Journal of POLISH CIMAC
Faculty of Ocean Engineering & Ship Technology
GDAŃSK UNIVERSITY OF TECHNOLOGY

ANALYSIS OF PARAMETERS OF RAIL VEHICLES

Leszek Knopik

University of Technology and Life Science in Bydgoszcz
Faculty of Management Science
Fordonska 430 st, 85-789 Bydgoszcz, Poland
e-mail: knopikl@utp.edu.pl

Abstract

In this paper, mixture of two normal distributions is proposed to accommodate the values of rail vehicles parameters. We also present the most commonly used maximum likelihood estimation to fit the two component mixture of normal distribution using data sets of rail vehicles.

Keywords: *rail vehicles, speed of train, mass of train, mixture of distributions, maximum likelihood method, normal distribution, EM algorithm*

1. Introduction

Monitoring the values of parameters of rail vehicles is a very important factor of safety in rail transportation. In this paper, we analyze some parameters of rail vehicles. 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:

(a) improvement of temperature of a bearing axle – function GM,
(b) non working brakes – function GH,
(c) exceeded pressure on axle (NO) or exceeded linear pressure (NL) – function OK,
(d) deformation of surface wheels – function PM (PO),

The system DSAT registers the following values of parameters:

(e) speed [km/h],
(f) number of axles,
(g) length of train [m],
(h) number of railway carriage,
(i) 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 in [4, 5, 6, 7]. Fitting the mixture distributions can be handled by variety of techniques, this includes graphical methods, the methods of moments, maximum likelihood and Bayesian approaches (see Titterington et al. [14], McLachan G.J. and Basford K.E [9], Lindsay [8], McLachlan and Peel [10], Furhwirth- Schnatter [3]). Now extensive advances have been introduced in the fitting of the mixture models especially 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 likelihood method is making by expectation maximization algorithm (EM algorithm). The key property of the EM algorithm has been established by Dempster et al. [1] and McLachan G.J. and Krishan [11]. 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.

2. Analysis of measurements

The research object is a real transportation rail system. In this rail system, the gauge registers four parameters for $n = 360$ of trains for 6 days.

It is known that the measurement parameters are dependent. For this purpose, we calculate the matrix of correlation of a random variables (X_1, X_2, X_3, X_4), where X_1 is speed, X_2 is number of axles, X_3 is the length of train, X_4 is the mass of train. The correlation matrix K of the random variable $X = (X_1, X_2, X_3, X_4)$ is given as

$$K = \begin{bmatrix} 1 & & & \\ -0.81601 & 1 & & \\ -0.75576 & 0.923385 & 1 & \\ -0.68109 & 0.782682 & 0.689303 & 1 \end{bmatrix}$$

All correlation coefficients are statistical significant under p – value, $p < 0.0001$. In Fig. 1, we illustrate the relation between the mass and the length of the train, however Fig. 2 illustrates the relation between the mass and the speed.

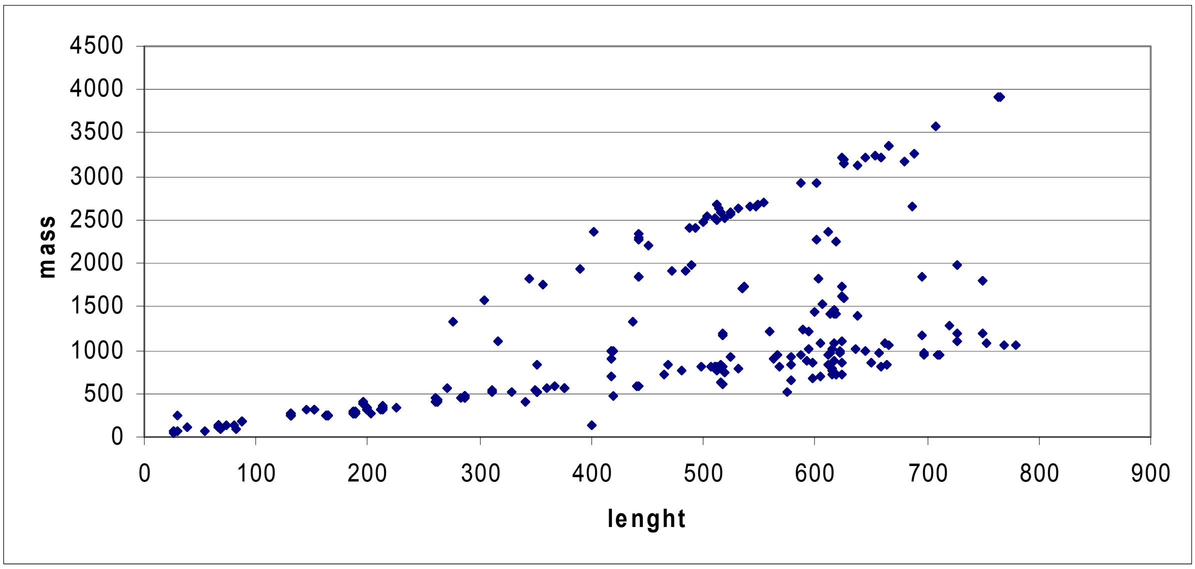

Fig. 1. Relation of mass versus length

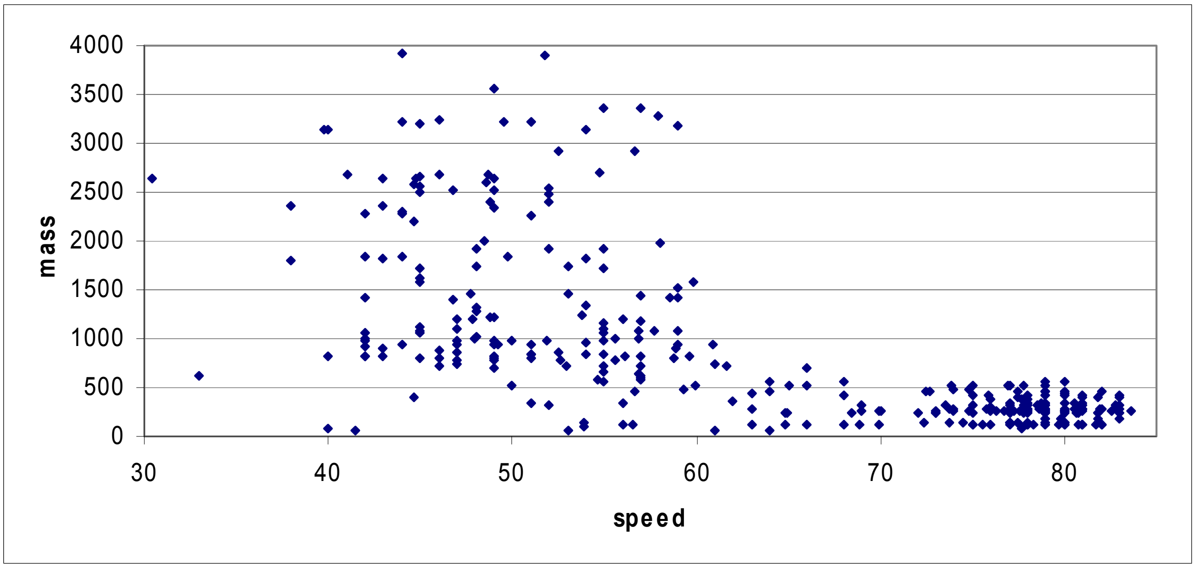

Fig. 2. Relation of mass versus speed

3. Model of distribution of parameters

The fact that the analyzed sets are heterogeneous caused that in order to analyze of 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 and Y 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 $\lambda_2(t)$. Distribution function of the mixture X and Y is described by the following formula:

$$F(x) = p\, F_1(x) + (1-p)\, F_2(x),$$

where p is the mixing parameter and $0 \le p \le 1$.
Analogously for the density function f(x) and the reliability function R(x), we can write as

$$f(x) = p\, f_1(x) + (1-p)\, f_2(x),$$

$$R(x) = p\, R_1(x) + (1-p)\, R_2(x).$$

The mean value of the random variable X is

$$EX = p\, m_1 + (1-p)\, m_2,$$

however the variance of X is

$$D^2X = p\, \sigma_1^2 + (1-p)\, \sigma_2^2.$$

The failure rate function of the mixture can be written as the mixture [4]:

$$\lambda(t) = \omega(t)\, \lambda_1(t) + (1-\omega(t))\, \lambda_2(t),$$

where $\lambda(t) = f(t) / R(t)$, $\omega(t) = pR_1(t) / R(t)$, $\lambda_1(t)$ and $\lambda_2(t)$ are the failure rate functions of the lifetimes X and Y. Understanding the shape of the failure rate function is important in reliability theory and practice.

Teicher [12, 13] introduced the concept of identifiability and developed a theory of identify mixtures. The concept of identifiability plays a vital role in the analysis of the finite mixture model. A mixture is identifiable if there exists a one to one correspondence between the mixing distributions and resulting mixture. The inference procedures on the mixture distributions can be meaningfully discussed only if the family of mixture distributions is identifiable.

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

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

We will estimate five parameters $m_1, m_2, \sigma_1, \sigma_2, p$ of the density (1). To estimate parameters $\Theta = (m_1, m_2, \sigma_1^2, \sigma_2^2, p)$ we will use the likelihood method. The likelihood function for the mixture (1) is

$$L(x_1, x_2, \ldots, x_n: m_1, m_2, \sigma_1^2, \sigma_2^2, p) = \prod_{i=1}^{n} f(x_i : m_1, m_2, \sigma_1^2, \sigma_2^2, p)$$

The logarithm of the likelihood function is

$$\ln L(x_1, x_2, \ldots, x_n: m_1, m_2, \sigma_1^2, \sigma_2^2, p) = \sum_{i=1}^{n} \ln f(x_i : m_1, m_2, \sigma_1^2, \sigma_2^2, p)$$

We compute the first partial derivative:

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

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

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

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

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

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 methods to find maximum likelihood parameters estimation.

4. Real data set

In this chapter of paper, we will estimate the parameters m_1, m_2, σ_1^2, σ_2^2, p of the mixture 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-KS we describe the value of the goodness of fit statistics λ- Kolmogorov.- Smirnov We used o procedure EM algorithm given for special case of normal mixtures by Hastie et al. [2]. The estimated parameters, K-S test statistics and p – values for four random variables are given in Table 1. All considered the parameters of rail vehicles shown good conformity 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	51.270	78.115	7.8284	2.6947	0.5315	0.3780	0.99
X_2 – axles	37.609	151.49	12.570	43.104	0.5310	0.6102	0.85
X_3–length	191.92	599.77	72.529	109.84	0.5476	0.9153	0.56
X_4– mass	381.66	2051.8	39.215	788.59	0.7381	0.8543	0.53

The graphs of the components (ft and ft-1) of the mixture and the density function (ft-2) of mixture are shown in Figure 3.

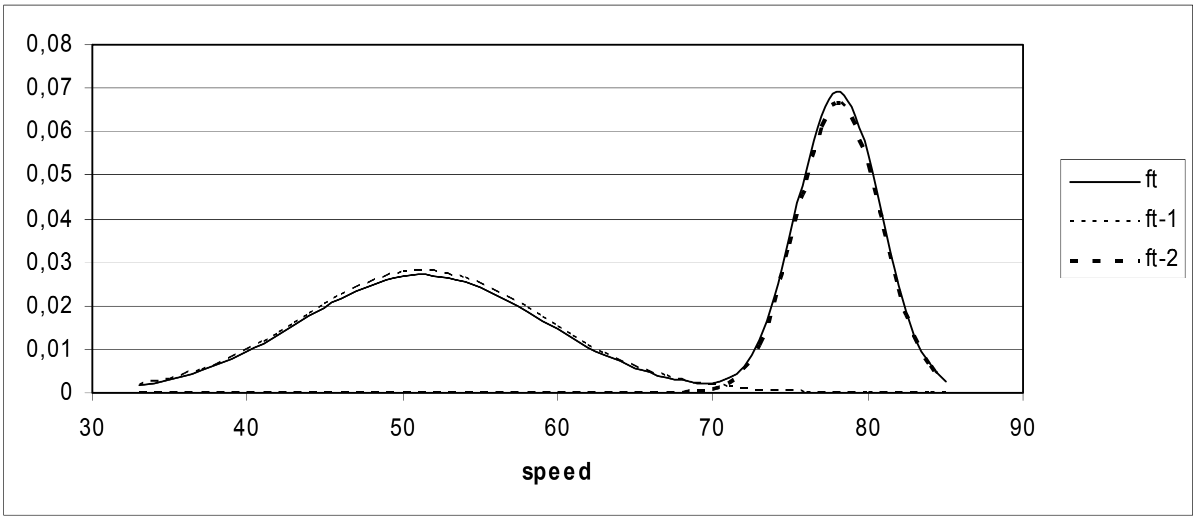

Fig. 3. The factors of mixture and the density function of mixture

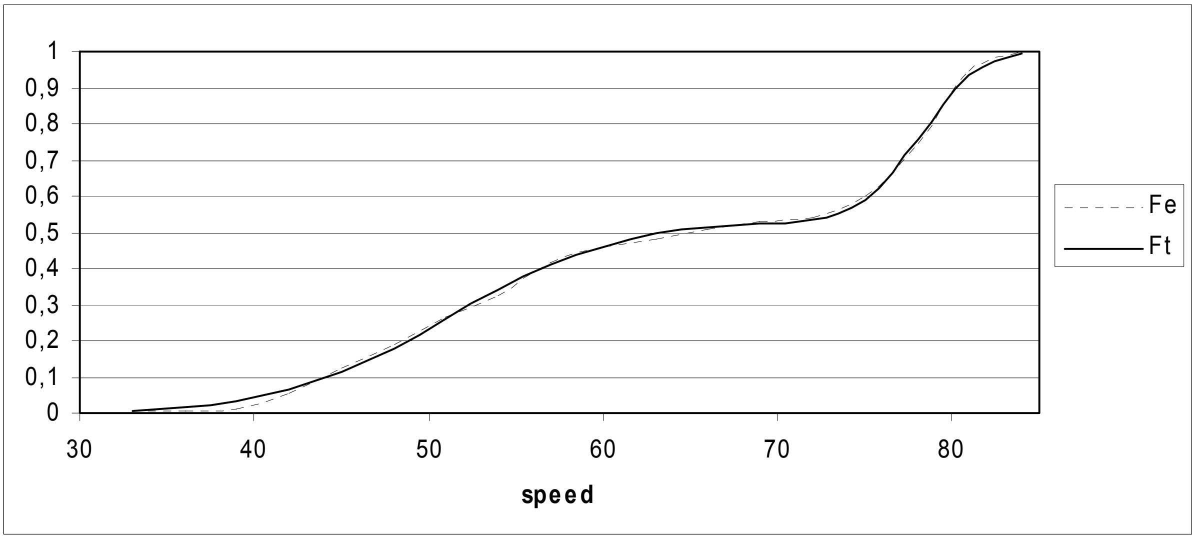

Fig. 4. The distribution function of speed (Fe) and the mixture (Ft)

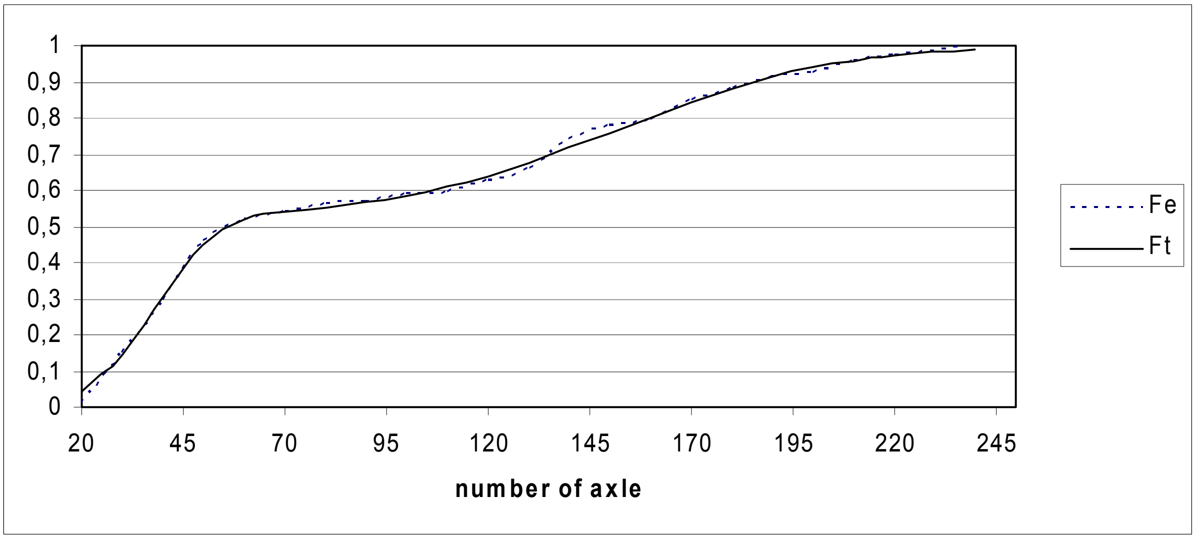

Fig. 5. The distribution function of number of axles (Fe) and the mixture (Ft)

The graphs the distribution function of the speed and distribution function of mixture are shown in Figure 4. We conclude that the mixture two-normal distributions is consistent model with the empirical distribution of the speed. The graphs of empirical distribution of the number of axles and the distribution function of mixture two-normal distributions are shown in Figure 5. In this case, we observe that both distribution are consistent too. The distribution functions for the length and the mass is given in Figure 6 and Figure 7.

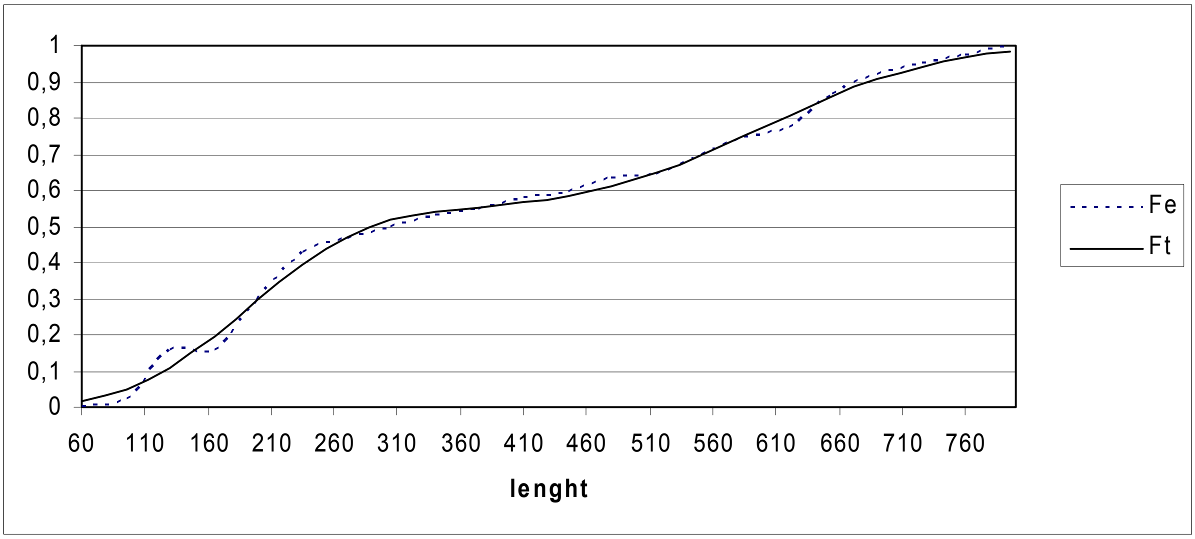

Fig. 6. The distribution function of the length (Fe) and the mixture (Ft)

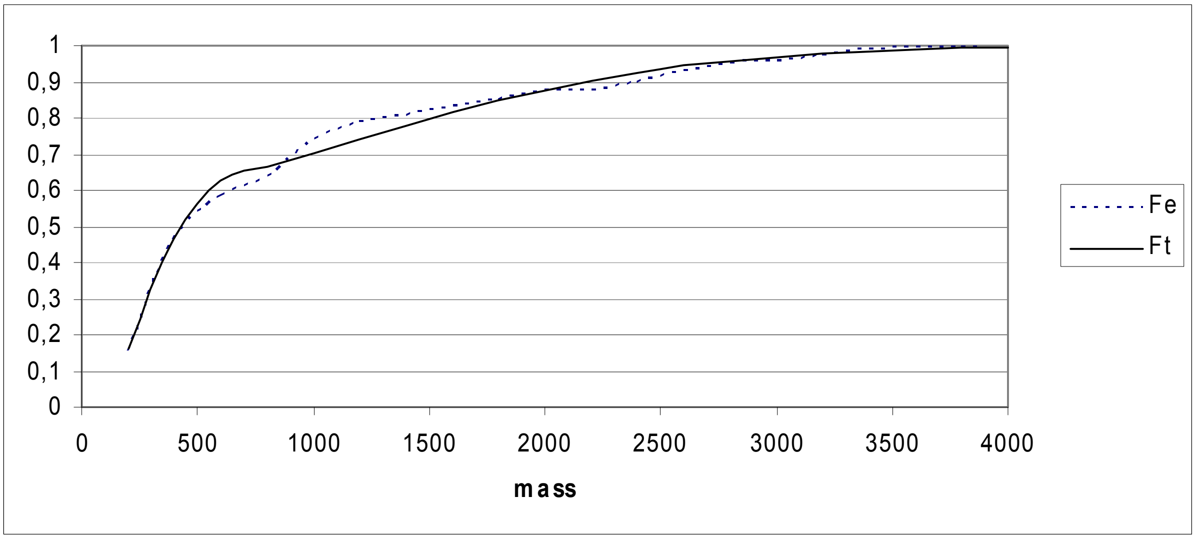

Fig. 7. The distribution function of the mass (Fe) and the mixture (Ft)

5. Conclusions

We use the mixture of two-normal distribution model for investigating complex distributions of the rail vehicles. It is shown that the mixture of two normal distributions is useful for exploring the complex distributions. Lastly, we fit the two component mixture normal distribution to data set using EM algorithm to maximize the likelihood function.

References

[1] Dempster, A.P., Larid, N.M. and Rubin, D.B., *Maximum likelihood from incomplete data via the EM algorithm*, Journal Statistical Society, Series B, 39, pp. 1–38, 1977.

[2] Hastie, T., Tibshirani, R. and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer Verlag, 2001.

[3] Fruhwirth-Schnatter, S.,Finite Mixture and Markov, *Switching Models*, New York, Springer, Verlag 2006.

[4] Knopik, L., *Mixture of distributions as a lifetime distribution of a technical object*, Scientific Problems of Machines Operation and Maintenance, Vol.162, No2, pp.53─61, 2010.
[5] Knopik, L., *Mixture of distributions as a lifetime distribution of a bus engine*, Journal of Polish CIMAC Vol.6, No1, pp. 119─129, 2011.
[6] Knopik, L., *Model of instantaneous failures*, Scientific Problems of Machines Operation and Maintenance, Vol.163, No4, pp. 123─133, 2011.
[7] Knopik, L., *Statistical analysis of failures*, Journal of Polish CIMAC Vol.7, No.2, pp. 91─95, 2012.
[8] Lindsay, B.G., *Mixture Model: Theory*, *Geometry and Applications*, Hayward Institute of Mathematical Statistics, 1995.
[9] MacLachan, G. J. and Basford K. E., *Mixture Model: Inference and Applications to Clustering,* Marcel Dekker, New York, 1988.
[10] MacLachlan, G. and Peel, D., *Finite Mixture Models*, John Wiley and Sons, New York, 2000.
[11] MacLachan, G.J. and Krishan, T., *The EM algorithm and extensions*, John Wiley & Sons, New York, 1997.
[12] Teicher, H., *Identifiability of mixtures*, The Annals of Mathematical Statistics, vol. 32, No.1, 244-249, 1961.
[13] Teicher, H., *Identifiability of finite mixtures*, The Annals of Mathematical Statistics, vol. 34, No.4, 1265- 1269, 1963.
[14] Titterington, D.M., Smith A.F.M. and Makov U.E., *Statistical Analysis of Finite Mixture Distribution*, John Wiley and Sons, New York, 1985.