Monte Carlo Simulation Based Structural Part Life Estimation According to Small Sample of Right-censored Data

Cheng-gang LI, Li-yang XIE* and Ning-xiang WU

MoE Key Laboratory of Vibration and Control of Aero-Propulsion System, Northeastern University, Shenyang, 110819 China

*Corresponding author

Keywords: Life distribution, Log-normal distribution, Right-censored data, Small size of sample.

Abstract. Life estimation normally requires large size sample of life data. When there is only right-censored life data and the sample size is small, additional assumption or special technique is required. The present paper applies Monte Carlo simulation to generate life samples for a given life distribution, by that a conservative relationship between the population percentile corresponding to $6\sigma$ and the minimum observation in a sample of size $n$ is characterized. Thereby, $(\mu-3\sigma)$ is estimated according to several right-censored data by conservatively taking the censoring time as the minimal observation in a complete life sample of the same size. Such a method is applied to the log-normal distribution of structural part life providing a known variation coefficient ($\sigma/\mu$). The variation coefficient of a particular type of structural part is assigned according to engineering experience.

Introduction

Structural part service life or remaining life estimation by means of parameter estimation method has to be performed based on a specific probability density function. Log-normal distribution use to be applied to describe the fatigue life of most of the structural parts. To estimate structural part life statistically, a certain number of experimental or service life data are necessary. When large sample of complete life data is available, life estimation can be performed conventionally. However, if the sample size is small or only censoring life data are available, life estimation will be very difficult.

In this field, M. Zhao et al [1] proposed an estimator for the mean residual life function with left-truncated and right-censored data, using the properties of the TJW estimator and the functional delta method. S. N. Sengupta [2] considered the problem of unbiased estimation of the mean life and the reliability for an exponential life distribution using time censored sample data and proved that there does not exist an unbiased estimator of the mean life based on a time censored sample.

Nonparametric estimation methods are also well investigated for product life estimation. A. C. McLain et al [3] proposed two nonparametric methods of estimating the conditional mean residual life function. One is a natural extension of the conditional survival model from Gonzalez-Manteiga and Cadarso-Suarez [4]. Based on the comparison between two estimators of the reliability function (i.e. the Kaplan-Meier estimator, and an estimator derived from the empirical MRL function), Y. Shen et al [5] proposed a nonparametric method for the estimation of decreasing mean residual life with type II censored data.

Besides, data-driven methods are more and more developed. T. H. Loutas et al [6] reported a data-driven approach for the remaining useful life estimation of rolling element bearings based on $\varepsilon$-Support Vector Regression. In another paper [7], L. L. Li et al [8] presented a general data-driven, similarity-based approach for residual useful life estimation for industrial components or systems.

Normal Distribution and Monte Carlo Sampling

The normal probability density function is
\[ f(t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{(t-\mu)^2}{2\sigma^2}\right) \]  

(1)

Where, \( t \) stands for life random variable, \( \mu \) stands for the mean, \( \sigma \) stand for the standard deviation.

As to an engineering problem of structural part life estimation, the information available is usually a certain number of life data or right-censored life data. Limited by the experiment or field observation cost and time, a common situation is that only a small size of life sample, or even only a small size of right-censored life sample available. This paper presents a Monte Carlo simulation assistant method of structural part life estimation in the situation that only a small size of right-censored life sample available.

**Small Size Right-censored Life Sample Based Normal Distribution Estimation Method**

Material specimen and product life usually follow log-normal distribution. That is, their logarithms following normal distribution. The present paper presents a method to estimate normal distribution according to small sample right-censored life data. Further, for the majority of structural parts, the variation coefficients, i.e. the ratios of standard deviation to mean value, can be assinged according to that of similar parts or knowledge about material property, manufacturing quality, etc. With such a precondition, the normal distributed log-life of a structural part is characterized by one independent parameter such as the mean, making the life probability distribution estimation easier.

To use right-censored life data to estimate normal distribution, a particular life parameter \( \mu - 3\sigma \), i.e. the percentile corresponding to the failure probability of 0.00135 is estimated, by that the mean and the standard deviation of the normal distribution can be calculated, given a specific variation coefficient.

**Necessary Sample Size for Percentile Estimation**

For a random variable, percentile can be confidently estimated with the minimal observation \( t_{(1,n)} \) if the sample size \( n \) is large enough. To determine the size of a sample by which a percentile can be accurately estimated as the minimal observation, the following median rank estimator of the distribution function \( \hat{F}(t) \) is applied in this paper:

\[ \hat{F}(t_i) = \frac{i}{i + (n+1-i)F_{2(n+1-i),2i,C}} \quad i = 1, 2, ..., n \]

(2)

Where, \( i \) - ordinal number of the life observations ranked from the minimum to the maximum; \( n \) - sample size; \( F_{2(n+1-i),2i,C} \) - F-distribution value corresponding to the upper tail area equaling to \( C \) with the degrees of freedom \( 2(n+1-i) \) and \( 2i \), respectively.

For the situation of \( C=0.5 \), Eq.2 is usually approximately as

\[ \hat{F}(t_{(n)}) \approx \frac{i - 0.3}{n + 0.4} \quad (i = 1, 2, ..., n) \]

(3)

When there are \( n \) observations available in which the minimum equals to \( t_{(1,n)} \), the probability that the product life is less than \( t_{(1,n)} \) equals to \( \hat{F}(t_{(n)}) \). To get the 0.00135×100th percentile, i.e. the life value \( \mu - 3\sigma \), the necessary sample size is

\[ n_{\mu - 3\sigma} = \frac{1 - 0.3}{0.00135} - 0.4 = 518 \]

**Relationship between \( (\mu - 3\sigma) \) and the Minimal Observations**

If the sample size \( n \) is less than 518, then the minimal observation \( t_{(1,n)} \) has to be reduced to approximately represent \( (\mu - 3\sigma) \). The difference between \( (\mu - 3\sigma) \) and \( t_{(1,n)} \) depends on both the mean and standard deviation of a specific normal distribution, a typical situation is shown in Fig.1.
To detect the difference between ($\mu$-3$\sigma$) and $t_{1,n}$ for a specific normal distribution, Monte Carlo sampling is applied to generate simulated observations of the random life. Shown in the following graphs are some of the Monte Carlo simulation results of the minimal observations in different size of samples from different normal distributions. It is shown that, for the relatively smaller sample size, all the minimal observations are less than ($\mu$-3$\sigma$).

In these graphs, the ten points with the abscissa of three are the minimal observations from the ten times of sampling with the sample size of three, respectively. For this situation, 3 observations ($t_1$, $t_2$, $t_3$) are randomly drawn from the normal distributed population $N(\mu,\sigma^2)$ in every trial, and the minimums $t_{1,3}$ is drawn in the graphs. The simulation results clearly show a tendency that when the sample size becomes small, the minimal observation may be much greater than ($\mu$-3$\sigma$), while when the sample size is large enough, the minimal observation will approach to ($\mu$-3$\sigma$). According to the
simulation results, a function can be selected to characterize the difference between \( \mu - 3\sigma \) and the minimal observation in a sample of size \( n \).

Basically, the decrement is determined by both the sample size and the scatter degree of the normal distribution, that is jointly determined by its mean \( \mu \) and standard deviation \( \sigma \). Considering stochastic issue of the Monte Carlo sampling, a conservative function should be applied to cover the dispersion of the majority of observations.

First, the upper boundary of the minimal observations from the different size of samples is described by an appropriate function. By trial and error, the following function is selected:

\[
T_{(1,n)} = (\mu - 3\sigma) + 50\nu[1 - \left(\frac{n}{n_{0.0135}}\right)^{0.5}]^{10}
\]

Where, \( \nu \) stands for the deviation ratio \( \sigma/\mu \).

Eq.4 is drawn in Fig.4 together with the Monte Carlo simulation results, i.e. the minimal observations. The simulation results illustrate that this function can well fit the relationship between \((\mu - 3\sigma)\) and the minimal observations in the samples with the sample size around 5.

When there are \( n \) life samples, \((\mu - 3\sigma)\) can be estimated from Eq.4 as

\[
(\mu - 3\sigma) = t_{(1,n)} - 50\nu[1 - \left(\frac{n}{n_{0.0135}}\right)^{0.5}]^{10}
\]

Normal Distribution Estimation According to Small Sample of Right-censored Life Data

The situation is that, there are five right-censored life data for a particular structural part, the censoring time is \( 1.0 \times 10^6 \) stress cycles. To estimate its life distribution, first it is assumed that the fatigue life of the structural part follows log-normal distribution, i.e. its logarithm follows normal distribution. Further, according to engineering experience, it is provided that the deviation ratio of normal distribution of the logarithmic life equals to 0.05.

With the input information of \( n=5 \), \( t^* = \log(1.0 \times 10^6) = 6 \), and \( \nu = 0.05 \), the pertinent \((\mu - 3\sigma)\) can be estimated as

\[
(\mu - 3\sigma) = t^* - 50\nu[1 - \left(\frac{n}{n_{0.0135}}\right)^{0.5}]^{10}
\]
distribution parameter can be estimated as \( \mu = 6.01, \sigma = 0.30 \). The related normal distribution and log-normal distribution are shown in Fig.5.

![Estimated normal distribution and corresponding log-normal distribution](image)

Figure 5. Life distribution curves estimated according to five right-censored data suspended at \( 1.0 \times 10^6 \).

**Conclusions**

A new method is presented to estimate the distribution parameters of normal or log-normal probability density function to describe structural part life distribution in the situation of vary small sample of right-censored life data. With the precondition of known variation ratio \( \sigma / \mu \) for a normal distribution, the distribution parameters, i.e. \( \mu \) and \( \sigma \) are estimated according to several, such as five right-censored life data. The normal distribution parameter estimation model is developed based on Monte Carlo simulation results, that provide the basis to characterize the relationship between the minimal observation and the percentile \( (\mu - 3\sigma) \) for arbitrary sample size \( n \). Conservatively assuming the right-censored life time as the minimal observation, life distribution can be estimated according the right-censored life data.

**Acknowledgement**

This research is subsidized by the Natural Science Foundation of China (Grant No. U1708255).

**References**

[1] M. Zhao, H. Jiang, X. Liu, A note on estimation of the mean residual life function with left-truncated and right-censored data, Statistics and Probability Letters 83 (2013) 2332–2336.

[2] S. N. Sengupta, B. K. Sinha, Estimation of mean life and reliability for exponential life distribution using time censored sample data, Statistics and Probability Letters 79 (2009) 899–906.

[3] A. C. McLain, S. K. Ghosh, Nonparametric estimation of the conditional mean residual life function with censored data, Lifetime Data Anal 17 (2011) 514-532.

[4] W. Gonzalez-Manteiga, C. Cadarso-Suarez, Asymptotic properties of a generalized Kaplan–Meier estimator with some applications, J Nonparametr Stat. 4 (1994) 65-78.

[5] Y. Shen, M. Xie, and L. C. Tang, Nonparametric Estimation of Decreasing Mean Residual Life With Type II Censored Data, IEEE Transactions on Reliability, 59 (2010) 38-44.

[6] T. H. Loutas, D. Roulias, Remaining Useful Life Estimation in Rolling Bearings Utilizing Data-Driven Probabilistic E-Support Vectors Regression, IEEE Tran. on Reli, 62 (2013) 821-832.

[7] Mohsen Mahoor, Amin Khodaei, Data Fusion and Machine Learning Integration for Transformer Loss of Life Estimation. University of Denver, 2017.

[8] L. L. Li, D. J. Ma and Z. G. Li, Residual Useful Life Estimation by a Data-Driven Similarity-Based Approach, Qual. Reliab. Engng. Int. 33 (2017) 231-239.