22nd International Conference
ENGINEERING MECHANICS 2016
Svratka, Czech Republic, 9 – 12 May 2016

STATISTICAL ANALYSIS OF PARAMETERS OF RAIL VEHICLES

L. Knopik*, K. Migawa, P. Kolber*****

Abstract: *In this paper we propose a mixture of two different normal distributions to model heterogeneous of rail vehicles parameters. Maximum likelihood estimations of the parameters of mixture are obtained by using expectation algorithm. Illustrative examples based on real data (speed, number of axles, length of train, number of railway carriage and mass of train) are given.*

Keywords: Statistical analysis, mixture of distributions, rail vehicles, speed of train, mass of train, maximum likelihood method, expectation algorithm

1. Introduction

Monitoring the values of parameters of rail vehicles is a very important factor of safety in rail transportation. Values of these parameters are collected by DSAT system. This system screens the values of parameters of rail vehicles with various types of construction of bearing axles and train brake. It is applicable to various diameters of the wheels. System DSAT is installed on a straight rail line. System DSAT finds the following symptoms:

improvement of temperature of a bearing axle,
no working brakes – function,
exceeded pressure on axle or exceeded linear pressure.
deformation of surface wheels – function.

The system DSAT registers the following values of parameters:

speed [km/h],
number of axles,
length of train [m],
number of railway carriage,
mass of train [t].

The values of these parameters are the heterogeneous sets. It is a result of the fact that the rail vehicles moving on the analyzed path execute different tasks, such as transportation of people and cargo. In this paper, we use the mixture model for investigating a complex distribution of parameters of the rail vehicles. The mixture model has a wide variety of applications in technical and life science. Because of their usefulness as extremely flexible method of modeling, finite mixture models have continued to receive increasing attention over the recent years, from both practical and theoretical points of view, and especially for lifetime distributions. The problem application of the mixture of distributions to lifetime analysis is considered by Knopik (2010). Fitting the mixture distributions can be handled by variety of techniques, this includes graphical methods, the methods of moments, maximum likelihood and Bayesian approaches (Titterington et al., 1985; McLachan & Basford, 1988; Lindsay, 1995; McLachlan & Peel, 2000; Furhwirth-Schnatter, 2006). Now extensive advances have been introduced in the fitting of the mixture models especialy via maximum likelihood method. Among all, the maximum likelihood method becomes the first preference due to the existence of an associated statistical theory. The maximum

* Assoc. Prof. Leszek Knopik, PhD.: Faculty of Management, UTP University of Science and Technology, Fordońska Street 430, 85-790 Bydgoszcz, Poland, knopikl@utp.edu.pl

** Assoc. Prof. Klaudiusz Migawa, PhD.: Faculty of Mechanical Engineering, UTP University of Science and Technology, Prof. S. Kaliskiego Street 7, 85-789 Bydgoszcz, Poland, klaudiusz.migawa@utp.edu.pl

*** RNDr. Piotr Kolber, DSc.: Faculty of Mechanical Engineering, UTP University of Science and Technology, Prof. S. Kaliskiego Street 7, 85-789 Bydgoszcz, Poland, piotr.kolber@utp.edu.pl

likelihood method is making by expectation maximization algorithm (EM algorithm). The key property of the EM algorithm has been established in by Dempster et al. (1977) and McLachan & Krishan (1997). The EM algorithm is a popular tool for solving maximum likelihood problems in the context of a mixture model. We will focus on maximum likelihood techniques in this paper since the estimates tend to converge to true parameters values under general conditions. Maximum likelihood estimation procedures seek to find the parameters values that maximize the likelihood function evaluated at the observations. The purpose of this paper is to show that the mixture of the different normal distributions is the appropriate model distribution for the heterogeneous data of value of rail vehicle parameters.

2. Model of distribution of parameters

The fact that the analyzed sets are heterogeneous caused that in order to analyze the probability distribution of parameters of the rail vehicles is not applicable to the various distributions such Weibull and gamma. In this paper, we analyze two-component mixture distribution of distributions as the distribution of examined parameters. Let X_1 and X_2 be the independent random variables with the density functions $f_1(x)$ and $f_2(x)$, the cumulative distribution functions $F_1(x)$ and $F_2(x)$, the reliability functions $R_1(x)$ and $R_2(x)$, the failure rate function (hazard function) $\lambda_1(t)$ and $\lambda_2(t)$. Reliability function of the mixture X_1 and X_2 is described by the following formula:

$$R(x) = p\, R_1(x) + (1- p)\, R_2(x) \tag{1}$$

where p is the mixing parameter and $0 \le p \le 1$.

The failure rate function of the mixture can be written as the mixture (Knopik, 2010):

$$\lambda(t) = \omega(t)\, \lambda_1(t) + (1 - \omega(t))\, \lambda_2(t) \tag{2}$$

where $\lambda(t) = f(t) / R(t)$, $\omega(t) = pR_1(t) / R(t)$. Understanding the shape of the failure rate function is important in reliability theory and practice.

The basic problem is to infer about unknown parameters, on the basis of a random sample of size n on the observable random variable X. The first opinion of the data from the DSAT system shows that the mixture of two normal distributions is a proper model for analyzed parameters. The density function of the mixture of two normal distributions can be written in the following form:

$$f\left(x; m_1, m_2, \sigma_1^2, \sigma_2^2, p\right) = \frac{p}{\sqrt{2\pi\sigma_1^2}} \exp[-\frac{(x - m_1)^2}{2\sigma_1^2}] + \frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp[-\frac{(x - m_2)^2}{2\sigma_2^2}] \tag{3}$$

We will estimate five parameters m_1, m_2, σ_1, σ_2, p of the density (3). To estimate parameters $\Theta = (m_1, m_2, \sigma_1^2, \sigma_2^2, p)$ we will use the likelihood method (see Hasti et al. 2001). The likelihood function for the mixture (3) is:

$$L\left(x_1, x_2, ..., x_n; m_1, m_2, \sigma_1^2, \sigma_2^2, p\right) = \prod_{i=1}^{n} f(x_i : m_1, m_2, \sigma_1^2, \sigma_2^2, p) \tag{4}$$

We compute the first partial derivative of the logarithm of likelihood function:

$$\frac{\partial \ln L}{\partial m_1} = \sum_{i=1}^{n} \left(\frac{1}{A} \frac{p}{\sqrt{2\pi\sigma_1^2}} \exp[-\frac{(x_i - m_1)^2}{2\sigma_1^2}] \times \frac{(x_i - m_1)}{\sigma_1^2} \right) = 0 \tag{5}$$

$$\frac{\partial \ln L}{\partial m_2} = \sum_{i=1}^{n} \left(\frac{1}{A} \frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp[-\frac{(x_i - m_2)^2}{2\sigma_2^2}] \times \frac{(x_i - m_2)}{\sigma_2^2} \right) = 0 \tag{6}$$

$$\frac{\partial \ln L}{\partial \sigma_1^2} = \sum_{i=1}^{n} \left(\frac{1}{A} [-\frac{p}{2\sqrt{2\pi}} (\sigma_1^2)^{-\frac{3}{2}} \exp[-\frac{(x_i - m_1)^2}{2\sigma_1^2}] + \frac{p}{\sqrt{2\pi\sigma_1^2}} \exp[-\frac{(x_i - m_1)^2}{2\sigma_1^2}] \frac{(x_i - m_1)^2}{2(\sigma_1^2)^2}] \right) = 0 \tag{7}$$

$$\frac{\partial \ln L}{\partial \sigma_2^2} = \sum_{i=1}^{n}\left(\frac{1}{A}[-\frac{p}{2\sqrt{2\pi}}(\sigma_2^2)^{-\frac{3}{2}}\exp[-\frac{(x_i-m_2)^2}{2\sigma_2^2}]+\frac{1-p}{\sqrt{2\pi\sigma_2^2}}\exp[-\frac{(x_i-m_2)^2}{2\sigma_2^2}]\frac{(x_i-m_2)^2}{2(\sigma_2^2)^2}]\right)=0 \quad (8)$$

$$\frac{\partial \ln L}{\partial p} = \sum_{i=1}^{n}\left(\frac{1}{A}(\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp[-\frac{(x_i-m_1)^2}{2\sigma_1^2}]-\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp[-\frac{(x_i-m_2)^2}{2\sigma_2^2}])\right)=0 \quad (9)$$

where $A = f(x_i; m_1, m_2, \sigma_1^2, \sigma_2^2, p)$.

To find the maximum log – likelihood function, we set the first partial derivative equal to zero. In finite mixture model, the EM algorithm has been used as an effective method to find maximum likelihood parameters estimation.

3. Real data set

In this chapter, we will estimate the parameters m_1, m_2, σ_1^2, σ_2^2, p of the mixture of two normal distributions for the random variable X_1 – speed of train, X_2 – number of axles, X_3 – length of train, and X_4 – mass of train. By λ we describe the value of the goodness of fit statistics Kolmogorov-Smirnov. We used procedure (EM algorithm) given for special case of normal mixtures by Hastie et al. (2001). The estimated parameters, K-S test statistics and p-values for four random variables are given in Table 1. All the considered the parameters of rail vehicles shown good conformity with the empirical distributions and the mixture distributions.

Tab.1: Values of parameters of mixtures

Random variable	Parameters of mixture					goodness of fit statistic λ-KS	p-value
	m_1	m_2	σ_1	σ_2	P		
X_1 – speed	51.27	78.11	7.82	2.69	0.531	0.3780	0.99
X_2 – axles	37.60	151.49	12.57	43.10	0.531	0.6102	0.85
X_3 – length	191.92	599.77	72.52	109.84	0.547	0.9153	0.56
X_4 – mass	381.66	2051.8	39.21	788.59	0.738	0.8543	0.53

The graphs of the empirical distribution functions (Fe) and the mixture distribution function (Ft) are shown in Figure 1.

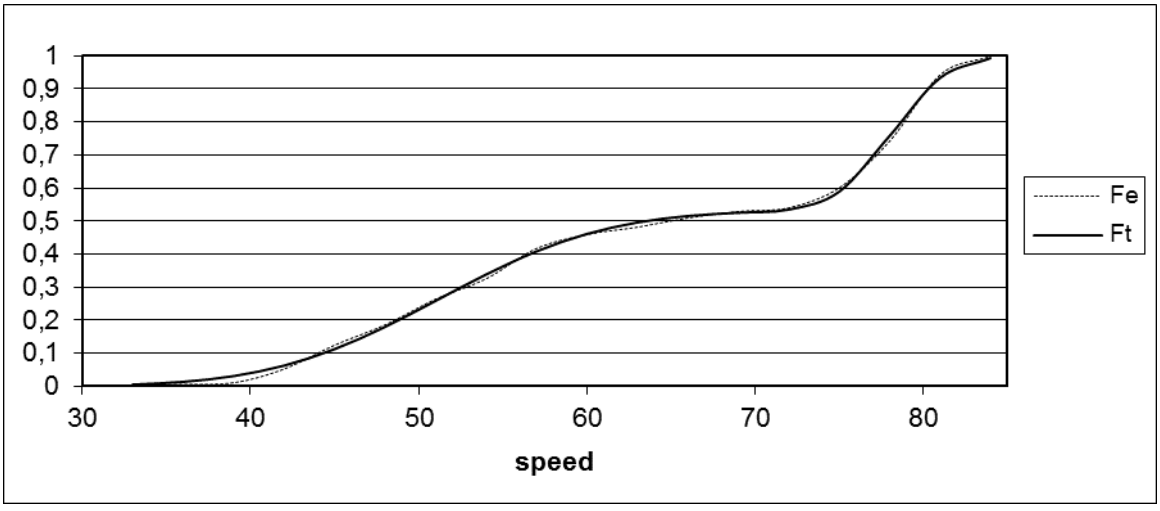
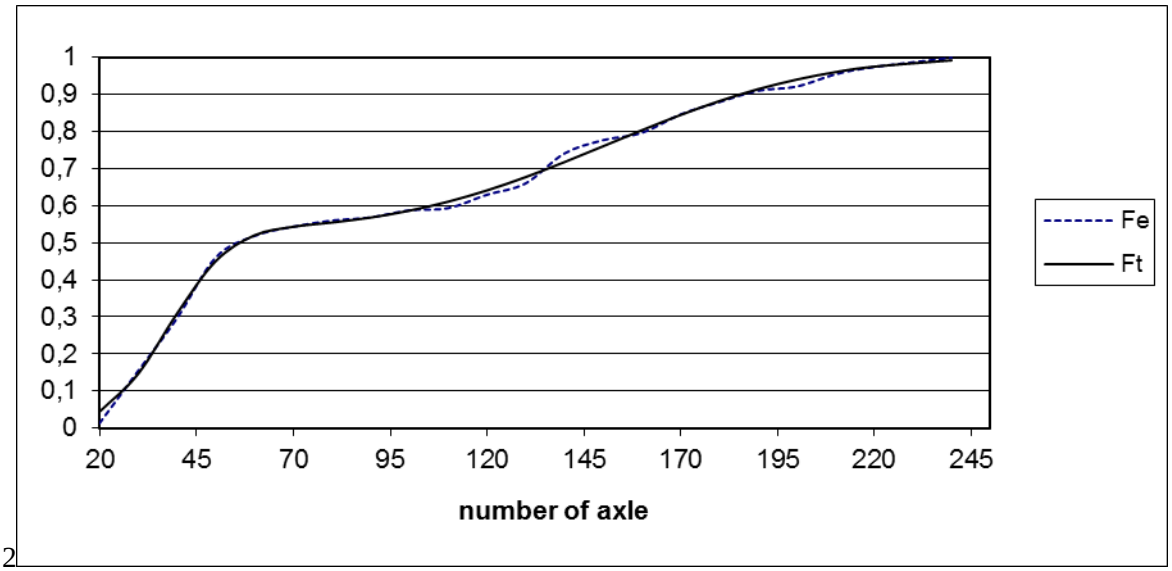


Fig. 1: Empirical distribution function and distribution function of mixture model for number of the speed

A graphical comparison of empirical distribution function and distribution function of the mixture model for number of axles is given in Figure 2.



2

Fig.2: Empirical distribution function and distribution function of mixture model for number of axles

4. Conclusions

We use the mixture of two-normal distributions for investigating complex probability distributions of parameters of the rail vehicles. It is shown that the mixture of the different normal distributions is useful for exploring the complex distributions. The probability distributions of all measured system parameters (speed, number of axles, length of rail vehicle, mass) are compatible with the calculated mixture of two normal distributions. Knowledge of the probability distributions of the load parameters of the railway line is useful for the design of the modernization of these lines. Lastly we fit the two-component mixture normal distribution to data set using EM algorithm to maximize the likelihood function.

References

Dempster, A.P., Larid, N.M. & Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal Statistical Society, Series B, 39, pp. 1-38.

Hastie, T., Tibshirani, R. & Friedman, J. (2001) The Elements of Statistical Learning: DataMining, Inference and Prediction. Springer Verlag.

Fruhwirth-Schnatter, S. (2006) Finite Mixture and Markov, Switching Models. Springer Verlag, New York.

Knopik, L. (2010) Mixture of distributions as a lifetime distribution of a technical object. Scientific Problems of Machines Operation and Maintenance, Vol. 162, No. 2, pp. 53-61.

Lindsay, B.G. (1995) Mixture Model*:* Theory, Geometry and Applications. Harward Institute of Mathematical Statistics.

MacLachan, G. J. & Basford, K. E. (1988) Mixture Model: Inference and Applications to Clustering. Marcel Dekker, New York.

MacLachlan, G. & Peel, D. (2000) Finite Mixture Models. John Wiley & Sons, New York.

MacLachan, G.J. & Krishan, T. (1997) The EM algorithm and extensions. John Wiley & Sons, New York.

Titterington, D.M., Smith, A.F.M. & Makov, U.E. (1985) Statistical Analysis of Finite Mixture Distribution. John Wiley & Sons, New York.