Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 48-3 (2015) 219–224

Bootstrap and Maximum Entropy Based Small-sample Product Lifetime Probability Distribution

Hailong Suo,, Jianmin Gao, Zhiyong Gao, Hongqvan Jiang, Rongxi Wang

Xi' an Jiaotong University, Xi' an , Shaanxi, China

State Key Laboratory for Manufacturing System Engineering

(email:hl_suo@163. com)

ABSTRACT: How to build the small-sample data distribution is a major issue to be addressed for analyzing large power equipment lifetime. To find a better product lifetime probability distribution function, a new method based on bootstrap and maximum entropy is proposed to study on the problem of small-sample product life probability distribution. The result shows that this new method is more accurate and feasible compared with bootstrap and maximum entropy respectively. Finally, an actual example of small-sample product lifetime distribution shows that this new approach has better guiding significance for practice.

Keywords: bootstrap; maximum entropy; nonlinear programming; small sample; probability density function

1. INTRODUCTION

Spare parts inventory management plays an important supporting role in guaranteeing stable operation of equipment. In the process of building a spare parts inventory model, different empirical distribution functions are chosen to analyze and predict depending on the product life data. However, failure lifetime data shows small-sample characteristics, and several distribution hypotheses may be validated by the method of Kolmogolov-Smirnov, therefore, this traditional approach is invalid and not feasible. In this case, to study on the personalized and reasonable product lifetime probability density function is necessary according to the actual lifetime data. The purpose is to provide scientific and objective guidance or advices to inventory management.

Srinivasan, V.S.(Srinivasan, V.S.2004)describes a theoretical method of estimating the most general form of probability density functions of random variables, and the estimated probability density function depends on the mean value of exponential distribution of the initial random variable. Lv Zhen(Lv Zhen et al. 2008) presents a method for determining fatigue life distribution under the assumption that fatigue life obeys three-parameter Weibull distribution, and an experimental data with 200 specimens is employed to demonstrate the good agreement with the large samples. Radu Florescu(Radu Florescu et al. 2006)exposes an original technique to determine the empirical probability density function and the empirical cumulative distribution function, and to estimate moments of any order for a small size random sample (m=3 - 10) of a continuous random variable.

In summary, there are two main methods for the present study of the small-sample probability distribution problem：
1) Suppose the probability distribution and verify; 2) Build a probability distribution function by oneself. For the first method, the error is often large for the small-sample problem of product lifetime probability distribution; for the second method, practical application is more difficult because it is based on the particular sample data. In this paper, a new method based on bootstrap and maximum entropy is proposed to study on the problem of small-sample product life probability distribution. It aims to make an objective and reliable analysis of unknown small-sample product life probability distribution for large power equipment.

10.1016/j.ifacol.2015.06.084

2. BOOTSTRAP METHOD

2.1 Methods and Principles

Bootstrap method is proposed as a new statistical analysis by professor B.Efron in 1979, and in the next years it is being developed. Gray bootstrap method (GBM)(Wang Yanqing et al. 2014) is proposed to solve the problem of estimating frequency-varying RVS with small samples for environment testing. Ouysse, Rachida(Ouysse, Rachida. 2013)describes a fast approximation for the small sample bias correction of the iterated bootstrap. Its approximation adapts existing fast approximation techniques of the bootstrap p-value and quantile functions to the problem of estimating the bias function. The characteristics of the bootstrap is: there is no need to make any assumptions to overall distribution, and it can be inferred with the sample data by the computer technology.

The main idea of bootstrap is:

1) $X_i (i = 1,2,3,\cdots,n)$ are independent identically distributed samples, $X_i \sim F(X)$,sample distribution function F_n is built by X_i ;

2) Random sampling with replacement from F_n ,generating N bootstrap samples, $X^*_{(1)}, X^*_{(2)},\cdots, X^*_{(N)}$

$$X^*_{(i)} = (x^*_{i1}, x^*_{i2},\cdots, x^*_{im})\ (i = 1,2,3,\cdots,N), \tag{1}$$

m is the bootstrap sample size;

3) For every $X^*_{(i)}$, estimate value of Statistics $\hat{\theta}$ should be calculated

$$\hat{\theta}^*(i) = S(X^{*i}),\ i = 1,2,3,\cdots,N \tag{2}$$

4) The estimated standard error $\hat{se}_N$ from bootstrap is the standard deviation of $\hat{\theta}^*(i)$

$$\hat{se}_N = \left(\sum_{i=1}^{N} \frac{[\hat{\theta}^*(i) - \hat{\theta}^*(\bullet)]}{N-1}\right)^{\frac{1}{2}} \text{Here,} \hat{\theta}^*(\bullet) = \sum_{i=1}^{N} \frac{\hat{\theta}^*(i)}{N} \tag{3}$$

2.2 Analysis of Examples

To verify the accuracy of the probability density function by the bootstrap method, three subsamples(sample size:15、30、60) are in turn selected from the overall sample of standard normal distribution, and results of 1000 re-sampling process by the bootstrap method are analyzed and compared with standard normal probability density function. The results are as follows:

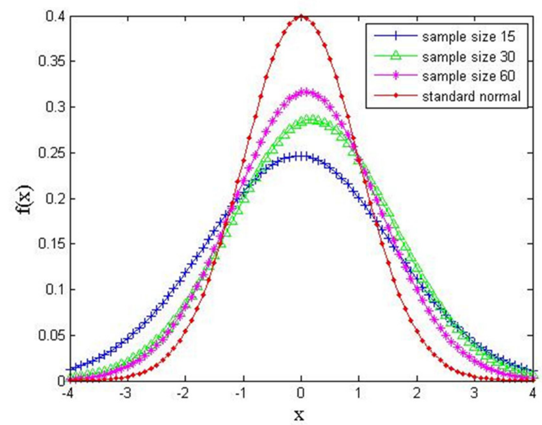

Figure 1 The probability density functions of different sample sizes based on bootstrap

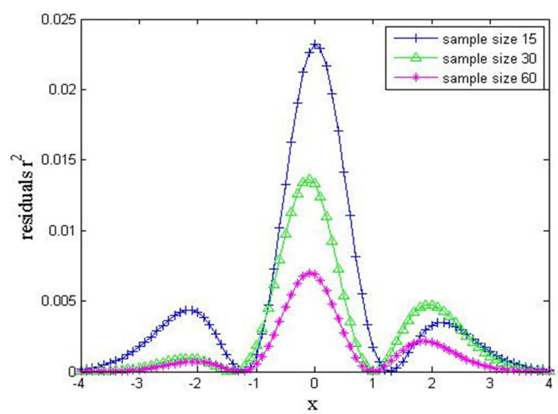

Figure 2 The probability density residuals of different sample sizes based on bootstrap

Figure 1 and figure 2 show: for the construction of standard normal probability density function with bootstrap method, the accuracy of the constructed sample density function has a positive correlation with sample size. However, the disadvantage of this approach is that whether the statistics is choiced in an appropriate way will affect the analysis for overall distribution of the samples.

3. MAXIMUM ENTROPY

3.1 Methods and Principles

The maximum entropy method is a reasoning view proposed by E.T.Jaynes. To infer the system state in the case of only grasping partial information, the legitimate state should be the state where entropy is the largest under the constraints, and it is the only impartial choice. In the case of only measurement data samples, there is no enough reason to choose some analytical distribution function.The form and parameters of measured distribution can be determined without bias by the method of maximum entropy. Majid Asadi(Majid Asadi et al.2014) introduces two new maximum entropy (ME) methods for modeling the distribution of time to an event.

The main idea of maximum entropy is:

Information entropy of probability density function $f(x)$ of random variable x is defined as follows：

$$H(x) = -\int_R f(x)\ln f(x)dx\,, \tag{4}$$

R for integral space

Given the objective function:

$$\max H = -\int_R f(x)\ln f(x)dx \tag{5}$$

Constraints for：

$$s.t.\begin{cases} \int_R f(x)dx = 1 \\ \int_R x^i f(x)dx = m_i,\ i = 1,2,\ldots,n \end{cases} \tag{6}$$

Where m is the order of the moment, m_i represents the i-order origin moment.

To obtain the maximum value of the above objective function under constraints, the Lagrange multiplier method is used to solve.

Set $\bar{H}$ as Lagrange function,

$$\bar{H}=H(x)+(\lambda_0+1)[\int_R f(x)dx-1]+\sum_{i=1}^{n}\lambda_i[\int_R x^i f(x)dx-m_i] \tag{7}$$

Where $\lambda_0,\lambda_1,\cdots,\lambda_n$ are Lagrange multipliers

Set $d\bar{H}/df(x)=0$,

$$-\int_R[\ln f(x)+1]dx+(\lambda_0+1)\int_R dx+\sum_{i=1}^{n}\lambda_i[\int_R x^i dx]=0 \tag{8}$$

To solve:

$$f(x) = \exp(\lambda_0 + \sum_{i=1}^{n}\lambda_i x^i) \tag{9}$$

Equation (9) is the analytic form of maximum entropy probability density function.

To solve the parameters λ_i in (9).

Equation(9) is substituted into the equations

$$\int_R f(x)dx = 1 \tag{10}$$

$$\int_R \exp(\lambda_0 + \sum_{i=1}^{n}\lambda_i x^i)dx = 1 \tag{11}$$

$$e^{-\lambda_0} = \int_R \exp(\sum_{i=1}^{n}\lambda_i x^i)dx \tag{12}$$

$$\lambda_0 = -\ln(\int_R \exp(\sum_{i=1}^{n}\lambda_i x^i)dx) \tag{13}$$

Equation (12) is differentiated to λ_i,

$$\frac{\partial\lambda_0}{\partial\lambda_i} = -\int_R x^i\exp(\lambda_0 + \sum_{i=1}^{n}\lambda_i x^i)dx = -m_i \tag{14}$$

Equation (13) is differentiated to λ_i

$$\frac{\partial\lambda_0}{\partial\lambda_1} = -\frac{\int_R x^i\exp(\sum_{i=1}^{n}\lambda_i x^i)dx}{\int_R \exp(\sum_{i=1}^{n}\lambda_i x^i)dx} \tag{15}$$

Simultaneous equation (14) and (15),

$$m_i = \frac{\int_R x^i\exp(\sum_{i=1}^{n}\lambda_i x^i)dx}{\int_R \exp(\sum_{i=1}^{n}\lambda_i x^i)dx} \tag{16}$$

To solve the optimal values satisfied (16)

it is converted into the following form:

$$R_i = 1 - \frac{\int_R x^i\exp(\sum_{i=1}^{n}\lambda_i x^i)dx}{m_i\int_R \exp(\sum_{i=1}^{n}\lambda_i x^i)dx} \tag{17}$$

$$R = \sum_{i=1}^{n} R_i^2 \to \min \tag{18}$$

At this point, the problem is converted to an unconditional extremum problem, and in this paper, the nonlinear quadratic programming method is taken to sovle this matter:

1) Given the initial solution λ_i;

2) Search direction d_k is determined under established rules ;

3) To determine the search compensation a_k;

4) Set $x_{k+1} = x_k + a_k d_k$

If it meets the termination conditions, stop the iteration and get the approximate optimal solution x_{k+1}; Otherwise, repeat the above steps.

3.2 Analysis of Examples

To verify the accuracy of the probability density function by maximum entropy method, three subsamples(sample size:15、30、60, the same as 1.1) are in turn selected from the overall sample of standard normal distribution. The results are analyzed and compared with standard normal probability density function as follows:

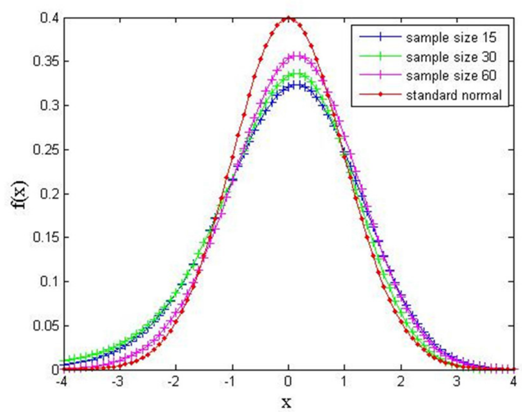

Figure 3 The probability density functions of different sample sizes based on maximum entropy

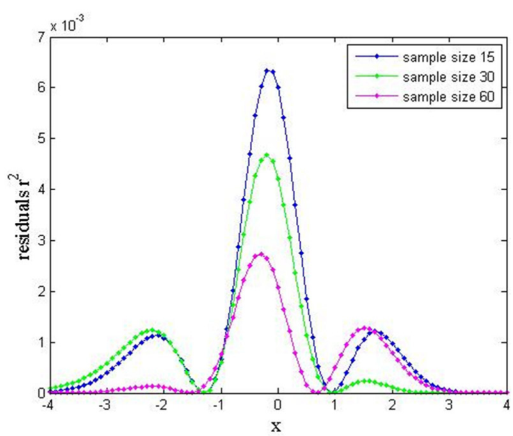

Figure 4 The probability density residuals of different sample sizes based on maximum entropy

Figure 3 and figure 4 show: for the construction of standard normal probability density function with maximum entropy method, the accuracy of the constructed sample density function also has a positive correlation with sample size. The shortage of this approach is that it would be not accurate enough to analyze overall sample probability distribution only with the small samples directly.

4. A NEW JOINT METHOD BASED ON BOOTSTRAP AND MAXIMUM ENTROPY

4.1 Method Description

In order to construct the optimal probability density function under the small-sample condition, a new joint method based on bootstrap and maximum entropy is proposed. It aims to get a more accurate probability density function under the situation of small sample size and unknown priori probability distribution. It can improve the accuracy of maximum entropy method after reasonable expansion for small sample by bootstrap method. Maximum entropy method can eliminate the dependence on statistics to construct polynomial expression for probability density distribution function.The steps are as follows:

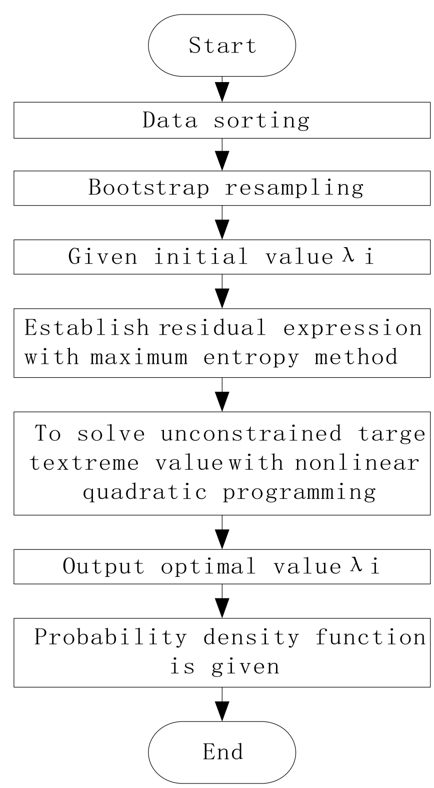

Figure 5 Analysis flowchart based on bootstrap and maximum entropy method

4.2 Analysis of Examples

To verify the accuracy of the probability density function by bootstrap and maximum entropy, three subsamples which are the same as the subsamples in 1.1 are in turn selected from the overall sample of standard normal distribution. The results are analyzed and compared with standard normal probability density function as follows:

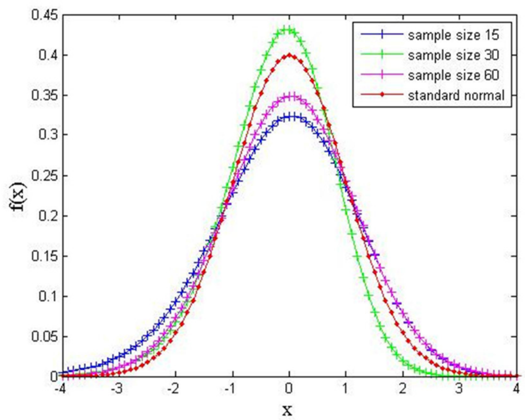

Figure 6 The probability density functions of different sample sizes based on bootstrap and maximum entropy

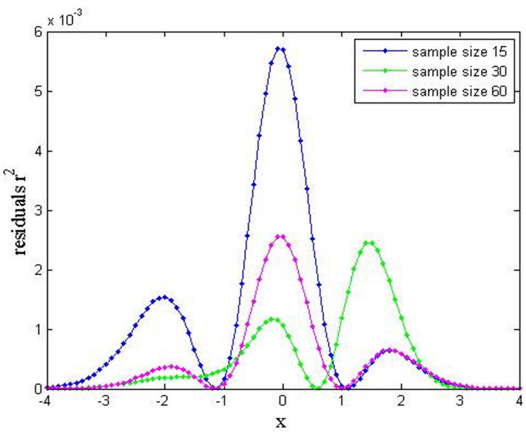

Figure 7 The probability density residuals of different sample sizes based on bootstrap and maximum entropy

Figure 6 and Figure 7 show: the accuracy of the constructed sample density function become more accurate with the sample size increasing.

Table 1 Cumulative residuals

	Bootstrap	Maximum Entropy	Bootstrap& Maximum Entropy
Sample size 15	0.3710	0.0921	0.0749
Sample size 30	0.2251	0.0669	0.0596
Sample size 60	0.1114	0.0429	0.0366

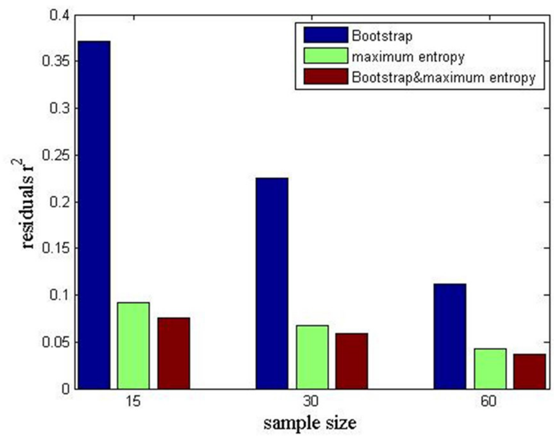

Figure 8 Contrast analysis diagram of cumulative residual under different methods and different sample sizes

From table 1 and figure 8 it can be shown: using above three approaches for the different sample sizes, the accuracy of the analysis is being improved and more accurate with the increasing sample size; However, for analyzing unknown probability distribution of the small sample size(less than 30), the accuracy of the joint method based on bootstrap and maximum entropy which is proposed in this paper is higher. The effect of the new method is obvious and outstanding, especially in the analysis of the small sample problem. From all of the above analysis, the new joint method in this article has its advantages and feasibility.

5. CASE APPLICATION

These are the product lifetime data of some large power equipment enterprise nearly the last five years in table 2, this is just a sample of 12 sets of data.

Table 2 The product lifetime data of an enterprise

Product Name	Commissioning date	Failure date	Lifetime (day)
Equipment 1	2009-07-19	2009-11-06	110
Equipment 2	2008-07-03	2009-11-15	500
Equipment 3	2010-05-26	2010-08-31	97
Equipment 4	2007-06-04	2010-12-08	1283
Equipment 5	2007-06-14	2011-03-30	1385
Equipment 6	2008-01-31	2011-02-20	1116

Equipment 7	2008-09-27	2012-03-05	1255
Equipment 8	2008-04-01	2012-04-10	1470
Equipment 9	2010-03-23	2012-06-12	812
Equipment10	2010-07-21	2012-07-20	730
Equipment11	2007-11-19	2012-09-13	1760
Equipment12	2010-03-15	2013-03-25	1106

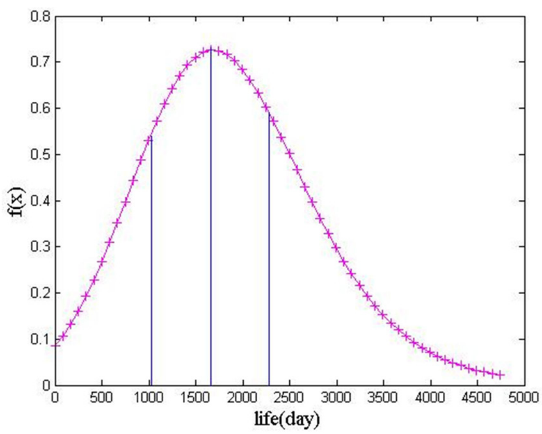

Figure 9 The actual case of probability density distribution of an enterprise product lifetime

Figure 9 is the he actual case of probability density distribution of an enterprise product lifetime using above joint method. It is shown that the point life=1663(d) is the extreme point of probability density curve, therefore it can be made a correlation analysis of the actual products lifetime. For example, if given the significance level α=0.5,the confidence interval on the confidence level of 50% should be in [1030 2277], and it can provide appropriate and better guidance or suggestion for inventory.

6. CONCLUSION

The small-sample probability density distribution is discussed and studied in this paper, respectively using three methods: bootstrap, maximum entropy and the new joint method based on bootstrap and maximum entropy. The results show that: when there is no priori knowledge and the sample size is less than 30 in analysis of small-sample probability density function for large power equipment product lifetime, the new joint method can be greatly close to the real distribution of overall samples. So that the joint method can provide a better operational guidance for engineering practice.

REFERENCES

Lv Zhen, Yao Weixing. (2008). Determining Fatigue Life Distribution with Small Samples [J]. Chinese Journal of Applied Mechanics, v 25, n 2.

Majid Asadi, Nader Ebrahimi, Ehsan S. Soofi, Somayeh Zarezadeh. (2014). New Maximum Entropy Methods for Modeling Lifetime Distributions. Naval Research Logistics, v 61, n 6, p 427-434.

Ouysse, Rachida. (2013). A fast iterated bootstrap procedure for approximating the small-sample bias. Communications in Statistics: Simulation and Computation, v 42, n 7, p 1472-1494.

Radu Florescu, PhD,and Nonel Thirer, PhD. (2006). Distribution Laws Of Small Size Sample0 Metrological Implementation. IEEE 24th Convention of Electrical and Electronics Engineers in Israel.

Srinivasan,V.S. (2004).Estimation of Probability Density Function of a Random Variable for Small Samples. Defence Science Journal, v 2, n 1, p 17-19.

Wang Yanqing, Wang Zhongyu, Sun Jianyong, Zhang,Jianjun,Zissimos.(2014).MourelatosGray bootstr -ap method for estimating frequency-varying random vibration signals with small samples. Chinese Journal of Aeronautics, v 27, n 2, p 383-389

FOUNDATION ITEM：
Project supported by the National Key Technology R&D Program(No. 2012BAF12B04).