An EM Algorithm for Estimating the Parameters of the Generalized Exponential Distribution under Unified Hybrid Censored Data

A. Habibi Rad* and M. Izanlo

Ferdowsi University of Mashhad

Abstract. The unified hybrid censoring is a mixture of generalized Type-I and Type-II hybrid censoring schemes. This article presents the statistical inferences on Generalized Exponential Distribution parameters when the data are obtained from the unified hybrid censoring scheme. It is observed that the maximum likelihood estimators can not be derived in closed form. The EM algorithm for computing the maximum likelihood estimators is proposed. We calculated the observed Fisher information matrix using the missing information principle which is useful for constructing the asymptotic confidence intervals. Simulations studies are performed to compare the performances of the estimators obtained under different schemes. Finally, a real data set has been analyzed for illustrative purposes.

Keywords. EM algorithm; Fisher information matrix; maximum likelihood estimators; unified hybrid censoring.

MSC 2010: 62N01, 62N02.

1 Introduction

Consider a life-testing experiment in which n identical units are placed on a life-test. Let $X_{1:n} \leqslant X_{2:n} \leqslant \cdots \leqslant X_{n:n}$ denote the corresponding lifetimes

---
* Corresponding author

from a generalized exponential (GE) distribution with probability density function (pdf) $f_\theta(x)$ and cumulative distribution function (cdf) $F_\theta(x)$. Generalized Type-I and Type-II hybrid censoring schemes (HCS) can be briefly described as follow:

Suppose n units are put on a life test. In generalized Type-I HCS, one fixes $r, k\epsilon\{1, 2, \ldots, n\}$ and $T\epsilon(0, \infty)$ such that $k < r$. If the k-th failure occurs before time T, the experiment is terminated at min $\{X_{r:n}, T\}$; if the k-th failure occurs after time T, the experiment is terminated at $X_{k:n}$. Thus, in generalized Type-I HCS, a minimum number k of failures are guaranteed. In generalized Type-II HCS, one fixes $r\epsilon\{1, 2, \ldots, n\}$ and $T_1, T_2\epsilon(0, \infty)$ such that $T_1 < T_2$. If the r-th failure occurs before time T_1, the experiment is terminated at T_1; if the r-th failure occurs between T_1 and T_2, the experiment is terminated at $X_{r:n}$; otherwise, the experiment is terminated at T_2. This hybrid censoring scheme guarantees that the experiment will be completed by time T_2, (see Chandrasekar et al., 2004).

A mixture of generalized Type-I and Type-II HCS is known as the unified HCS and it can be described as follows. Suppose n identical units are put to test under the same environmental conditions and the lifetime of each unit is independent and identically distribution (i.i.d) random variables. Fix $k, r\epsilon\{1, 2, \ldots, n\}$ and $T_1 < T_2\epsilon(0, \infty)$ such that $k < r$. If k-th failure occurs before time T_1, the experiment terminate at $\min\{\max\{X_{r:n}, T_1\}, T_2\}$; if the k-th failure occurs between T_1 and T_2, the experiment terminate at $\min\{X_{r:n}, T_2\}$ and if the k-th failure occurs after time T_2, then the experiment terminate at $X_{k:n}$. Under this censoring scheme, we can guarantee that the experiment would be completed at most in time T_2 with at least k failure and if not, we can guarantee exactly k failures.

Balakrishnan et al. (2008), first introduced the unified HCS and analyzed the data under the assumption of exponential lifetime distribution of the experimental units. They also obtained exact confidence intervals for the mean of the exponential distribution under the unified HCS.

Thus, under the unified HCS, we have the following six cases:

(1) $0 < X_{k:n} < X_{r:n} < T_1 < T_2$ the experiment terminate at T_1,

(2) $0 < X_{k:n} < T_1 < X_{r:n} < T_2$ terminate at $X_{r:n}$,

(3) $0 < X_{k:n} < T_1 < T_2 < X_{r:n}$ terminate at T_2,

(4) $0 < T_1 < X_{k:n} < X_{r:n} < T_2$ terminate at $X_{r:n}$,

(5) $0 < T_1 < X_{k:n} < T_2 < X_{r:n}$ terminate at T_2,

$$(6) \quad 0 < T_1 < T_2 < X_{k:n} < X_{r:n} \quad \text{terminate at } X_{k:n},$$

where $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ denote the observed failure times of the experimental units. Based on the observed unified HCS data, the likelihood function for six cases can be combined and written as follow:

$$L(\theta, \underline{x}) = \frac{n!}{(n-d)!} \prod_{i=1}^{d} f(x_{i:n})[1 - F(c)]^{(n-d)}, \tag{1}$$

where c $(c \epsilon \{T_1, T_2, x_{r:n}, x_{k:n}\})$ is the stopping point and d $(d \epsilon \{D_1, D_2, k, r\})$ is the number of failures until time c. So that

$$
\begin{aligned}
&\text{for case } (1): &&d = D, &&\text{and} \quad c = T_1,\\
&\text{for cases } (2), (4): &&d = r, &&\text{and} \quad c = x_{r:n},\\
&\text{for cases } (3), (5): &&d = D_2, &&\text{and} \quad c = T_2,\\
&\text{for case } (6): &&d = k, &&\text{and} \quad c = x_{k:n},
\end{aligned}
$$

and D_j denote the number of failures until time T_j, $j = 1, 2$ and for case (1) $D = D_1 = D_2$.

In this paper, we consider the analysis of the unified HCS lifetime data when the lifetime of each experimental units follows two-parameter GE distribution.

The GE distribution is one of the most common distribution which is used to analyze several lifetime data. The two-parameter GE was originally proposed by Gupta and Kundu (1999) and it has received considerable attention in the recent years because of its flexibility and wide scale applicability. See, for example, the review article by Gupta and Kundu (2007) on recent development of the GE distribution. The two-parameter GE distribution with the shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$ has the pdf for $x > 0$ as:

$$f_{GE}(x; \alpha, \lambda) = \alpha \lambda e^{-\lambda x}(1 - e^{-\lambda x})^{\alpha-1}; \tag{2}$$

and the corresponding cdf is

$$F_{GE}(x; \alpha, \lambda) = (1 - e^{-\lambda x})^{\alpha}.$$

The aim of this article is the point and interval estimating of the unknown parameters based on the frequentist approach. It is observed that the maximum

likelihood estimates can not be obtained in closed form. Newton-Raphson algorithm is one of the standard methods to determine the maximum likelihood estimates of the parameters (Izanlo and Habibi, 2009). To employ the algorithm, the second derivatives of the log-likelihood are required. Sometimes the calculations of the derivatives based on the unified HCS censored samples are complicated. Unfortunately it does not converge all the time even from good starting values To avoid such computation, we propose to use the EM algorithm instead, similary as in Ng et al. (2002) to compute the MLEs. Also, by using the missing information principle we calculate the observed Fisher information matrix, which can be used for constructing the asymptotic confidence intervals of the unknown parameters. Simulations are performed to compare the performances of the different schemes and a real data set has been analyzed for illustrative purposes.

The rest of the paper is organized as follows. The maximum likelihood estimators and EM algorithm are provided In Section 2. The Fisher information matrix is in Section 3. Simulation results and analyzed of one data set are presented in Section 4. Finally conclusions appear in Section 5.

2   Maximum Likelihood Estimators

In this section based on the observed data in (2), ignoring the additive constant, the log-likelihood function for six cases is given by

$$l(\alpha, \lambda | \underline{x}) = d \ln \alpha + d \ln \lambda - \lambda \sum_{i=1}^{d} x_{i:n} + (\alpha - 1) \sum_{i=1}^{d} \ln(1 - e^{-\lambda x_{i:m}})$$
$$+ (n - d) \ln\{1 - (1 - e^{-\lambda c})^{\alpha}\}, \tag{3}$$

where c and d have defined in (1).
Taking derivatives with respect to α and λ of (3) and putting them equal to zero we obtain

$$\frac{\partial l}{\partial \alpha} = \frac{d}{\alpha} + \sum_{i=1}^{d} \ln(1 - e^{-\lambda x_{i:m}}) - (n - d) \frac{\ln(1 - e^{-\lambda c})(1 - e^{-\lambda c})^{\alpha}}{1 - (1 - e^{-\lambda c})^{\alpha}} = 0,$$

$$\frac{\partial l}{\partial \lambda} = \frac{d}{\lambda} - \sum_{i=1}^{d} x_{i:n} + (\alpha - 1) \sum_{i=1}^{d} \frac{x_{i:n} e^{-\lambda x_{i:n}}}{1 - e^{-\lambda x_{i:n}}}$$

$$- (n - d) \frac{\alpha c e^{-\lambda c}(1 - e^{-\lambda c})^{\alpha-1}}{1 - (1 - e^{-\lambda c})^{\alpha}} = 0. \qquad (4)$$

It is clear that the equations in (4) are implicit, we need some numerical techniques to solve the simultaneous equations. We suggest to use the EM algorithm to compute the MLEs and it is described below.

2.1 EM Algorithm

The EM algorithm, originally proposed by Dempster et al. (1977), is a very powerful tools in handling the incomplete data problem. We treat this problem as a missing value problem similarly as in Ng et al., (2002). Let us denote the observed and the censored data by $X = (X_{1:n}, X_{2:n}, \ldots, X_{d:n})$ and $Z = (Z_1, Z_2, \ldots, Z_{n-d})$ respectively. Here for a given d, $Z_1, Z_2, \ldots, Z_{n-d}$ are not observable. The censored data vector Z can be thought of as missing data. The combination of $W = (X, Z)$ forms the complete data set. The log-likelihood function based on the complete log-lifetime W and ignoring the additive constant is

$$l_c(\alpha, \lambda; W) = n \ln \alpha + n \ln \lambda - \lambda \left(\sum_{i=1}^{d} x_{i:n} + \sum_{i=1}^{n-d} z_i \right)$$

$$+ (\alpha - 1) \left\{ \sum_{i=1}^{d} \ln(1 - e^{-\lambda x_{i:n}}) + \sum_{i=1}^{n-d} \ln(1 - e^{-\lambda z_i}) \right\}. \qquad (5)$$

In E-step one needs to compute the pseudo log-likelihood function as $E(l(\alpha, \lambda; W)|X)$. Therefore,

$$E(l_c(\alpha, \lambda; W)|X) = n \ln \alpha + n \ln \lambda - \lambda \sum_{i=1}^{d} x_{i:n} + (\alpha - 1) \sum_{i=1}^{d} \ln(1 - e^{-\lambda x_{i:n}})$$

$$- \lambda \sum_{i=1}^{n-d} E(Z_i|Z_i > c) + (\alpha - 1)$$

$$\times \sum_{i=1}^{n-d} E \left\{ \ln(1 - e^{-\lambda Z_i})|Z_i > c \right\}. \qquad (6)$$

We know the conditional pdf of Z_i, for $i = 1, \ldots, n - d$, when $X_{1:n} = x_{1:n}, \ldots, X_{d:n} = x_{d:n}$ is (Ng et al. 2002)

$$f_{Z|X}(z_i | X_{1:n} = x_{1:n}, \ldots, X_{d:n} = x_{d:n}) = \frac{f_W(z_i; \alpha, \lambda)}{1 - F_W(c; \alpha, \lambda)}; \qquad z_i > c, \quad (7)$$

Z_i and Z_j for $i \neq j$ are conditionally independent. Based on (7), we can write

$$A(c; \alpha, \lambda) = E(Z_i | Z_i > c) = \frac{\alpha \lambda}{1 - F_{GE}(c; \alpha, \lambda)} \times \int_c^\infty x e^{-\lambda x} (1 - e^{-\lambda x})^{\alpha - 1} \, dx,$$

$$= -\frac{\alpha}{\lambda(1 - F_{GE}(c; \alpha, \lambda))} u(\lambda c, \alpha),$$

where

$$u(\lambda c, \alpha) = \int_0^{e^{-\lambda c}} (1 - z)^{\alpha - 1} \ln z \, dz,$$

and

$$B(c; \alpha, \lambda) = E(\ln(1 - e^{-\lambda z_i}) | Z_i > c),$$

$$= \frac{\alpha \lambda}{1 - F_{GE}(c; \alpha, \lambda)} \times \int_c^\infty e^{-\lambda x} (1 - e^{-\lambda x})^{(\alpha - 1)} \ln(1 - e^{-\lambda x}) \, dx,$$

$$= \frac{1}{\alpha \{1 - F_{GE}(c; \alpha, \lambda)\}} [(1 - e^{-\lambda c})^\alpha \{1 - \alpha \ln(1 - e^{-\lambda c})\} - 1].$$

Now the M-step involves the maximization of (6), if at the k-th stage the estimate of (α, λ) is $(\alpha^{(k)}, \lambda^{(k)})$, then $(\alpha^{(k+1)}, \lambda^{(k+1)})$ can be obtained by maximizing

$$g(\alpha, \lambda) = n \ln \alpha + n \ln \lambda - \lambda \sum_{i=1}^d x_{i:n} + (\alpha - 1) \sum_{i=1}^d \ln(1 - e^{-\lambda x_{i:n}})$$

$$- \lambda(n - d) A(c; \alpha^{(k)}, \lambda^{(k)}) + (\alpha - 1)(n - d) B(c; \alpha^{(k)}, \lambda^{(k)}). \quad (8)$$

Taking derivatives with respect to α and λ of (8) and putting them equal to zero, first find $\lambda^{(k+1)}$ by solving a fixed point type equation as

$$h(\lambda) = \lambda,$$

so that the function $h(\lambda)$ is defined as follow;

$$h(\lambda) = \left[\frac{1}{n} \sum_{i=1}^{d} x_{i:n} + \frac{n-d}{n} A(c; \alpha^{(k)}, \lambda^{(k)}) - \frac{1}{n} \{\hat{\alpha}(\lambda) - 1\} \sum_{i=1}^{d} \frac{x_{i:n} e^{-\lambda x_{i:n}}}{1 - e^{-\lambda x_{i:n}}} \right]^{-1},$$

where

$$\hat{\alpha}(\lambda) = -\frac{n}{\sum_{i=1}^{d} \ln(1 - e^{-\lambda x_{i:n}}) + (n-d)B(c; \alpha^{(k)}, \lambda^{(k)})}.$$

Then, $\alpha^{(k+1)}$ is obtained as $\alpha^{(k+1)} = \hat{\alpha}(\lambda^{(k+1)})$.

3 Fisher Information Matrices

In this section we present the observed Fisher information matrix obtained using the missing value principle of Louis (1982). The observed Fisher information matrix can be used to construct the asymptotic confidence intervals. The idea of missing information principle is as follows;

Observed information = Complete information - Missing information.

Let us use the following notation; $\theta = (\alpha, \lambda)$, $X =$ the observed data, $W =$ the complete data, $I_X(\theta) =$ the observed information, $I_W(\theta) =$ the complete information, $I_{W|X} =$ the missing information, then

$$I_X(\theta) = I_W(\theta) - I_{W|X}(\theta). \tag{9}$$

The complete information $I_W(\theta)$ is given by

$$I_W(\theta) = -E \left[\frac{\partial^2 l(W; \theta)}{\partial \theta^2} \right].$$

The Fisher information matrix of the censored observations can be written as

$$I_{W|X}(\theta) = -(n-d) E_{Z|X} \left[\frac{\partial^2 \ln f_{Z|X}(z|X, \theta)}{\partial \theta^2} \right],$$

The elements of the matrix for complete data set are available in Gupta and Kundu (2001). They are as follow;

$$a_{11} = \frac{n}{\alpha^2}.$$

$$a_{22} = \frac{n}{\lambda^2} + \frac{n\alpha(\alpha-1)}{\lambda^2} \int_0^\infty x^2 e^{-2x}(1-e^{-x})^{\alpha-3}dx.$$

$$a_{12} = a_{21} = -\frac{n\alpha}{\lambda} \int_0^\infty x e^{-2x}(1-e^{-x})^{\alpha-2}dx.$$

Using (7), we present $I_{W|X}$. If

$$I_{W|X}(\theta) = (n-d) \begin{bmatrix} b_{11}(c;\alpha,\lambda) & b_{12}(c;\alpha,\lambda) \\ b_{21}(c;\alpha,\lambda) & b_{22}(c;\alpha,\lambda) \end{bmatrix},$$

then

$$b_{11}(c;\alpha,\lambda) = \frac{1}{\alpha^2} - [\ln(1-e^{-\lambda c})]^2 \frac{(1-e^{-\lambda c})^\alpha}{\{1-(1-e^{-\lambda c})^\alpha\}^2},$$

$$b_{22}(c;\alpha,\lambda) = \frac{1}{\lambda^2} + (\alpha-1)h_1(c;\alpha,\lambda)$$
$$- \frac{\alpha c^2 e^{-\lambda c}(1-e^{-\lambda c})^{\alpha-2}\{\alpha e^{-\lambda c} - 1 + (1-e^{-\lambda c})^\alpha\}}{\{1-(1-e^{-\lambda c})^\alpha\}^2},$$

$$b_{12}(c;\alpha,\lambda) = h_2(c;\alpha,\lambda)$$
$$- \frac{c e^{-\lambda c}(1-e^{-\lambda c})^{\alpha-1}\{1 + \alpha \ln(1-e^{-\lambda c}) - (1-e^{-\lambda c})^\alpha\}}{\{1-(1-e^{-\lambda c})^\alpha\}^2},$$

$$b_{21}(c;\alpha,\lambda) = b_{12}(c;\alpha,\lambda),$$

where

$$h_1(c;\alpha,\lambda) = \frac{1}{\lambda^2\{1-(1-e^{-\lambda c})^\alpha\}}$$
$$\times \int_{(1-e^{-\lambda c})^\alpha}^1 \{\ln(1-u^{1/\alpha})\}^2 \left\{1-u^{\frac{1}{\alpha}}\right\} u^{\frac{-2}{\alpha}} du,$$

$$h_2(c;\alpha,\lambda) = \frac{1}{\lambda\{1-(1-e^{-\lambda c})^\alpha\}}$$
$$\times \int_{(1-e^{-\lambda c})^\alpha}^1 \{\ln(1-u^{1/\alpha})\} \left\{1-u^{\frac{1}{\alpha}}\right\} u^{\frac{-1}{\alpha}} du.$$

Naturally, the asmptotic variance covariance matrix of $\hat{\theta}$ can be obtained by inverting $I_X(\hat{\theta})$.

4 Simulations and Data Analysis

4.1 Simulations

In this section, we present some simulation results to compare the performances of the different schemes, in terms of their mean squared error (MSE) and their coverage percentages. All the programs are written in R.

In each case, we generated a sample from GE distribution with $\alpha = 7$, $\lambda = 0.05$ and $n = 50$. The simulation is carried out for different choices of k, r, T_1 and T_2 values. We have estimated the α and λ (Tables 1 and 2) using the MLEs, we have used the EM algorithm for computing the MLEs, and computed the coverage percentages of the confidence intervals using the observed Fisher information matrix (Tables 4 and 6). For comparison purpose, we replicated the process 10.000 times and report them in Tables 1-6.

The results from Tables 1-6 are as follows:

 (i) for fixed r, T_1 and T_2 when k increase, estimation of the parameters near to the real value, Tables 1 and 2.

 (ii) for fixed r, k and T_2 when T_1 decrease, the $MSEs$ increase and the coverage percentages decrease, Tables 3 and 4.

(iii) for fixed r, k and T_1 when T_2 increase, the $MSEs$ decreases and the coverage percentages increase, Tables 5 and 6.

4.2 Data Analysis

For illustrative purposes, we present here a data analysis using the proposed methods. The data set is taken from Lawless (1982, p. 228). The data given here arose in tests on endurance of deep groove ball bearings. The data are the numbers of million revolution before failure for each of the 23 ball bearings in the life test and they are: 17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.80, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04 and 173.40. It has been observed by Gupta and Kundu (2001) and Kundo and Pradhan (2009) that the two-parameter GE distribution can be used quite effectively to analyze this data set.

Table 1. Estimation of the (α, λ) for unified HCS when T_2 is 95.

r	k	T_1			
		4	30	75	90
19	11	(1.529,0.0131)	(1.529,0.0131)	(7.146,0.0488)	(7.651,0.0508)
33	11	(4.387,0.03707)	(4.364,0.0370)	(7.157,0.0488)	(7.678,0.0509)
	19	(4.367,0.03705)	(4.374,0.0371)	(7.168,0.0489)	(7.708,0.0510)
	27	(4.371,0.0370)	(4.400,0.0372)	(7.152,0.0489)	(7.676,0.0509)
47	17	(7.701,0.0509)	(7.689,0.0509)	(7.717,0.0510)	(7.699,0.0509)
	29	(7.670,0.0508)	(7.674,0.0510)	(7.724,0.0511)	(7.764,0.0511)
	43	(7.721,0.0509)	(7.705,0.0510)	(7.724,0.0510)	(7.716,0.0511)

Table 2. Estimation of the (α, λ) for unified HCS when T_1 is 30.

r	k	T_2			
		50	65	95	130
23	17	(1.884,0.0177)	(1.901,0.0179)	(1.899,0.0178)	(1.902,0.0179)
33	27	(3.044,0.0281)	(4.344,0.0369)	(4.389,0.0372)	(4.384,0.0371)
41	19	(2.897,0.0265)	(6.072,0.0445)	(6.902,0.0482)	(6.918,0.0483)
	25	(2.985,0.0274)	(6.065,0.0445)	(6.915,0.0483)	(6.872,0.0481)
	37	(5.812,0.0439)	(6.299,0.0457)	(6.9071,0.0457)	(6.939,0.0483)

Table 3. $MSEs$ of the (α, λ) for unified HCS when T_2 is 95.

r	k	T_1			
		4	30	75	90
19	11	$(29.935, 1.36 \times 10^{-3})$	$(29.935, 1.36 \times 10^{-3})$	$(6.829, 7.02 \times 10^{-5})$	$(7.1364, 5.46 \times 10^{-5})$
33	11	$(7.7710, 1.92 \times 10^{-4})$	$(7.8837, 1.92 \times 10^{-4})$	$(7.0425, 7.13 \times 10^{-5})$	$(7.3515, 5.63 \times 10^{-5})$
	19	$(7.8624, 1.92 \times 10^{-4})$	$(7.8426, 1.92 \times 10^{-4})$	$(6.9267, 7.00 \times 10^{-5})$	$(7.2757, 5.61 \times 10^{-5})$
	27	$(7.8810, 1.94 \times 10^{-4})$	$(7.7310, 1.89 \times 10^{-4})$	$(6.7423, 6.99 \times 10^{-5})$	$(7.1615, 5.48 \times 10^{-5})$
47	17	$(7.144, 5.28 \times 10^{-5})$	$(7.318, 5.30 \times 10^{-5})$	$(7.665, 5.52 \times 10^{-5})$	$(7.484, 5.42 \times 10^{-5})$
	29	$(7.197, 5.19 \times 10^{-5})$	$(6.850, 5.31 \times 10^{-5})$	$(7.121, 5.42 \times 10^{-5})$	$(7.597, 5.33 \times 10^{-5})$
	43	$(7.411, 5.39 \times 10^{-5})$	$(7.400, 5.42 \times 10^{-5})$	$(7.429, 5.45 \times 10^{-5})$	$(7.263, 5.37 \times 10^{-5})$

Table 4. Coverage percentages of the (α, λ) for unified HCS when T_2 is 95.

r	k	T_1			
		4	30	75	90
19	11	(0.0000,0.0000)	(0.0000,0.0000)	(0.8385,0.8154)	(0.9067,0.8810)
33	11	(0.3159,0.3388)	(0.3100,0.3332)	(0.8424,0.8211)	(0.9045,0.8789)
	19	(0.3112,0.3371)	(0.3080,0.3381)	(0.8445,0.8208)	(0.9087,0.8803)
	27	(0.3092,0.3321)	(0.3285,0.3485)	(0.8464,0.8212)	(0.9075,0.8798)
47	17	(0.9126,0.8840)	(0.9104,0.8848)	(0.9061,0.8791)	(0.9072,0.8818)
	29	(0.9116,0.8893)	(0.9127,0.8895)	(0.9167,0.8860)	(0.9140,0.8877)
	43	(0.9079,0.8859)	(0.9140,0.8838)	(0.9101,0.8845)	(0.9134,0.8857)

Table 5. $MSEs$ of the (α, λ) for unified HCS when T_1 is 30.

r	k	T_2			
		50	65	95	130
23	17	$(26.193,1.05\times10^{-3})$	$(26.014,1.03\times10^{-3})$	$(26.029,1.04\times10^{-3})$	$(26.005,1.03\times10^{-3})$
33	27	$(16.556,5.32\times10^{-4})$	$(8.029,1.99\times10^{-4})$	$(7.747,1.89\times10^{-4})$	$(7.779,1.90\times10^{-4})$
41	19	$(18.324,6.40\times10^{-4})$	$(6.304,1.16\times10^{-4})$	$(5.268,5.21\times10^{-5})$	$(4.953,5.09\times10^{-5})$
	25	$(17.542,5.91\times10^{-4})$	$(6.338,1.17\times10^{-4})$	$(5.049,5.08\times10^{-5})$	$(4.911,5.09\times10^{-5})$
	37	$(4.271,7.66\times10^{-5})$	$(5.494,8.18\times10^{-5})$	$(4.893,5.04\times10^{-5})$	$(5.084,5.17\times10^{-5})$

Table 6. Coverage percentages of the (α, λ) for unified HCS when T_1 is 30.

r	k	T_2			
		50	65	95	130
23	17	(0.0000,0.0000)	(0.0000,0.0000)	(0.0000,0.0000)	(0.0000,0.0000)
33	27	(0.0691,0.1112)	(0.3019,0.3305)	(0.3273,0.3410)	(0.3184,0.3382)
41	19	(0.0740,0.1116)	(0.6794,0.6805)	(0.8614,0.8695)	(0.8671,0.8736)
	25	(0.0815,0.1205)	(0.6854,0.6825)	(0.8669,0.8767)	(0.8649,0.8742)
	37	(0.7148,0.7472)	(0.7487,0.7432)	(0.8670,0.8772)	(0.8690,0.8716)

We have created six artificially unified HCS data sets from the above uncensored data set, using the following censoring schemes:

(1): $T_1 = 80$, $T_2 = 100$, $k = 10$, $r = 14$

(2): $T_1 = 80$, $T_2 = 100$, $k = 10$, $r = 17$

(3): $T_1 = 80$, $T_2 = 100$, $k = 10$, $r = 19$

(4): $T_1 = 65$, $T_2 = 100$, $k = 13$, $r = 18$

(5): $T_1 = 65$, $T_2 = 95$, $k = 13$, $r = 21$

(6): $T_1 = 65$, $T_2 = 85$, $k = 19$, $r = 22$

In the all cases we have estimated the unknown parameters (α, λ) using the MLEs, for computing the MLEs we have used the EM algorithm as described in Section 2, and also computed 95% confidence intervals using the observed Fisher information matrix as provided in Section 3. The results are reported in Table 7.

table 7. The MLEs, 95% confidence intervals of the α and λ for six schemes.

Schemes	(1)	(2)	(3)	(4)	(5)	(6)
α	3.3795	4.0249	4.2837	4.3425	3.9359	4.6243
	(1.106,5.653)	(1.331,6.719)	(1.439,7.129)	(1.449,7.236)	(1.315,6.557)	(1.563,7.686)
λ	0.0239	0.0271	0.0282	0.0285	0.0266	0.0297
	(0.0131,0.0347)	(0.0162,0.0380)	(0.0174,0.0391)	(0.0176,0.0395)	(0.0159,0.0374)	(0.0188,0.0406)

Based on the complete data set (without censored data) the MLEs of α and λ are 5.278 and 0.0323, respectively.

5 Conclusions

In this paper, we consider the classical inference procedure for the unified HCS GE distribution parameters. It is observed that the maximum likelihood estimates can be obtained by solving two non-linear equation, but they can not be obtained in closed form. Although, the standard Newton-Raphson algorithm can be employed to solve the non-linear equation, but we propose to use the EM algorithm to compute the MLEs. Using the missing information principle we calculate the observed Fisher information matrix for constructing the asymptotic confidence intervals of the unknown parameters. Also it is observed that for fixed r, k and T_1 when T_2 increases, and for fixed r, k and T_2 when T_1 increases, the $MSEs$ decrease and coverage percentages increase.

Acknowledgement

References

Balakrishnan, N., Rasouli, A. and Sanjari Farsipour, N. (2008). Exact likelihood inference based on an unified hybrid censored sample from the exponential distribution, *Journal of Statistical Computation and Simulation*, **78**, 475-788.

Chandrasekar, B., Childs, A., and Balakrishnan, N. (2004). Exact likelihood inference for the exponential distribution under generalized type-I and type-II hybrid censoring, *Naval Research Logistics*, **51**, 994-1004.

Dempster, A.P., Laird, N.M. and Rubin, D.B., (1977). Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc. Ser. B*, **39**, 1-39.

Gupta, R.D. and Kundu, D. (1999). Generalized exponential distribution, *Australian and New Zealand Journal of Statistics*, **41**, 173-188.

Gupta, R.D. and Kundu, D. (2001). Exponentiated exponential family; an alternative to Gamma and Weibull, *Biometr. Journal,* **33**, 117-130.

Gupta, R.D. and Kundu, D. (2007). Generalized exponential distribution: existing methods and recent developments, *Journal of the Statistical Planning and Inference,* **137**, 3537-3547.

Izanlo, M. and Habibi Rad, A. (2009). Estimating the parameters of the generalized exponential distribution Based on Unified Hybrid Censored Data, *Journal of Statistical Sciences*, Spring and Summer, **3**, 1-16.

Kundu, D. and Pradhan, B. (2009), Estimating the parameters of the generalized exponential distribution in presence of hybrid censoring, *Communications in Statistics - Theory and Methods*, **38**, 2030-2041.

Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley & sons, New York.

Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, **44**, 226-233.

Ng, T., Chan, C.S. and Balakrishnan, N. (2002). Estimation of parameters from progressively censored data using EM algorithm. *Computational Statistics & Data Analysis*, **39**, 371-386.

A. Habibi Rad
Department of Statistics,
School of Mathematical Sciences,
Ferdowsi University of Mashhad,
P. O. Box 91775-1159,
Mashhad, Iran.
email: *ahabibi@um.ac.ir*

M. Izanlo
Department of Statistics,
School of Mathematical Sciences,
Ferdowsi University of Mashhad,
P. O. Box 91775-1159,
Mashhad, Iran.

آلگوریتم EMی برای براورد پارامترهای توزیع نمایی تعمیم‌یافته تحت داده‌های سانسور شده‌ی هیبرید واحد شده

آرزو حبیبی راد و معصومه ایزانلو

دانشگاه فردوسی مشهد

چکیده. سانسور هیبرید واحد شده ترکیبی از دو سانسور هیبرید تعمیم‌یافته‌ی نوع I و II می‌باشد. هدف این مقاله، استنباط آماری در مورد پارامترهای توزیع نمایی تعمیم‌یافته تحت سانسور هیبرید واحد شده می‌باشد و از آن جا که براورد ماکسیمم درستنمایی پارامترها فرم بسته‌ای ندارند، لذا برای حل مشکل، استفاده از روش آلگوریتم EM را پیشنهاد می‌دهیم. سپس با کمک ماتریس اطلاع فیشر مشاهده‌ها، بازه‌ی اطمینان مجانبی را برای پارامترها به دست آورده و در یک مطالعه‌ی شبیه‌سازی، طرح‌های مختلف با هم مقایسه می‌شوند. در انتها با یک مثال واقعی هدف مقاله بیشتر توضیح داده می‌شود.

واژگان کلیدی. الگوریتم EM؛ ماتریس اطلاع فیشر؛ براورد ماکسیمم درستنمایی؛ سانسور هیبرید واحد شده.

نسخه‌ی کامل این مقاله، به زبان انگلیسی در صفحه‌های ۱۴۹ تا ۱۶۲ آمده است.