Point Estimation Method of Electromagnetic Flowmeters
Life Based on Randomly Censored Failure Data

Zhen Zhou, Wenzhi Jiang, Yongchao Guan, Bin Niu, Dezhong Ma
School of Measurement Technology and Communication Engineering, Harbin University
of Science and Technology, Harbin, China
Tel.: +86 13009861061, fax: +86-0451-86392366
E-mail: zhzh49@126.com

Received: 18 June 2014   /Accepted: 31 July 2014   /Published: 31 August 2014

Abstract: This paper analyzes the characteristics of the enterprise after-sale service records for field failure data, and summarizes the types of field data. Maximum likelihood estimation method and the least squares method are presented for the complexity and difficulty of field failure data processing, and Monte Carlo simulation method is proposed. Monte Carlo simulation, the relatively simple calculation method, is an effective method, whose result is closed to that of the other two methods. Through the after-sale service records analysis of a specific electromagnetic flowmeter enterprises, this paper illustrates the effectiveness of field failure data processing methods.

Keywords: Failure data, Reliability, Monte Carlo simulation, Estimation method, Product life.

1. Introduction

Reliability data is the basis of reliability analysis and evaluation. The products reliability data from survey and collection could serve as product reliability evaluation and failure analysis data samples. Then feeding back the results to the design, manufacturing and management, and we’ll ultimately achieve the purpose of improving product reliability, and provide important information for future new product design [1-2]. The enterprise products service record contains a large number of product field failure data. These data help to get the reliability level of the working products under field conditions, and help to analyze the failure phase, and the occurrence site. These data are important information for failure mode analysis and reliability assessment.

Some field data are from the service center, and some are from the users sample survey. In order to reflect the actual situation more accurately, the data should be true and reliable, but the professionalism of the site personnel recording, job responsibility, the ability to identify failure modes, the design of sample table, normative industry terminology, or the incomplete recorded information, would affect the authenticity of the data field, and therefore pose difficulty and complexity to the field data processing [3-4]. But to really understand the reliability level of working products under field conditions, we must first analyze the reliability of field data [5-6].

2. Characteristics of Field Failure Data

Now we mostly apply laboratory or field data to the assessment of reliability indicators. The statistical analysis of field data is discussed in the article, since the application of field data can be fully in line with actual conditions of use, save a lot of materials, energy, labor and other costs and show significant economic benefits [7]. Field data shows the following characteristics.
1) The sources of information are very complex. The same type of product can be used in different fields, due to different operators, different environments, and different occurrence time, reasons, or positions, producing inaccurate or incomplete records.

2) The starting time is different from each other. Due to the different times units purchasing products, the starting moment of use differs, and therefore they would not work simultaneously as simulation under laboratory conditions, even for the same product.

3) Some products, though with no failure, lose information midway for various reasons. Under the actual conditions of use, the product information will be lost midway.

4) When the statistical ends, some products are failed, but some are not.

### 3. Several Types of Field Failure Data

Product field data are irregularly censored in incomplete life test, including the following basic conditions.

1) Fixed number censoring data. Let \( n \) products which are independent of the same type start timing at the same time \( t = 0 \), and terminate at the moment the \( r \)th product fails. In this way, the product life data of \( r \) units ahead are available: \( T_0 \leq T_1 \leq \ldots \leq T_r \), and that is censored data.

2) Fixed time censoring data. Similar to censored, the number of failure at a predetermined time \( t_0 \) is a random variable. If \( r \) units of product are failed when observation is terminated, then we could get the data \( T_0 \leq T_1 \leq \ldots \leq T_r \leq t_0 \), that is time censored situation.

3) Randomly censored data [4]. Product reliability data on-site basically belongs to the randomly censored data, which is composed of interrupted observational data and fault data randomly.

Suppose the life of field product is \( T_1, T_2, \ldots, T_n \), which is independent and identically distributed on the distribution function \( F(t; \theta) \) or its density function \( f(t; \theta) \) of its overall \( T \), \( \theta \) is its unknown parameter. And suppose the censored time \( \{L_i\} \) of corresponding product is independent, and there is the distribution function \( \{G_i(t), i = 1, 2, \ldots, n\} \) or density function \( \{g_i(t), i = 1, 2, \ldots, n\} \), assuming that the distribution of the censored time is unrelated to the unknown parameters. Also suppose \( \{T_i\} \) and \( \{L_i\} \) are independent. If introducing

\[
X_i = \min \{T_i, L_i\} (i = 1, 2, \ldots, n)
\]

we can get "random variables" (or two-dimensional random variable), that is

\[
(X_i, \delta) \quad (i = 1, 2, \ldots, n),
\]

indicating whether the product is censored. Thus, product censored time is a random variable, that is, random censoring.

4) Doubly censored data. Suppose the product life data on-site \( T_1, T_2, \ldots, T_n \) are independent and identically distributed, \( t_{(1)} \leq t_{(2)} \leq \ldots \leq t_{(n)} \) are their observations, assuming samples available as part of the interception data observational, which means intercepting related information in the middle period of working time, with both start time and down time unknown, which includes the left and right censored, double censored.

### 4. Analysis of Field Failure Data

Point and interval estimation methods can be used in parameter estimation of distribution model in reliability engineering. By this means, the unknown sample is given closed to the true value of the parameter by observed samples. There are many methods of point estimates, such as the probability plots estimate method [8], least squares [9-10], and maximum likelihood estimation method [11-12]. Probability plots estimate method is simple and easy, one can estimate all the parameters needed in the probability paper, convenient for the majority of engineering and technical personnel to grasp, which is widely used. But there are too many human factors in drawing, the least squares method, and other quantitative estimation methods eliminate this drawback. The precision of maximum likelihood estimation method is high, but with a large amount of calculation and complexity. The best linear unbiased estimation and best linear invariant estimation must rely on special forms, which are dedicated and large, so they can’t be used without the table [13]. Besides, least squares estimation fit all kinds of conditions, both complete - samples and incomplete samples, both censored and time censored situation. Currently, the Monte Carlo simulation method is also a popular parameter estimation method [14-16]. This article gives three methods of parameter estimation: the maximum likelihood estimation, least squares estimation, and Monte Carlo simulation.

#### 4.1. The Maximum Likelihood Estimation Method

Suppose the joint density function of the sample \( X_1, X_2, \ldots, X_n \) is \( f(x; \theta) \), where parameter \( \theta \) is unknown, for a given \( x \), we name
\[ L(\theta, x) = f(x; \theta) \]  
for the likelihood function parameters. Accordingly, we call \( \ln L(\theta, x) \) the Logarithmic likelihood function of parameter \( \theta \).

If there is statistical quantity \( \hat{\theta} \), which can make
\[ L(\hat{\theta}, x) = \max_{\theta} L(\theta, x), \] (5)
we call \( \hat{\theta} \) as the maximum likelihood estimator (MLE) of \( \theta \).

Generally, the common method to find the MLE of parameter \( \theta \) is to solve the likelihood equation:
\[ \frac{\partial \ln L(\theta, x)}{\partial \theta_i} = 0, \quad i = 1, 2, \ldots, k, \] (6)
where \( k \) is the number of unknown parameter \( \theta \).

### 4.2. Least Squares Estimation Method

Assuming a linear equation as:
\[ Y = b_0 + b_1 x, \] (7)
its residual sum of squares is
\[ Q(b_0, b_1) = \sum_{i=1}^{n} (Y_i - b_0 - b_1 x)^2, \] (8)
If
\[ \delta = \begin{cases} \frac{\partial \theta}{\partial b_0} = 0 \\ \frac{\partial \theta}{\partial b_1} = 0 \end{cases} \] (9)
we obtain this
\[ \begin{cases} b_1 = \frac{r \sum_{i=1}^{r} X_i Y_i - \sum_{i=1}^{r} X_i \sum_{i=1}^{r} Y_i}{r \sum_{i=1}^{r} X_i^2 - (\sum_{i=1}^{r} X_i)^2} \\ b_0 = \frac{1}{r} (\sum_{i=1}^{r} Y_i - b_1 \sum_{i=1}^{r} X_i) \end{cases} \] (10)
where \( r \) is the number of failure data, and \( X_i, Y_i \) is the value corresponding to scatter. Take exponential distribution as an example, as
\[ F(t) = 1 - e^{-\lambda t - \gamma}, \] (11)
after the conversion, we can get
\[ \ln \frac{1}{1 - F(t)} = \lambda t - \mu \] (12)
If \( Y = \ln \frac{1}{1 - F(t)} \), \( x = t \), \( Y \) and \( X \) keeping a linear relationship, we could obtain that \( \lambda = b_1, \gamma = -\frac{b_0}{b_1} \).

Thus the average life expectancy of the product is:
\[ m = \frac{1}{\lambda} + \gamma \] (13)

### 4.3. Monte Carlo Simulation Method

If the product life \( T \) subjects to two-parameter exponential distribution, its distribution function is
\[ F_T(t) = P[T \leq t] = 1 - e^{-\lambda(t - \gamma)}, \] (14)
where \( \lambda, \gamma \) are for the scale parameter and position parameters respectively, besides, \( \lambda > 0, \gamma > 0, t > 0 \). Taking \( T \) on transformation \( X = t \), the distribution function of the transformation on \( X \) is
\[ F_X(x) = P[X < x] = 1 - e^{-b_1 x + b_0}, \] (15)
where \( \lambda = b_1, \gamma = -\frac{b_0}{b_1} \). And then the \( X \) linear transformation is:
\[ Y = b_1 X + b_0 \] (16)
So, the distribution function for converting \( Y \) is
\[ F_Y(y) = P[(b_1 X + b_0) < y] = 1 - e^{-y}, \] (17)

Truncating test \((n, r)\) with fixed number \( r \), the sample of \( T \) was \( T_i, i = 1, 2, \ldots, r, r \leq n \), thereby obtaining the order statistics \( T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(r)} \).

Calculate the average and covariance of the above equation,
\[ E(Y_{(i)}) = b_1 E(X_{(i)}) + b_0 \]
\[ \text{cov}(Y_{(i)}, Y_{(j)}) = b_1^2 \text{cov}(X_{(i)}, X_{(j)}) \] (18)

Monte Carlo stochastic simulation method gives the calculated value. According to formula (17), assigning
\[ U = F_Y(y) = 1 - e^{-y} \] (19)
we could get to know that \( U \) obeys uniform distribution on the interval \( U(0,1) \).
Then anti-solving $Y = \ln \frac{1}{1 - F(t)}$ according to formula (19). In this case $Y$ is given by the expression of $U$. It can first be computer-generated uniformly distributed random sample $U_i$, sorting of $U_{(i)}$, on the interval (0,1), besides, giving the corresponding $Y_i$, and $Y_{(i)}$, finally we can find the corresponding parameter. Monte Carlo simulation algorithm is as follows:

For the given $n$, $r$, and simulation times $N$, $N$ sets of samples may be generated. Repeating the following steps a) ~ b) would help to generate $N$ sets of analog sample values $\{Y_{(i),d=1,2,\ldots,r,j=1,2,\ldots,N}\}$:

a) Generating $n$ random variable samples $U_i$ subjected to uniform distribution $U(0,1)$;

b) According to $Y = \ln \frac{1}{1 - U}$, calculate a set of samples of random variables $Y_1, Y_2, \ldots, Y_n$;

c) By sorting, order statistics would be obtained $Y_{(1)}, Y_{(2)}, \ldots, Y_{(r)}, \ldots, Y_{(n)}$;

d) Fixed data censoring, obtain the samples $\{Y_{(i),j=1,2,\ldots,n}\}$.

5. The Example Analysis

Taking valid samples selected in a certain type of electromagnetic flowmeters after-sale service records in an enterprise for analysis, the sample data are shown in Table 1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Sales volume /sets</th>
<th>Sales date</th>
<th>Rework number /sets</th>
<th>Rework Date</th>
<th>Life time /days</th>
<th>Packet No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>2011-1-25</td>
<td>1</td>
<td>2011-9-24</td>
<td>242</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>2011-1-30</td>
<td>1</td>
<td>2011-6-20</td>
<td>141</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>2011-2-26</td>
<td>1</td>
<td>2011-8-8</td>
<td>163</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2011-3-30</td>
<td>1</td>
<td>2011-6-27</td>
<td>89</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>2011-3-30</td>
<td>1</td>
<td>2011-7-3</td>
<td>95</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>2011-4-20</td>
<td>1</td>
<td>2011-10-23</td>
<td>186</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>42</td>
<td>2011-4-30</td>
<td>1</td>
<td>2011-8-20</td>
<td>112</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>42</td>
<td>2011-4-30</td>
<td>1</td>
<td>2011-9-7</td>
<td>130</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>2011-5-20</td>
<td>1</td>
<td>2011-7-2</td>
<td>43</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>2011-6-28</td>
<td>1</td>
<td>2011-9-19</td>
<td>83</td>
<td>8</td>
</tr>
</tbody>
</table>

5.1. An Example for Maximum Likelihood Estimation

Using Weibull++7.0 reliability analysis software, and maximum likelihood estimation method for the above sample data analysis, we could obtain that two-parameter exponential distribution is the best distribution. The distribution of exponential probability is shown in Fig. 1.

5.2. An Example for Least Squares Estimation Method

According to formula (8), the residual sum of squares is:

$$Q(b_0, b_1) = 0.000339$$

According to formula (9), the value of $b_0$ and $b_1$ are:

$$b_0 = -0.01566, b_1 = 4.05 \times 10^{-4}$$

So, $\lambda = b_1 = 4.05 \times 10^{-4}, \gamma = -\frac{b_0}{b_1} = 39$ (days)

And the average life expectancy of the product is:

$$MTBF = \frac{1}{\lambda} + \gamma = 2508$$ (days)

5.3. An Example for Monte Carlo Simulation Method

Electromagnetic flowmeter maintenance data could be viewed as resulting from the number fixed
censored program \((n,r) = (150,10)\), and the samples are \(T: 43, 83, 89, 95, 112, 130, 141, 163, 186, 242\), (unit: day). Assuming that electromagnetic flowmeter subjects to two-parameter exponential distribution, with Table 2 showing the results of the three methods, we can see that the Monte Carlo simulation method \((N = 1000)\), is also an effective method, and the results of it is close to the other two methods of results.

<table>
<thead>
<tr>
<th>Method</th>
<th>(\lambda)</th>
<th>(Y)</th>
<th>MTBF (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum likelihood estimation method</td>
<td>0.000400</td>
<td>43</td>
<td>2543</td>
</tr>
<tr>
<td>Least squares estimation method</td>
<td>0.000405</td>
<td>39</td>
<td>2508</td>
</tr>
<tr>
<td>Monte Carlo simulation method</td>
<td>0.000368</td>
<td>25</td>
<td>2742</td>
</tr>
</tbody>
</table>

6. Conclusions

This paper presents several effective reliability analysis methods for the complexity and difficulty of field failure data processing, which are verified by an electromagnetic flowmeter sale data. The results show that the maximum likelihood estimation and least squares estimation results of the assessment are closed. And although the Monte Carlo simulation method has its own advantages, the results have a little bias, indicating that this assessment still needs to be improved. Through the study of these three assessment methods, we initially achieve that the lifetime of such electromagnetic flowmeter products subjects to two-parameter exponential distribution. This conclusion can further provide the basis for reliability life test and reliability growth test.

Acknowledgements

The work was supported by Natural Science Foundation of Heilongjiang Province of China (F201305) and National High-tech R&D Program of China (863 Program) (2008AA042207).

References