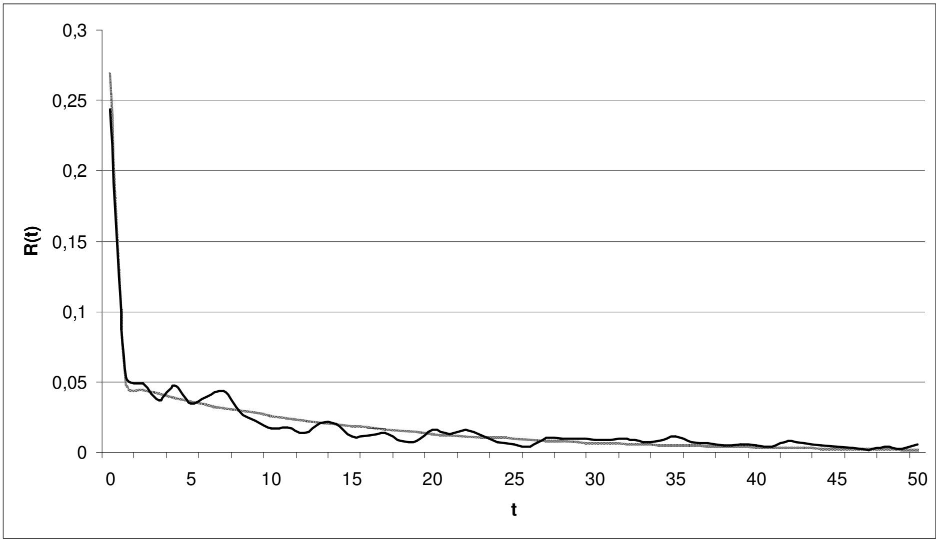

Fig. 1. Plots empirical and mixture density function
Rys. 1. Wykres gęstości empirycznej i mieszaniny

Conclusions

In this paper, we studied the lifetime model for instantaneous failures of bus engines. The estimation of parameters is approached by the method of maximum likelihood and the expected information matrix is derived. The estimates for α can be obtained as the solution of equation (12). The confidence intervals for p, α, and β depend on the information matrix $I(p, \alpha, \beta)$. Furthermore, this confidence interval can be used for a large sample. An application to real data set shows that this model may be applicable in practice. When the parameters are estimated, it is possible to apply further calculations, such as MTTF (Mean Time to Failure), burn-in time, the failure rate function, and the replacement time. The development of efficient parameter estimation methods for these mixture distributions and their application for times to failure modelling are topics for further study.

References

[1] Aitchison I.: On the distribution of a positive random variable having a discrete probability mass at origin, Journal of the American Statistical Associations, vol. 50, 1955, pp. 901-908.

[2] Jayade V.P., Parasad M. S.: Estimations of parameters of mixed failure time distribution. Communications Statistics, Theory and Methods, vol. 19, 1996, pp. 4667-4677.

[3] Kale B.K., Muralidharan K.: Optimal estimating equations in mixture distributions accommodating instantaneous or early failures, Journal Indian Statistical Associations, vol. 38, 2000, pp. 317-329.
[4] Kleyle R.M., Dahiya R.L.: Estimation of parameters of mixed failure time distribution from censored data, Communications Statistics, Theory and Methods, vol. 4, 1975, pp. 873-882.
[5] Muralidharan K.: Test for mixing proportion in mixture of a degenerate and exponential distributions, Journal Indian Statistical Associations, vol. 37, 1999, pp. 105-119.
[6] Muralidharan K.: The UMVUE and Bayes estimate of reliability of mixed failure time distribution, Communications Statistics, Simulation Computations, vol. 29, No 2, 2000, pp. 603-158.
[7] Muralidharan K.: Modified gamma distribution with singularity at zero, Communications Statistics, Simulation Computations, vol. 31, No 1, 2002, pp.143-158.
[8] Muralidharan K., Lathika P.: Analysis of instantaneous and early failures in Wiebull distribution, Metrika vol. 64, 2006, pp. 305-316.
[9] Knopik L.: Mixture of distributions as a lifetime distribution of a technical object, Scientific Problems of Machines Operation and Maintenance, 2010, vol. 45, 2(165), pp. 53-60.
[10] Wdzięczny A., Woropay M., Muślewski Ł.: Division and investigation of damages to technical object and their influence on the reliability of operation and maintenance systems Scientific Problems of Machines Operation and Maintenance, vol. 44, 2(154), 2008, pp. 31-43.
[11] Smalko Z., Jaźwiński J.: Using beta distribution for assessment of availability and reliability of technical transport facilities, Journal of KONBIN, vol. 11-12, No 1, 87-100.

Model dla uszkodzeń nagłych

S t r e s z c z e n i e

Rozkłady czasów życia są ważne w badaniach niezawodnościowych. Podczas testowania czasów życia istnieje wiele sytuacji, gdy element (obiekt techniczny) ulega natychmiastowemu uszkodzeniu i czas uszkodzenia jest zapisywany jako czas zerowy. Proponuje się mieszaninę rozkładu jednopunktowego i rozkładu Weibulla jako rozkład czasu życia. Do estymacji parametrów mieszaniny stosuje się metodę największej wiarygodności. Metodę zilustrowano przykładem numerycznym czasów między uszkodzeniami silników autobusowych.