Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Application of the Bayes' theory to the failure rate evaluation of electrical apparatus

JIN Shao-hua†, LU Jian-guo, WAN Yan-ping, SUN Shu-guang
(*Department of Mathematics, Hebei University of Technology, Tianjin 300130, China*)
†E-mail: jin_0510@126.com
Received Dec. 19, 2006; revision accepted Jan. 5, 2007

Abstract: The Weibull distribution has been widely used in reliability fields. A mixed Weibull distribution represents a population that consists of several Weibull subpopulations. In this paper, a new approach which combines the least-squares method with Bayes' theorem, takes advantage of the parameter estimation for single Weibull distribution is developed to estimate the parameters of each subpopulation. The estimates given by this paper also satisfy the maximum likelihood equation. The estimates of the failure rate of the mixed Weibull population are given. An actual test data is computed by using the proposed method. The Kolmogorov-Smirnov goodness-of-fit test turns out that the proposed method yields more accurate result.

Key words: Bayesian estimation, Failure rate, Least-squares method, Maximum likelihood principle, Mixed-Weibull distribution
doi:10.1631/jzus.2007.A0501 **Document code:** A **CLC number:** TM7321

INTRODUCTION

In reliability engineering, it is known that electrical equipment usually has more than one failure mode (Liu, 2002a; 2002b). It has been recognized for more than five decades that the mixed Weibull distribution (Weibull, 1951) is an appropriate distribution to use in modeling the lifetimes of the units that have more than one failure cause. However, due to the lack of a systematic statistical procedure for fitting an appropriate distribution to such a mixed data set, it has not been widely used. A mixed Weibull distribution represents a population that consists of several Weibull subpopulations. The reliability estimation for a mixed distribution is much more difficult than that for a single distribution. The difficulties are partly caused by the involvement of more unknown parameters.

In this paper, a new approach which combines the least-squares method with Bayesian theorem, takes advantage of the parameter estimation for the single Weibull distribution is developed to estimate the parameters of each subpopulation. The estimates given by this paper also satisfy the maximum likelihood equation. The estimates of the failure rate of the mixed Weibull population are given. An actual test data is computed by using the proposed method. The Kolmogorov-Smirnov goodness-of-fit test turns out that the proposed method yields more accurate results.

MIXED WEIBULL DISTRIBUTION

If the population consists of a mixture of m independent subpopulations with no correlation and each subpopulation has its own unique failure mode and distribution, the lifetime distribution for the mixed population can be expressed by

$$f(t) = p_1 f_1(t) + \cdots + p_m f_m(t), \tag{1}$$

where $f(t)$ is probability density function of the mixed population; $f_i(t)$ is probability density function of the ith subpopulation, i=1,2,…,m; p_i is mixing weight of the ith subpopulation, $p_i \in (0,1)$, i=1,2,…,m, and $\sum_{i=1}^{m} p_i = 1$.

If a subpopulation can be described by a single Weibull distribution, i.e.

$$f_i(t)=(\beta_i/\eta_i)(t/\eta_i)^{\beta_i-1}\exp[-(t/\eta_i)^{\beta_i}], \tag{2}$$

Eq.(1) becomes a mixed m-Weibull distribution with $3m-1$ parameters $\beta_1,\eta_1,\beta_2,\eta_2,\ldots,\beta_m,\eta_m,p_1,\ldots,p_{m-1}$.

CURRENT METHODS

Currently, two major estimation methods are used for the mixed Weibull distribution: the graphical method (Mann *et al.*, 1974) and the Maximum Likelihood Estimation (MLE) method (Nelson, 1990). The graphical parameter estimation method is very popular for the mixed Weibull distribution due to its simplicity and visibility. It depends on visual inspection of the data plots, which fails in most well-mixed cases. Also, it is hard to use for small sample sizes, which is the case electrical engineers often encounter.

The MLE estimate has excellent statistical characteristics. It can find simultaneously all parameters that maximize the likelihood function of the observed sample. For a mixed Weibull population, the MLE is very complex. It has to be pointed out that, for small size samples, the MLE estimates tend to be highly biased and should be used very carefully.

ESTIMATION OF THE FAILURE RATE FOR THE MIXED WEIBULL POPULATION

Determination of the subsamples by Bayes' theorem

If a reliability life test is carried out on N units of an electrical product, which has m failure modes, a times-to-failure sample $\{t_i, i=1,2,\ldots,N\}$ is obtained. Assume that $t_1<t_2<\ldots<t_N$.

At time t_i, a failure is obtained. To split this failure point rationally, the concept of belonging probability, $p_j(t_i)$, which is the posterior probability that this failure belongs to the jth subpopulation ($j=1,2,\ldots,m$), is introduced by Kececioglu (1993). By definition

$$p_j(t_i)=p\left\{T\in f_j(t)\,|\,t_i-\Delta t/2<T<t_i+\Delta t/2\right\},\quad j=1,2,\cdots,m;\ i=1,2,\cdots,N. \tag{3}$$

Applying Bayes' theorem, yields

$$p_j(t_i)=p\left\{t_i-\Delta t/2<T<t_i+\Delta t/2\,|\,T\in f_j(t)\right\}\cdot p\left\{T\in f_j(t)\right\}\Big/\left\{\sum_{j=1}^{m}p\left\{t_i-\Delta t/2<T<t_i+\Delta t/2\,|\,T\in f_j(t)\right\}\cdot p\left\{T\in f_j(t)\right\}\right\}, \tag{4}$$

where $p\{T\in f_j(t)\}=p_j$ ($1\le j\le m$).

Then, the probability that a failure occurring at time t_i belongs to subpopulation j ($1\le j\le m$) is

$$p_j(t_i)=\frac{p_jf_j(t_i)}{\sum_{j=1}^{m}p_jf_j(t_i)}=\frac{p_jf_j(t_i)}{f(t_i)},\ i=1,2,\cdots,N. \tag{5}$$

Note that for each failure point, the sum of all belonging probabilities must be unity, i.e. $\sum_{j=1}^{m}p_j(t_i)=1$ ($1\le i\le N$).

By means of the posterior belonging probabilities, the failure occurring at t_i can be divided into m portions: $100p_j(t_i)$ percentage of this failure belongs to subpopulation j ($1\le j\le m$). It can be claimed that there were $p_j(t_i)$ failures expected at time t_i if only subpopulation j was put in test with the sample size of Np_j. In other words, $p_j(t_i)$ can be considered as the failed unit number at time t_i in subpopulation j. Pooling these "fractional failures" versus their corresponding occurrence times under the same subpopulation yields the following m subsamples:

$$i\ (1\le i\le m):\{(t_1,p_i(t_1)),(t_2,p_i(t_2)),\cdots,(t_N,p_i(t_N))\}. \tag{6}$$

Application of the least-squares method

In subsample j ($1\le j\le m$), $p_j(t_1)$ units fail at time t_1, $p_j(t_2)$ units fail at time $t_2,\ldots,p_j(t_N)$ units fail at time t_N. For each subpopulation, its corresponding subsample can be seen as a grouped data sample. For each subsample, the conventional estimation method, for single-Weibull distribution can be used to estimate the parameters of each subpopulation.

The accumulated failure probability of the ith failure in the jth ($1\le j\le m$) subpopulation will be

$$F_j(t_i)=\sum_{K=1}^{i}p_j(t_K)\Big/\sum_{l=1}^{N}p_j(t_l). \tag{7}$$

Therefore, the following m data sets will be obtained:

Subpopulation

$$i\ (1\le i\le m): \{(t_1, F_i(t_1)), (t_2, F_i(t_2)), \cdots, (t_N, F_i(t_N))\}. \quad (8)$$

The unreliability, for a Weibull distribution (Lu and Su, 2003) can be written in the form of

$$\ln\ln\left[\frac{1}{1-F_j(t_i)}\right] = \beta_j \ln t_i - \beta_j \ln \eta_j, \quad (9)$$

or in the linearized form of

$$Y_j(i) = \beta_j X(i) + b_j, \quad (10)$$

where $Y_j(i)=\ln\{-\ln[1-F_j(t_i)]\}$, $X(i)=\ln t_i$, $b_j=-\beta_j \ln\eta_j$.

Applying the least-squares method, the distribution parameters are given by

$$\hat{\beta}_j = \frac{\sum_{i=1}^{N} X(i)Y_j(i) - \frac{1}{N}\left[\sum_{i=1}^{N} X(i)\cdot\sum_{i=1}^{N} Y_j(i)\right]}{\sum_{i=1}^{N} X^2(i) - \frac{1}{N}\left[\sum_{i=1}^{N} X(i)\right]^2}, \quad (11)$$

$$\hat{b}_j = \frac{1}{N}\sum_{i=1}^{N} Y_j(i) - \hat{\beta}_j \frac{1}{N}\sum_{i=1}^{N} X(i),\ 1\le j\le m, \quad (12)$$

$$\hat{\eta}_j = \exp(-\hat{b}_j / \hat{\beta}_j),\ j=1,2,\ldots,m. \quad (13)$$

On the other hand, the posterior mixing weight can be obtained from

$$\hat{p}_j = \frac{1}{N}\sum_{i=1}^{N} p_j(t_i),\ 1\le j\le m. \quad (14)$$

For the mixed m-Weibull distribution, the likelihood function is

$$L = \prod_{i=1}^{N} f(t_i) = \prod_{i=1}^{N}\left[\sum_{j=1}^{m} p_j f_j(t_i)\right], \quad (15)$$

or

$$l = \ln L = \sum_{i=1}^{N}\ln\left[\sum_{j=1}^{m} p_j f_j(t_i)\right]. \quad (16)$$

Taking the partial derivative with respect to p_k, yields

$$\frac{\partial l}{\partial p_k} = \sum_{i=1}^{N}\frac{f_k(t_i)-f_m(t_i)}{f(t_i)},\ 1\le k\le m-1. \quad (17)$$

Substituting Eq.(4) into Eq.(17), yields

$$\frac{\partial l}{\partial p_k} = \sum_{i=1}^{N}\frac{p_k(t_i)}{p_k} - \sum_{i=1}^{N}\frac{p_m(t_i)}{p_m}. \quad (18)$$

Then, substituting Eq.(14) into Eq.(18), yields

$$\frac{\partial l}{\partial p_k} = N\cdot\frac{\hat{p}_k}{\hat{p}_k} - N\cdot\frac{\hat{p}_m}{\hat{p}_m} = 0. \quad (19)$$

By Eq.(19), the estimates given by Eq.(14) satisfy the maximum likelihood equation.

By these two methods, if the population consists of a mixture of two Weibull distributions, the probability density function of the mixed population can be expressed by

$$\begin{aligned} f(t) &= p_1(\beta_1/\eta_1)(t/\eta_1)^{\beta_1-1}\exp\left[-(t/\eta_1)^{\beta_1}\right] \\ &+ (1-p_1)(\beta_2/\eta_2)(t/\eta_2)^{\beta_2-1}\exp\left[-(t/\eta_2)^{\beta_2}\right]. \end{aligned} \quad (20)$$

The corresponding unreliability of Eq.(20) is

$$\begin{aligned} F(t) &= \int_0^t f(t)\mathrm{d}t \\ &= \int_0^t p_1(\beta_1/\eta_1)(t/\eta_1)^{\beta_1-1}\exp\left[-(t/\eta_1)^{\beta_1}\right]\mathrm{d}t \\ &+ \int_0^t (1-p_1)(\beta_2/\eta_2)(t/\eta_2)^{\beta_2-1}\exp\left[-(t/\eta_2)^{\beta_2}\right]\mathrm{d}t. \end{aligned} \quad (21)$$

Letting $x=t^{\beta_1}/\eta_1^{\beta_1}$ in the first integration of Eq.(21), $y=t^{\beta_2}/\eta_2^{\beta_2}$ in the second integration of Eq.(21). We have

$$\begin{aligned} F(t) &= p_1\int_0^{(t/\eta_1)^{\beta_1}} \mathrm{e}^{-x}\mathrm{d}x + (1-p_1)\int_0^{(t/\eta_2)^{\beta_2}} \mathrm{e}^{-y}\mathrm{d}y \\ &= 1 - p_1\exp\left[-(t/\eta_1)^{\beta_1}\right] - (1-p_1)\exp\left[-(t/\eta_2)^{\beta_2}\right]. \end{aligned} \quad (22)$$

The corresponding reliability function is

$$\begin{aligned} R(t) &= 1 - F(t) \\ &= p_1\exp\left[-(t/\eta_1)^{\beta_1}\right] + (1-p_1)\exp\left[-(t/\eta_2)^{\beta_2}\right]. \end{aligned} \quad (23)$$

Finally, the corresponding failure rate function is

$$\lambda(t)=\frac{f(t)}{R(t)}=\left\{p_1(\beta_1/\eta_1)(t/\eta_1)^{\beta_1-1}\exp\left[-(t/\eta_1)^{\beta_1}\right]\right.$$
$$\left.+(1-p_1)(\beta_2/\eta_2)(t/\eta_2)^{\beta_2-1}\exp\left[-(t/\eta_2)^{\beta_2}\right]\right\} \quad (24)$$
$$\Big/\left\{p_1\exp\left[-(t/\eta_1)^{\beta_1}\right]+(1-p_1)\exp\left[-(t/\eta_2)^{\beta_2}\right]\right\}.$$

Similarly, the estimation of the failure rate function for the mixed m-Weibull distribution is

$$\lambda(t)=\left\{p_1(\beta_1/\eta_1)(t/\eta_1)^{\beta_1-1}\exp\left[-(t/\eta_1)^{\beta_1}\right]+\cdots\right.$$
$$\left.+p_m(\beta_m/\eta_m)(t/\eta_m)^{\beta_m-1}\exp\left[-(t/\eta_m)^{\beta_m}\right]\right\} \quad (25)$$
$$\Big/\left\{p_1\mathrm{e}^{-(t/\eta_1)^{\beta_1}}+\cdots+p_m\mathrm{e}^{-(t/\eta_1)^{\beta_m}}\right\}.$$

PROPOSED ALGORITH

The original estimation problem is that of having a data sample from a mixed population, which has a probability density function given in Eqs.(1) and (2), and then of estimating the $3m-1$ parameters, $\beta_1,\eta_1,\beta_2,\eta_2,\ldots,\beta_m,\eta_m,p_1,\ldots,p_{m-1}$, such that the distribution fits the data "best".

In practice, the quality of the data is not always good. However, the idea of the least-squares method is to find the "best" fit. Regardless of how good or bad the quality of the data, the "best" fitting line always exists. According to the least-squares principle, the "best" fitting line minimizes the residual variation around the line. Generally, the correlation coefficient, ρ, provides a good measure of how well the line fits the data. The larger the absolute value of ρ, the better the fitted line. Based on this discussion, the "best" parameter estimates can be obtained by employing the least-squares principle to iterate on the $\beta_1,\eta_1,\beta_2,\eta_2,\ldots,\beta_m,\eta_m,p_1,\ldots,p_{m-1}$ values to maximize the absolute value of ρ.

With any chosen set of $\beta_1,\eta_1,\beta_2,\eta_2,\ldots,\beta_m,\eta_m,p_1,\ldots,p_{m-1}$, the correlation coefficients of ρ_j ($1\le j\le m$) are determined by the following equation:

$$\rho_j=\left\{\sum_{i=1}^{N}X(i)Y_j(i)-\frac{1}{N}\sum_{i=1}^{N}X(i)\cdot\sum_{i=1}^{N}Y_j(i)\right\}$$
$$\Big/\left\{\left\{\sum_{i=1}^{N}X^2(i)-\frac{1}{N}\left[\sum_{i=1}^{N}X(i)\right]^2\right\}\right.$$
$$\left.\cdot\left\{\sum_{i=1}^{N}Y_j^2(i)-\frac{1}{N}\left[\sum_{i=1}^{N}Y_j(i)\right]^2\right\}\right\},\ 1\le j\le m. \quad (26)$$

Since m lines are simultaneously fitted to m subsamples from m-Weibull subpopulations and every parameter has an effect on m correlation coefficients, the sum of the squares of these m correlation coefficients might be the "best" measure for the degree of fitting, i.e.

$$\rho=\sum_{j=1}^{m}\rho_j^2. \quad (27)$$

Therefore, employing the iterative procedure, the estimates of $\beta_1,\eta_1,\beta_2,\eta_2,\ldots,\beta_m,\eta_m,p_1,\ldots,p_{m-1}$, can be obtained by maximizing the value of ρ, starting from the graphical estimates to save search time.

EXAMPLE

Table 1 shows the life test data of ten relays. If the mixed two-Weibull distribution is used to represent the Times to Failure distribution, we can determine the mixed population's parameters using the proposed approach and the graphic method. The Kolmogorov-Smirnov goodness-of-fit test was conducted to compare the results. By using the proposed approach, we can estimate the failure rate of the mixed Weibull population.

Table 1 Failure data from a life test for example

Failure order i	Times to failure
1	3.0
2	28.5
3	71.6
4	91.1
5	129.1
6	157.8
7	188.9
8	226.1
9	278.0
10	367.2

Solutions to the example

1. Graphic estimation

By graphic method, the parameter estimates are $\hat{p}_1=0.3,\ \hat{\beta}_1=0.6,\ \hat{\eta}_1=39.4,\ \hat{\beta}_2=2.3,\ \hat{\eta}_2=233.51.$

2. The proposed estimation

To start the proposed approach, we use the graphic estimates as the initial values of β_1^0, η_1^0, $\beta_2^0, \eta_2^0, p_1^0$. Using the proposed approach yields (the calculation details are omitted here): $\hat{p}_1 = 0.3$, $\hat{\beta}_1 = 0.5$, $\hat{\eta}_1 = 50.1$, $\hat{\beta}_2 = 1.9$, $\hat{\eta}_2 = 193.1$.

3. Comparison

A comparison is made by conducting the Kolmogorov-Smirnov goodness-of-fit test on the parameter values obtained by the graphical method and the proposed approach as follow:

The graphic method yields D_{max}=0.075; the proposed approach yields D_{max}=0.031.

It is easy to see that the proposed approach may yield more accurate parameter estimates.

The estimation of the failure rate function of the mixed Weibull population of the example using the proposed approach is

$$\begin{aligned}\hat{\lambda}(t) = &\Big\{0.003(t/50.1)^{-0.5}\exp\left[-(t/50.1)^{0.5}\right] \\ &+0.007(t/193.1)^{0.9}\exp\left[-(t/193.1)^{1.9}\right]\Big\} \\ &\Big/\Big\{0.3\exp\left[-(t/50.1)^{0.5}\right] + 0.7\exp\left[-(t/193.1)^{1.9}\right]\Big\}.\end{aligned} \tag{28}$$

CONCLUSION

The proposed method makes full use of the information on the distribution's behavior of m subpopulations over the whole test duration. The population sample data are split into subpopulation data sets over the whole test duration by using the posterior belonging probability of each observation to each subpopulation. Then, with the concept of accumulated failure probability, the proposed approach combines the least-squares method with Bayes' theorem, takes advantage of the parameter estimation for single Weibull distribution to each derived subgroup data set, and estimates the parameters of each subpopulation. The estimates given by this paper also satisfy the maximum likelihood equation. The estimates of its failure rate function are given.

References

Kececioglu, D., 1993. Reliability and Life Testing Handbook 1. PTR Prentice Hall, New Jersey.

Liu, B.Z., 2002a. Field survey and analysis on reliability of low voltage electrical apparatus. *Low Voltage Electrical Apparatus*, **2**:11-13.

Liu, B.Z., 2002b. Field survey and analysis on reliability of low voltage electrical apparatus (continued). *Low Voltage Electrical Apparatus*, **3**:14-18.

Lu, J.G., Su, X.P., 2003. Theory and Application of Reliability Optimization Design of Electromagnetic System on Electrical Apparatus. China Machine Press, Beijing (in Chinese).

Mann, N.R., Schafer, R.E., Singpurwalla, N.D., 1974. Methods for Statistical Analysis of Reliability and Life Data. John Willey & Sons, New York, p.69-78.

Nelson, W., 1990. Accelerated Testing-statistical Models, Test Plans, and Data Analysis. John Willey & Sons, New York.

Weibull, W., 1951. A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, **18**:293-297.