

Estimating Mixed Exponential Distributions Based on Hybrid Censored Samples *

TIAN YuZhu

(School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang, 471023, China)

HAN XueFeng

(School of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo, 454000, China)

TIAN MaoZai

(Center for Applied Statistics, Renmin University of China, Beijing, 100872, China)

Abstract: The hybrid censoring scheme is a mixture of type-I and type-II censoring schemes. It is a popular censoring scheme in the literature of life data analysis. Mixed exponential distribution (MED) models is a class of favorable models in reliability statistics. Nevertheless, there is no much discussion to focus on parameters estimation for MED models with hybrid censored samples. We will address this problem in this paper. The EM (Expectation-Maximization) algorithm is employed to derive the closed form of the maximum likelihood estimators (MLEs). Finally, Monte Carlo simulations and a real-world data analysis are conducted to illustrate the proposed method.

Keywords: mixed exponential distribution (MED); hybrid censoring; maximum likelihood estimation (MLE); EM algorithm

2010 Mathematics Subject Classification: 62F10; 62N05

Citation: Tian Y Z, Han X F, Tian M Z. Estimating mixed exponential distributions based on hybrid censored samples [J]. Chinese J. Appl. Probab. Statist., 2017, 33(2): 191–202.

§1. Introduction

The type-I and type-II censoring schemes are the two most popular schemes in life testing. In type-I censoring scheme, the experiment can not terminate until a pre-specified time reaches. On the other hand, the type-II censoring scheme requires the experiment

*The project was partly supported by the National Natural Science Foundation of China (Grant Nos. 11501167; 11271368), the Key Scientific Research Project of Henan Province Universities (Grant No. 15A110025) and the Doctoral Scientific Research Foundation of Henan University of Science and Technology (Grant No. 4010-13480049).

Received September 24, 2015. Revised June 6, 2016.

to continue until a pre-specified number of failures occurs. Combining the type-I censoring with type-II censoring, the hybrid censoring scheme performs flexible and has become quite eye-catching in the literature of life testing. For example, Ebrahimi^[1] discussed the prediction intervals of future failures for exponential distribution under hybrid censoring scheme, Childs et al.^[2] investigated the exact likelihood inference of the exponential distribution based on the hybrid censored samples, Kundu^[3] considered statistical inference of Weibull distribution under hybrid censoring, Kundu and Pradhan^[4] discussed parameters estimation of generalized exponential distribution in presence of hybrid censored data, Dube et al.^[5] considered estimation of the hybrid censored log-normal distributions. Recently, Balakrishnan and Kundu^[6] presented a review on hybrid censoring, Dey and Pradhan^[7] discussed generalized inverted exponential distribution under hybrid censoring.

Mixture model is an important class of statistical models in reliability analysis. Mixture models under the complete samples, conventional type-I and type-II censoring schemes have been studied by many authors. Many literatures can be referred, such as [8–13]. In this paper, we are interested in parameters estimation of MED based on hybrid censoring scheme. The pdf (probability density function) and cdf (cumulative distribution function) of MED with K components can be given respectively as follows

$$f(x; p, \lambda) = \sum_{k=1}^K p_k \lambda_k e^{-\lambda_k x}, \quad x \geq 0, \quad (1)$$

$$F(x; p, \lambda) = 1 - \sum_{k=1}^K p_k e^{-\lambda_k x}, \quad x \geq 0, \quad (2)$$

where $p = (p_1, p_2, \dots, p_{K-1})$, $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_K)$; $0 < p_k < 1$, $k = 1, 2, \dots, K-1$, $p_K = 1 - \sum_{k=1}^{K-1} p_k$; $\lambda_k > 0$, $k = 1, 2, \dots, K$. The remaining sections are arranged as follows. In Section 2, we consider the MLEs of MED models under the hybrid censoring. In Section 3, we propose estimation procedure of the EM algorithm. In Section 4, some simulations are implemented to illustrate the proposed procedures. In Section 5, a real-world data analysis is provided for further illustration purpose. In the last section, we draw some conclusions.

§2. The Likelihood Function

2.1 The Censoring Scheme

The life-testing experiment of hybrid censoring scheme can be described as follows: Suppose that n identical units are put on test. The experiment is terminated when a pre-

assigned number r , out of n units have failed or a pre-determined time T has been reached. Therefore, in the hybrid censoring scheme, the experimental time and the number of failures will not exceed T and r , respectively. Let X_1, X_2, \dots, X_n be n lifetimes from model (1) and $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ be the corresponding order statistics. The number of failures and observation times are denoted by D and $C = \min(X_{r:n}, T)$, respectively. Thus, the observed sample is represented by $(X_{1:n}, X_{2:n}, \dots, X_{D:n}, C, D)$. Note that when $D = 0$, no failure information is observed. It is obvious that $X_{(D+1):n}, X_{(D+2):n}, \dots, X_{n:n}$ are not can be observed. The experimental time T is supposed to be bounded, and it depends on the maximum experimental time for the experimenter to afford.

2.2 The Likelihood

Under the hybrid censoring scheme, we consider the maximum likelihood estimators of model (1). Let c and d be the observed values of C and D , respectively. The likelihood function is

$$\begin{aligned} l(p, \lambda) &= \prod_{i=1}^d f(x_{i:n}; p, \lambda) \cdot (1 - F(c; p, \lambda))^{n-d} \\ &= \prod_{i=1}^d \left(\sum_{k=1}^K p_k \lambda_k e^{-\lambda_k x_{i:n}} \right) \cdot \left(\sum_{k=1}^K p_k e^{-\lambda_k c} \right)^{n-d}. \end{aligned}$$

However, the MLEs of parameters in this case are hard to obtained directly due to its complex likelihood. The EM algorithm^[14] is employed to address this problem.

§3. The Proposed Procedure

Suppose X_1, X_2, \dots, X_n are n identical independent samples from model (1). Denote

$$\begin{aligned} f_{kj} &= \lambda_k e^{-\lambda_k x_j}, \quad s_{kj} = e^{-\lambda_k x_j}, \quad k = 1, 2, \dots, m, \\ f_j &= \sum_{k=1}^K p_k f_{kj}, \quad s_j = \sum_{k=1}^K p_k s_{kj}, \quad j = 1, 2, \dots, n. \end{aligned}$$

Let $I_j = (I_{j1}, I_{j2}, \dots, I_{jK})$ be an indicator vector of X_j , where I_{jk} is dichotomous variable only taking 1 if X_j comes from the k -th component, and 0 otherwise. Denote $I = (I_1, I_2, \dots, I_n)$ as a big indicator vector composed of n indicator vectors of each life variable X_j . We notice that random vector $I_j = (I_{j1}, I_{j2}, \dots, I_{jK})$ follows a multinomial distribution. I_j is not observable which can be deemed as the missing data. Denote $I_j^{(1)} = (I_{j1}^{(1)}, I_{j2}^{(1)}, \dots, I_{jK}^{(1)})$ and $I_j^{(2)} = (I_{j1}^{(2)}, I_{j2}^{(2)}, \dots, I_{jK}^{(2)})$ as the indicator vectors of the complete data and the censored data, respectively.

For the failure observation X_j , the joint density of X_j and $I_j^{(1)}$ is

$$g(x_j, I_j^{(1)} | p, \lambda) = \prod_{k=1}^K [p_k f_{kj}]^{I_{jk}^{(1)}}.$$

For the censored variable X_j , the joint density of X_j and $I_j^{(2)}$ is

$$g(x_j, I_j^{(2)} | p, \lambda) = \prod_{k=1}^K [p_k s_{kj}]^{I_{jk}^{(2)}}.$$

According to the Bayesian theorem, the conditional probabilities of $I_j^{(1)}$ and $I_j^{(2)}$ on X_j are, respectively

$$P(I_{jk}^{(1)} = 1 | x_j, p, \lambda) = \frac{p_k f_{kj}}{f_j}, \quad P(I_{jk}^{(2)} = 1 | x_j, p, \lambda) = \frac{p_k s_{kj}}{s_j}, \quad k = 1, 2, \dots, K.$$

For the above life variables X_1, X_2, \dots, X_n , we conduct the life-testing experiment of hybrid censoring in Section 2. The complete failure times of D units are denoted as $X_{1:n}, X_{2:n}, \dots, X_{D:n}$. The observed data can be denoted as $X = (X_{1:n}, X_{2:n}, \dots, X_{D:n})$. Denote $Z = \{Z_j, j = 1, 2, \dots, n - D\}$, where Z_j stands for the j -th censored variable at the failure time C . Z is not observable variable which can also be deemed as the missing data. The whole missing data can be denoted as (I, Z) and the complete data can be denoted as $W = (X, I, Z)$. Next, we employ the EM algorithm to obtain the MLEs of the unknown parameters.

The joint pdf of the complete data W can be obtained as follows

$$f(p, \lambda | W) \propto \prod_{i=1}^d \prod_{k=1}^K (p_k \lambda_k e^{-\lambda_k x_{i:n}})^{I_{ik}^{(1)}} \cdot \prod_{j=1}^{n-d} \prod_{k=1}^K (p_k \lambda_k e^{-\lambda_k z_j})^{I_{jk}^{(2)}}.$$

The log-likelihood function of the above complete data is

$$\ln f(p, \lambda | W) = \sum_{k=1}^K \left[\sum_{i=1}^d I_{ik}^{(1)} (\ln p_k + \ln \lambda_k - \lambda_k x_{i:n}) + \sum_{j=1}^{n-d} I_{jk}^{(2)} (\ln p_k + \ln \lambda_k - \lambda_k z_j) \right].$$

The EM algorithm is described as follows:

Given initial values $p^{(0)}, \lambda^{(0)}$ of parameters p, λ , we can obtain parameter estimators via the following two steps:

E step: Given the $(h-1)$ -th iteration values $p^{(h-1)}, \lambda^{(h-1)}$, the Q function of the h -th iteration is

$$\begin{aligned} & Q(p, \lambda | p^{(h-1)}, \lambda^{(h-1)}, W) \\ &= E[\ln f(p, \lambda | W) | p^{(h-1)}, \lambda^{(h-1)}, W] \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^K \left[\sum_{i=1}^d (\ln p_k + \ln \lambda_k - \lambda_k x_{i:n}) \cdot \mathbb{E}(\mathbb{E}(I_{ik}^{(1)} | p^{(h-1)}, \lambda^{(h-1)}, Z)) \right. \\
&\quad + \sum_{j=1}^{n-d} ((\ln p_k + \ln \lambda_k) \cdot \mathbb{E}(\mathbb{E}(I_{jk}^{(2)} | p^{(h-1)}, \lambda^{(h-1)}, Z)) \\
&\quad \left. - \lambda_k \cdot \mathbb{E}(\mathbb{E}(z_j I_{jk}^{(2)} | p^{(h-1)}, \lambda^{(h-1)}, Z))) \right] \\
&= \sum_{k=1}^K \left[\sum_{i=1}^d (\ln p_k + \ln \lambda_k - \lambda_k x_{i:n}) \cdot a_{ki}^{(h-1)}(x_{i:n}) \right. \\
&\quad \left. + \sum_{j=1}^{n-d} ((\ln p_k + \ln \lambda_k) \cdot \mathbb{E}(b_{kj}^{(h-1)}(z_j)) - \lambda_k \cdot \mathbb{E}(z_j \cdot b_{kj}^{(h-1)}(z_j))) \right],
\end{aligned}$$

where

$$\begin{aligned}
a_{ki}^{(h-1)}(x) &= \frac{p_k^{(h-1)} \cdot f_{ki}^{(h-1)}(x)}{f_i^{(h-1)}(x)}, \quad f_{ki}^{(h-1)}(x) = \lambda_k^{(h-1)} e^{-\lambda_k^{(h-1)} x}, \\
f_i^{(h-1)}(x) &= \sum_{k=1}^K p_k^{(h-1)} \cdot f_{ki}^{(h-1)}(x), \quad k = 1, 2, \dots, K, \quad i = 1, 2, \dots, d, \\
b_{kj}^{(h-1)}(x) &= \frac{p_k^{(h-1)} \cdot s_{kj}^{(h-1)}(x)}{s_j^{(h-1)}(x)}, \quad s_{kj}^{(h-1)}(x) = e^{-\lambda_k^{(h-1)} x}, \\
s_j^{(h-1)}(x) &= \sum_{k=1}^K p_k^{(h-1)} \cdot s_{kj}^{(h-1)}(x), \quad k = 1, 2, \dots, K, \quad j = 1, 2, \dots, n-d.
\end{aligned}$$

In the above Q function, the conditional pdf(s) of all censored data Z_j are

$$\begin{aligned}
p_j(z) &= p_j(z | p^{(h-1)}, \lambda^{(h-1)}, W) \\
&= \left[\sum_{k=1}^K p_k^{(h-1)} \lambda_k^{(h-1)} e^{-\lambda_k^{(h-1)} z} \right] / \left[\sum_{k=1}^K p_k^{(h-1)} e^{-\lambda_k^{(h-1)} \cdot c} \right], \\
&\quad z \in [c, +\infty), \quad j = 1, 2, \dots, n-d.
\end{aligned}$$

Then, we have

$$\begin{aligned}
&Q(p, \lambda | p^{(h-1)}, \lambda^{(h-1)}, W) \\
&= \sum_{k=1}^K \left[\left(\sum_{i=1}^d \ln p_k + \ln \lambda_k - \lambda_k x_{i:n} \right) \cdot a_{ki}^{(h-1)}(x_{i:n}) + (n-d)(\ln p_k + \ln \lambda_k) \right. \\
&\quad \left. \cdot \int_c^\infty b_{kj}^{(h-1)}(x) \cdot p_j(x) dx - \lambda_k(n-d) \cdot \int_c^\infty x \cdot b_{kj}^{(h-1)}(x) \cdot p_j(x) dx \right] \\
&= \sum_{k=1}^K (\ln p_k + \ln \lambda_k) \left(\sum_{i=1}^d \Delta 1_{ki}^{(h-1)} + (n-d) \Delta 2_k^{(h-1)} \right) \\
&\quad - \lambda_k \left(\sum_{i=1}^d x_{i:n} \cdot \Delta 1_{ki}^{(h-1)} + (n-d) \Delta 3_k^{(h-1)} \right),
\end{aligned}$$

where

$$\begin{aligned}\Delta 1_{ki}^{(h-1)} &= a_{ki}^{(h-1)}(x_{i:n}), \\ \Delta 2_k^{(h-1)} &= \int_c^\infty b_{kj}^{(h-1)}(x) \cdot p_j(x) dx, \\ \Delta 3_k^{(h-1)} &= \int_c^\infty x \cdot b_{kj}^{(h-1)}(x) \cdot p_j(x) dx.\end{aligned}$$

M step: We optimize numerically the Q function of E-step regarding to p and λ to derive the updated estimators $p^{(h)}$ and $\lambda^{(h)}$. Let

$$\begin{aligned}\frac{\partial Q}{\partial \lambda_k} &= \frac{1}{\lambda_k} \left(\sum_{i=1}^d \Delta 1_{ki}^{(h-1)} + (n-d) \Delta 2_k^{(h-1)} \right) - \left(\sum_{i=1}^d x_{i:n} \Delta 1_{ki}^{(h-1)} + (n-d) \Delta 3_k^{(h-1)} \right) \\ &= 0, \quad k = 1, 2, \dots, K,\end{aligned}\tag{3}$$

$$\begin{aligned}\frac{\partial Q}{\partial p_k} &= \frac{1}{p_k} \left[\sum_{i=1}^d \Delta 1_{ki}^{(h-1)} + (n-d) \Delta 2_k^{(h-1)} \right] - 1 / \left(1 - \sum_{l=1}^{K-1} p_l \right) \\ &\quad \cdot \left(\sum_{i=1}^d \Delta 1_{Ki}^{(h-1)} + (n-d) \Delta 2_K^{(h-1)} \right) \\ &= 0, \quad k = 1, 2, \dots, K-1.\end{aligned}\tag{4}$$

From (3), we obtain

$$\begin{aligned}\hat{\lambda}_k^{(h)} &= \left[\sum_{i=1}^d \Delta 1_{ki}^{(h-1)} + (n-d) \Delta 2_k^{(h-1)} \right] / \left[\sum_{i=1}^d x_{i:n} \cdot \Delta 1_{ki}^{(h-1)} + (n-d) \Delta 3_k^{(h-1)} \right], \\ &\quad k = 1, 2, \dots, K.\end{aligned}\tag{5}$$

From (4), we have

$$\begin{aligned}& p_k \left(\sum_{i=1}^d \Delta 1_{Ki}^{(h-1)} + (n-d) \Delta 2_K^{(h-1)} \right) + \sum_{l=1}^{K-1} p_l \left(\sum_{i=1}^d \Delta 1_{ki}^{(h-1)} + (n-d) \Delta 2_k^{(h-1)} \right) \\ &= \sum_{i=1}^d \Delta 1_{ki}^{(h-1)} + (n-d) \Delta 2_k^{(h-1)},\end{aligned}$$

where $k = 1, 2, \dots, K-1$.

From the above equations, the h -th iteration values in M-step with respect to parameters p_1, p_2, \dots, p_{K-1} are the solution vector of the linear equation group $AP = b$, where P , A , b are given as follows

$$\begin{aligned}P &= (p_1, p_2, \dots, p_{K-1})^\top, \quad A_{K-1} = (a_{ls}), \\ a_{ls} &= \begin{cases} \sum_{i=1}^d (\Delta 1_{li}^{(h-1)} + \Delta 1_{Ki}^{(h-1)}) + (n-d)(\Delta 2_l^{(h-1)} + \Delta 2_K^{(h-1)}), & l = s; \\ \sum_{i=1}^d \Delta 1_{li}^{(h-1)} + (n-d) \Delta 2_l^{(h-1)}, & l \neq s, \end{cases}\end{aligned}$$

$$b = \left(\sum_{i=1}^d \Delta 1_{1i}^{(h-1)} + (n-d)\Delta 2_1^{(h-1)}, \dots, \sum_{i=1}^d \Delta 1_{K-1,i}^{(h-1)} + (n-d)\Delta 2_{K-1}^{(h-1)} \right)^\top.$$

If $\sum_{i=1}^d \Delta 1_{li}^{(h-1)} + (n-d)\Delta 2_l^{(h-1)} > 0$, $l = 1, 2, \dots, K$, we can easily prove that $\text{rank}(A) = K - 1$. Thus, the only solution of parameter vector P of the h -th iteration in the M-step is

$$\hat{p}^{(h)} = (\hat{p}_1^{(h)}, \hat{p}_2^{(h)}, \dots, \hat{p}_{m-1}^{(h)})^\top = A^{-1}b. \quad (6)$$

From the above (5) and (6), we can update $(p^{(h)}, \lambda^{(h)})$ by repeating E-step and M-step till the total errors approach the pre-assigned constraints. It needs to be emphasized that, in practical applications, we can try to run the EM algorithm several times at different starting values for obtaining more stable estimates.

§4. Simulations

In this section, some simulations are conducted to illustrate the performance of the proposed method. Suppose X_i , $i = 1, 2, \dots, n$ are n identical independently distributed samples generated from model (1), we consider MED models with two mixture components under the hybrid censoring scheme. In order to compare results between different censoring schemes, the true values of parameters are uniformly set as $p_1 = 0.4$, $\lambda_1 = 2.4$, $\lambda_2 = 0.8$, whereas the initial values are supposed to be $p_1^{(0)} = 0.3$, $\lambda_1^{(0)} = 2$, $\lambda_2^{(0)} = 0.5$ for simplicity.

Table 1 Mean(s) and RMSE(s) when $T = 0.8$

| n | r | Mean | | | RMSE | | |
|-----|-----|-------|-------------|-------------|-------|-------------|-------------|
| | | p_1 | λ_1 | λ_2 | p_1 | λ_1 | λ_2 |
| 30 | 15 | 0.330 | 3.869 | 0.997 | 0.089 | 2.126 | 0.434 |
| | 20 | 0.340 | 3.877 | 0.749 | 0.085 | 1.922 | 0.246 |
| | 30 | 0.341 | 3.891 | 0.699 | 0.085 | 1.928 | 0.183 |
| 50 | 20 | 0.344 | 3.597 | 1.025 | 0.071 | 1.363 | 0.330 |
| | 30 | 0.343 | 3.666 | 0.761 | 0.070 | 1.448 | 0.198 |
| | 40 | 0.345 | 3.709 | 0.683 | 0.070 | 1.498 | 0.163 |
| | 50 | 0.345 | 3.709 | 0.683 | 0.070 | 1.498 | 0.163 |
| 80 | 40 | 0.345 | 3.377 | 0.804 | 0.063 | 1.126 | 0.195 |
| | 50 | 0.341 | 3.303 | 0.671 | 0.067 | 1.059 | 0.182 |
| | 60 | 0.340 | 3.285 | 0.635 | 0.068 | 1.041 | 0.179 |
| | 70 | 0.340 | 3.286 | 0.635 | 0.068 | 1.042 | 0.179 |

Consider three different censoring schemes for $T = 0.8, 1.5, 3$. Repeat the estimation process 1 000 times for distinct sample sizes $n = 30, 50, 80$ with pre-assigned failure number r . The estimated means and root mean square errors (RMSEs) over $s = 1000$ times simulations are presented in Table 1–3. All simulations are carried out via Matlab2013.

Table 2 Mean(s) and RMSE(s) when $T = 1.5$

| n | r | Mean | | | RMSE | | |
|-----|-----|-------|-------------|-------------|-------|-------------|-------------|
| | | p_1 | λ_1 | λ_2 | p_1 | λ_1 | λ_2 |
| 30 | 15 | 0.374 | 2.681 | 1.212 | 0.050 | 0.834 | 0.509 |
| | 20 | 0.376 | 2.731 | 1.019 | 0.049 | 0.742 | 0.381 |
| | 30 | 0.384 | 2.835 | 0.788 | 0.051 | 0.827 | 0.225 |
| 50 | 20 | 0.379 | 2.639 | 1.269 | 0.039 | 0.508 | 0.523 |
| | 30 | 0.378 | 2.661 | 1.043 | 0.039 | 0.521 | 0.356 |
| | 40 | 0.378 | 2.712 | 0.798 | 0.042 | 0.581 | 0.224 |
| | 50 | 0.379 | 2.727 | 0.737 | 0.044 | 0.596 | 0.168 |
| 80 | 40 | 0.380 | 2.590 | 1.001 | 0.032 | 0.358 | 0.312 |
| | 50 | 0.377 | 2.586 | 0.886 | 0.034 | 0.344 | 0.215 |
| | 60 | 0.373 | 2.597 | 0.773 | 0.036 | 0.332 | 0.158 |
| | 70 | 0.372 | 2.610 | 0.710 | 0.037 | 0.336 | 0.141 |

Table 3 Mean(s) and RMSE(s) when $T = 3$

| n | r | Mean | | | RMSE | | |
|-----|-----|-------|-------------|-------------|-------|-------------|-------------|
| | | p_1 | λ_1 | λ_2 | p_1 | λ_1 | λ_2 |
| 30 | 15 | 0.376 | 2.368 | 1.082 | 0.048 | 0.585 | 0.351 |
| | 20 | 0.376 | 2.387 | 1.035 | 0.047 | 0.580 | 0.320 |
| | 30 | 0.386 | 2.484 | 0.850 | 0.047 | 0.612 | 0.235 |
| 50 | 20 | 0.379 | 2.379 | 1.965 | 0.034 | 0.403 | 0.317 |
| | 30 | 0.379 | 2.387 | 1.019 | 0.034 | 0.398 | 0.280 |
| | 40 | 0.380 | 2.419 | 0.940 | 0.034 | 0.410 | 0.228 |
| | 50 | 0.383 | 2.442 | 0.814 | 0.036 | 0.404 | 0.156 |
| 80 | 40 | 0.383 | 2.377 | 0.982 | 0.027 | 0.251 | 0.224 |
| | 50 | 0.382 | 2.377 | 0.955 | 0.028 | 0.249 | 0.202 |
| | 60 | 0.381 | 2.380 | 0.918 | 0.029 | 0.243 | 0.174 |
| | 70 | 0.381 | 2.389 | 0.862 | 0.027 | 0.241 | 0.147 |

From Table 1–3, it can be seen that, the EM algorithm is relatively effective for estimating MED model (1) under different hybrid censored schemes. We can see that for fixed n and r , as T increases, the biases, the RMSEs for most of estimated parameters

decline as expected. Similarly, we also observe that for fixed n and T as r increases, overall, the biases and the RMSEs become smaller for all parameters. And for fixed r and T as n increases, the biases and the RMSEs of all parameters decline.

§5. Data Analysis

In this example, we analyze a set of failure times of the air conditional system. This data set has been utilized by many authors such as [15], [16] and [13]. Mokhtari et al.^[16] asserted that the Weibull distribution can provide a good fit to the failure data. Tian et al.^[13] fitted this data better by using MED model with two mixture components. In their results, the MLEs are $\hat{p}_1 = 0.346$, $\hat{\lambda}_1 = 0.065$, $\hat{\lambda}_2 = 0.012$. The p -value is 0.697 with the K-S distance 0.124 by Kolmogorov-Smirnov goodness-of-fit test. Figure 1 provides the empirical survival function and the fitted failure rate function.

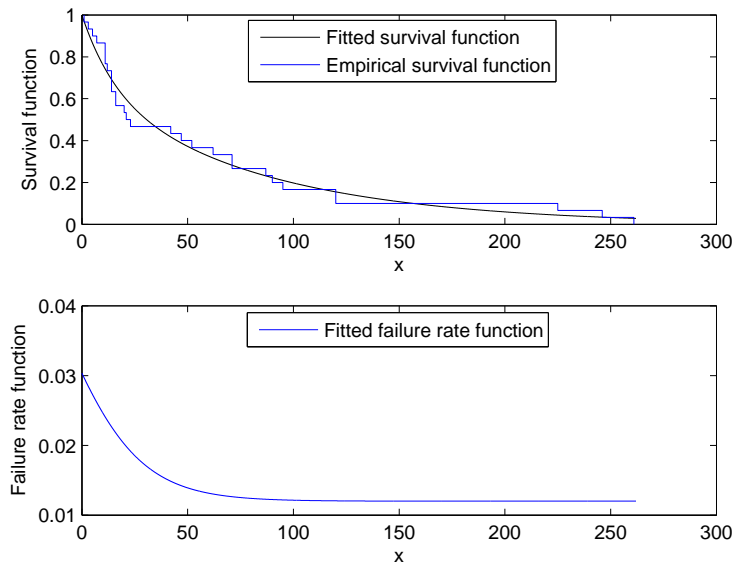


Figure 1 Fitted survival function and failure rate function

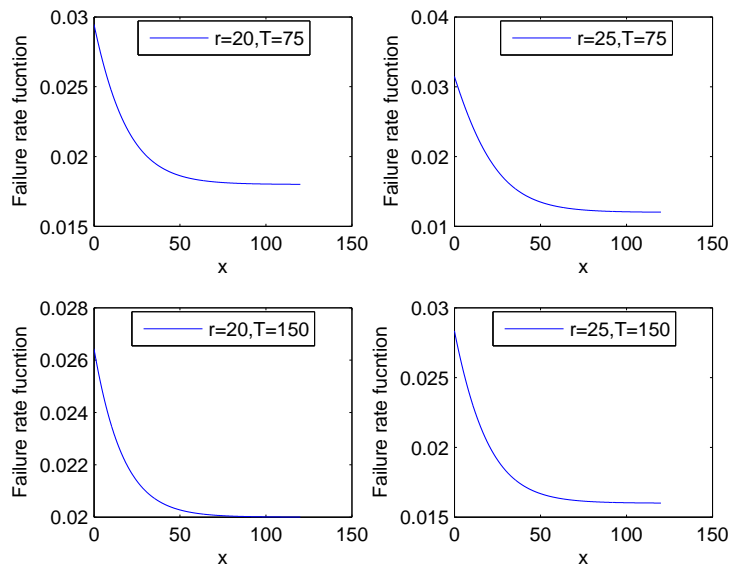
Next, we generate the artificially hybrid censored data from the above data set by taking $r = 20, 25$ and $T = 75, 150$ which result in a total of 4 censoring schemes. Based on the proposed estimation procedure, the estimates (Est) and estimated standard errors (S.E.), 95% quantile confidence intervals (C.I.) of 500 times Bootstrap samples^[17] under 4 set-ups are listed in Table 4.

From the Table 4, we can see that, for fixed censoring value T as r increases (the

Table 4 Est (S.E.) and 95% C.I. under 4 set-ups

| | | $T = 75$ | | $T = 150$ | |
|------------|-------------|---------------|---------------|---------------|---------------|
| | | $r = 20$ | $r = 25$ | $r = 20$ | $r = 25$ |
| Est (S.E.) | p_1 | 0.185(0.057) | 0.330(0.081) | 0.099(0.049) | 0.200(0.087) |
| | λ_1 | 0.080(0.032) | 0.071(0.017) | 0.085(0.053) | 0.077(0.024) |
| | λ_2 | 0.018(0.007) | 0.012(0.003) | 0.020(0.005) | 0.016(0.004) |
| 95% C.I. | p_1 | (0.085,0.325) | (0.065,0.497) | (0.023,0.243) | (0.069,0.451) |
| | λ_1 | (0.041,0.127) | (0.041,0.111) | (0.026,0.158) | (0.031,0.131) |
| | λ_2 | (0.008,0.037) | (0.007,0.021) | (0.013,0.031) | (0.008,0.025) |

bigger r means lower censoring rate), the estimated S.E. decrease and the 95% C.I. of scales λ_1 and λ_2 parameters become shorter. However, for mixing rate p_1 , the conclusion seems somewhat inconsistent. And for fixed r as T increases, overall, the estimated S.E. decrease for most of parameters but the trends of the 95% C.I. are not very uniform for smaller r and bigger r . The fitted failure rate functions under different censoring set-ups are provided in Figure 2. From Figure 2, we can see the estimated hazard functions are different for different r and T . For example, for the same r , as T becomes bigger, the hazard function also gets higher. For the same T , the bigger r (higher censoring rate), the lower failure rate function we can suffer. The opposite is also true in practical lifetime analysis.

**Figure 2 Fitted failure rate function under different censoring set-ups**

§6. Conclusion

We investigate the parameters estimation of MED models (1) under the hybrid censored data. The MLEs via the EM-Algorithm are obtained and some Monte Carlo simulations are carried out to investigate the performance of the proposed estimation technique. Finally, a real-world data analysis is conducted to illustrate the developed method. Another important issue for mixture models is to select proper model order (the number of mixture components). In this paper, the number of mixture components K of MED model is assumed to be known and we do not discuss too much on how to select the order. However, in practical applications, K is really unknown. Estimating the order is a challenging job in modeling mixture models. Many statistical methods for selecting K have been proposed, for instance, AIC (Akaike's information criterion), BIC (Bayesian information criterion), graph clustering approach. In addition, some new penalized log-likelihood methods (such as [18] and [19]) have also been employed to conduct order selection of mixture models. In future work, we will carry out order selection via incorporating the above methods into of mixture lifetime models under complex censoring schemes.

References

- [1] Ebrahimi N. Prediction intervals for future failures in the exponential distribution under hybrid censoring [J]. *IEEE Trans. Reliab.*, 1992, **41**(1): 127–132.
- [2] Childs A, Chandrasekhar B, Balakrishnan N, et al. Exact likelihood inference based on type-I and type-II hybrid censored samples from the exponential distribution [J]. *Ann. Inst. Statist. Math.*, 2003, **55**(2): 319–330.
- [3] Kundu D. On hybrid censored Weibull distribution [J]. *J. Statist. Plann. Inference*, 2007, **137**(7): 2127–2142.
- [4] Kundu D, Pradhan B. Estimating the parameters of the generalized exponential distribution in presence of hybrid censoring [J]. *Comm. Statist. Theory Methods*, 2009, **38**(12): 2030–2041.
- [5] Dube S, Pradhan B, Kundu D. Parameter estimation of the hybrid censored log-normal distribution [J]. *J. Stat. Comput. Simul.*, 2011, **81**(3): 275–287.
- [6] Balakrishnan N, Kundu D. Hybrid censoring: models, inferential results and applications [J]. *Comput. Statist. Data Anal.*, 2013, **57**(1): 166–209.
- [7] Dey S, Pradhan B. Generalized inverted exponential distribution under hybrid censoring [J]. *Stat. Methodol.*, 2014, **18**: 101–114.
- [8] Mendenhall W, Hader R J. Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data [J]. *Biometrika*, 1958, **45**(3/4): 504–520.
- [9] Al-Hussaini E K. Bayesian prediction under a mixture of two exponential components model based on type I censoring [J]. *J. Appl. Statist. Sci.*, 1999, **8**(2-3): 173–185.

- [10] McLachlan G, Peel D. *Finite Mixture Models* [M]. New York: Wiley, 2000.
- [11] Jaheen Z F. On record statistics from a mixture of two exponential distributions [J]. *J. Stat. Comput. Simul.*, 2005, **75**(1): 1–11.
- [12] Jones P W, Ashour S K. Bayesian estimation of the parameters of the mixed exponential distribution from censored samples [J]. *Biom. J.*, 2007, **18**(8): 633–637.
- [13] Tian Y Z, Tian M Z, Zhu Q Q. Estimating a finite mixed exponential distribution under progressively type-II censored data [J]. *Comm. Statist. Theory Methods*, 2014, **43**(17): 3762–3776.
- [14] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm [J]. *J. Roy. Statist. Soc. Ser. B*, 1977, **39**(1): 1–38.
- [15] Linhart H, Zucchini W. *Model Selection* [M]. New York: Wiley, 1986.
- [16] Mokhtari E M, Rad A H, Yousefzadeh F. Inference for Weibull distribution based on progressively type-II hybrid censored data [J]. *J. Statist. Plann. Inference*, 2011, **141**(8): 2824–2838.
- [17] Efron B, Tibshirani R J. *An Introduction to the Bootstrap* [M]. New York: Chapman and Hall, 1993.
- [18] Chen J H, Khalili A. Order selection in finite mixture models with a nonsmooth penalty [J]. *J. Amer. Statist. Assoc.*, 2009, **104**(485): 187–196.
- [19] Luo R H, Wang H S, Tsai C L. On mixture regression shrinkage and selection via the MR-LASSO [J]. *Int. J. Pure Appl. Math.*, 2008, **46**(3): 403–414.

混合删失样本下混合指数分布的估计

田玉柱

韩学峰

(河南科技大学数学与统计学院, 洛阳, 471023)

(河南理工大学数学与信息科学学院, 焦作, 454000)

田茂再

(中国人民大学应用统计中心, 北京, 100872)

摘 要: 混合删失方案是I型和II型删失方案的一个组合, 该删失方案在寿命数据分析已非常流行. 混合指数分析模型是可靠性统计中非常受欢迎的一类模型. 然而, 带混合删失样本的混合指数分布的参数估计问题还没有出现相关的讨论, 本文将来处理这个问题. 文章利用EM算法获得了参数极大似然估计的显示表达式. 最后, 我们开展了一些数值模拟和实际数据分析来展示所提方法的估计效果.

关键词: 混合指数分布; 混合删失; 极大似然估计; EM算法

中图分类号: O212.2; O213.2