Noname manuscript No.
(will be inserted by the editor)

Likelihood Analysis and Stochastic EM Algorithm for Left Truncated Right Censored Data and Associated Model Selection From the Lehmann Family of Life Distributions

Debanjan Mitra · Debasis Kundu · Narayanaswamy Balakrishnan

Received: date / Accepted: date

Abstract The Lehmann family of distributions includes Weibull, Gompertz, and Lomax models as special cases, all of which are quite useful for modeling lifetime data. Analyses of left truncated right censored data from the Lehmann family of distributions is carried out in this article. In particular, the special cases of Weibull, Gompertz, and Lomax distributions are considered. Maximum likelihood estimates of the model parameters are obtained. The steps of the stochastic expectation maximization algorithm are developed in this context. Asymptotic confidence intervals for the model parameters are constructed using the missing information principle, and two parametric bootstrap approaches. Through extensive Monte Carlo simulations, performances of the inferential methods are examined. The effect of model misspecification is also assessed within this family of distributions through simulations. As it is quite important to select a suitable model for a given data, a study of model selection based on left truncated right censored data is carried out through extensive Monte Carlo simulations. Two examples based on real datasets are provided for illustrating the models and methods of inference discussed here.

Keywords Maximum likelihood estimators · Stochastic EM algorithm · Left truncation · Right censoring · Weibull distribution · Gompertz distribution · Lomax distribution · Missing information principle · Asymptotic confidence intervals · Bootstrap method · Model selection

Debanjan Mitra
Indian Institute of Management Udaipur, Rajasthan 313001, India
E-mail: debanjan.mitra@iimu.ac.in

Debasis Kundu
Indian Institute of Technology Kanpur, Uttar Pradesh 208016, India
E-mail: kundu@iitk.ac.in

Narayanaswamy Balakrishnan
McMaster University, Hamilton, Ontario L8S 4K1, Canada
E-mail: bala@mcmaster.ca

1 Introduction

Wang, Yu, and Jones (2010) considered lifetime distributions that belong to a certain family called the Lehmann family of distributions. The same family was also considered by Marshall and Olkin (2007), and they termed it as the "frailty parameter" family. The distribution function of this family is given by

$$F(t;\lambda,\alpha) = 1 - [1 - G(t;\lambda)]^{\alpha}, \quad t > 0, \tag{1}$$

where $G(\cdot;\lambda)$ is a distribution function that depends only on λ. Special members of this family are Weibull, Gompertz and Lomax distributions, all of which are commonly used to model lifetime data. Details on these special members of the family are given in Section 3.

Left truncated right censored (LTRC) data is commonly observed in survival and reliability experiments (Klein and Moeschberger 2003). Recently, Hong, Meeker, and McCalley (2009) discussed analysis of LTRC data. The structure of the LTRC data as considered by Hong, Meeker, and McCalley (2009) is quite general, in the sense that it allows units of different types - left-truncated, untruncated, censored, and a mixture of some of these - to coexist. More details on this structure is given in the next section. Motivated by their work, Balakrishnan and Mitra (2011, 2012, 2013, 2014a, 2014b) discussed likelihood inference via the expectation maximization (EM) algorithm for LTRC data from lognormal, Weibull, and gamma distributions.

In this article, we model LTRC data using the Lehmann family of distributions. The special members of the family, namely, Weibull, Gompertz and Lomax, are discussed in particular. We employ the stochastic expectation maximization (St-EM) algorithm for estimating the model parameters. We also discuss maximum likelihood estimation. Asymptotic confidence intervals are constructed using Louis' missing information principle (Louis 1982). Two parametric bootstrap approaches are also used for this purpose. Then, through Monte Carlo simulations, we assess the performance of these methods of point and interval estimation. Using simulations, we also assess the effect of model misspecification on the probability of failure of a right censored unit by a future time point. We observe that model misspecification may lead to substantially inaccurate inferential results. In view of this, we realize the need for a study on model selection.

The importance of selecting a suitable model for given data cannot be overstated. Several researchers have contributed to the relevant literature. Some of the earliest works are by Cox (1961, 1962). Since then, several classical and Bayesian methods have been proposed; see Pereira and Pereira (2016) for a comprehensive review. In particular, Marshall, Meza, and Olkin (2001) studied probability of model selection based on data originating from some special models such as Weibull, lognormal, gamma, exponential, and geometric extreme exponential, to investigate whether data can correctly identify its parent distribution. Recently, in a series of papers, Dey and Kundu (2009, 2010, 2012) have studied procedures to discriminate between some well-known lifetime models; see also Kundu (2005), and Kundu and Manglick (2004) in this regard.

Following the work of Marshall, Meza, and Olkin (2001), we study probabilities of specific models being selected for a given LTRC data. It is to be noted that very recently, Emura and Shiu (2016) have studied an AIC-based procedure for model selection based on LTRC data. However, the candidate models considered by Emura and Shiu are different than the models considered here. Also, in this paper, we discuss two methods - likelihood-based and distance-based - for model selection, and then compare their performance.

The main contributions of this paper are as follows:
(i) Modeling LTRC data using a general family of distributions like the Lehmann family. No researcher has studied LTRC data from the Lehmann family so far. Also, among the specific members of the family, to the best of our knowledge, no researcher has analyzed LTRC data from the Gompertz and Lomax distributions which are widely used in survival and reliability studies;
(ii) Use of the St-EM algorithm for parameter estimation based on LTRC data. The St-EM algorithm is a very powerful variant of the traditional EM algorithm. The St-EM algorithm avoids some of the major drawbacks of the EM algorithm while retaining its advantages;
(iii) A detailed study of model selection based on LTRC data using likelihood-based and distance-based criteria, when the candidate models belong to the Lehmann family. The problem of model selection was studied for complete data by Marshall, Meza, and Olkin (2001); not much research has been carried out towards model selection based on incomplete data. In this paper, we observe that the procedures similar to those used by Marshall, Meza, and Olkin (2001) can be used for LTRC data as well.

The rest of this paper is organized as follows. In Section 2, the structure of the data, and the notations used are provided. In Section 3, maximum likelihood estimation is discussed. The St-EM algorithm for parameter estimation based on LTRC data is detailed in Section 4, along with the construction of asymptotic confidence intervals for model parameters. A Monte Carlo simulation study for assessing the performance of the inferential methods is presented in Section 5. In this section, results of the study on model misspecification are also presented. The methodological and numerical details of the study on model selection are presented in Sections 6 and 7, respectively. In Section 8, two illustrative examples based on real datasets are provided. Finally, some concluding remarks are made in Section 9.

2 Structure of data, and notations used

The notion of LTRC data considered here follows the work of Hong, Meeker, and McCalley (2009) who analyzed life data of power transformers which were already in operation in different parts of the US. The record keeping on transformer lifetimes started in the year 1980, and continued until 2008. For machines installed before 1980, only those failures were observed that occurred after 1980 and before 2008, and installation years of those machines were also available. No information was available for machines which were installed and failed before 1980. Complete information was available on machines which were installed after 1980, and failed before 2008. However, failures after 2008 were not observed due to right censoring. Therefore, the years 1980 and 2008 can be taken as the left truncation point, and the right censoring point, respectively. This results in the LTRC structure of the data. A "left truncated" unit is a unit which was installed before 1980, but failed after 1980 - essentially indicating that the unit's lifetime is observable only if it is above a threshold. This is a fairly general structure that can accommodate different types of data. Note that, in an LTRC dataset with this structure, there may be four different types of units: left truncated and right censored, left truncated but not right censored, untruncated but right censored, and untruncated and not right censored.

Below are the notations we use in this article:
T_i: Latent lifetime of the i-th unit, $i = 1, ..., n$
C_i: Right censoring time; for the i-th unit, C_i is the time until 2008 (the right censoring point) starting from its installation year
τ_{Li}: Left truncation time; for the i-th unit, τ_{Li} is the time between its installation year and 1980 (the left truncation point), if its date of installation is before 1980. For units installed after 1980, τ_{Li} is undefined
Y_i: Observed lifetime of the i-th unit, $Y_i = \min(T_i, C_i)$, $Y_i > \tau_{Li}$
δ_i: Censoring indicator; 0 if the i-th unit is censored, 1 if not censored; that is, using the usual notation of the indicator function $I(\cdot)$, we have $\delta_i = I(T_i \leq C_i)$
ν_i: Truncation indicator; 0 if the i-th unit is truncated, 1 if not truncated
S_0, S_1: Index sets for censored units and failed units, respectively; that is, we have $S_k = \{i : \delta_i = k, i = 1, ..., n\}$, where $k = 0, 1$
R_0, R_1: Index sets for truncated and untruncated units, respectively; that is $R_k = \{i : \nu_i = k, i = 1, ..., n\}$, where $k = 0, 1$
Wei(λ, α): Weibull distribution with scale parameter α and shape parameter λ
Gom(λ, α): Gompertz distribution with scale parameter λ and shape parameter α
Lom(λ, α): Lomax distribution with scale parameter λ and shape parameter α
$\boldsymbol{\theta}$: Vector of model parameters; in this context, $\boldsymbol{\theta} = (\lambda, \alpha)$.

Let m and n_1 be the cardinalities of S_1 and T_0, respectively, i.e., $|S_1| = m$ and $|T_0| = n_1$. Finally, we assume that C_i is independent of T_i, which is the usual assumption of independent censoring (Klein and Moeschberger (2003)).

To illustrate different combinations of truncation and censoring that may occur within this general structure of LTRC data, we present simulated data in Table 1. Units 1 and 2 are untruncated, as they were installed after the left truncation point. While unit no. 2 failed in 2007, unit no. 1 was still operating at the end of the study, and hence regarded as a right censored unit. Units 3 and 4 are left truncated, as they were installed before the left truncation point

Table 1: An illustration of LTRC data

Serial no.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime
1	1994	1	*	*	0	14
2	1985	1	*	2007	1	22
3	1975	0	5	2005	1	30
4	1978	0	2	*	0	30

but their lifetimes exceeded the respective thresholds (5 and 2, respectively) which we call their left truncation times. Among these two units, unit no. 3 failed in 2005 (thus not censored), but unit no. 4 got censored at the end of the study.

3 Maximum likelihood estimation

Choosing $G(t; \lambda) = 1 - e^{-t^\lambda}$ in (1), we get the Weibull model as a special member of the family, with its distribution function as

$$F_W(t; \lambda, \alpha) = 1 - e^{-\alpha t^\lambda}, \quad t > 0.$$

We need to set $G(t; \lambda) = 1 - e^{-(e^{\lambda t}-1)}$ to get the distribution function of the Gompertz model as

$$F_G(t; \lambda, \alpha) = 1 - e^{-\alpha(e^{\lambda t}-1)}, \quad t > 0.$$

In fact, this is originally the distribution function of a Gompertz distribution that is truncated at 0. The actual Gompertz distribution has the entire real line as its support; see Gompertz (1825). This truncated Gompertz distribution, being defined on the positive side of the real line, is suitable to be used as a lifetime model. We shall refer to this truncated Gompertz distribution as the Gompertz distribution here. Finally, choosing $G(t; \lambda) = \frac{\lambda t}{1+\lambda t}$, we get the distribution function for the Lomax model as

$$F_L(t; \lambda, \alpha) = 1 - \left[\frac{1}{1 + \lambda t}\right]^\alpha, \quad t > 0.$$

Note that while λ and α are the scale and shape parameters, respectively, for the Gompertz and Lomax models, for the Weibull model they are the shape and scale parameters, respectively. It may be mentioned here that the Lomax model is also known as the Pareto model of Type-II.

The density function of the Lehmann family is given by

$$f(t; \lambda, \alpha) = \alpha g(t; \lambda)[1 - G(t; \lambda)]^{\alpha-1}, \tag{2}$$

where $g(\cdot; \lambda)$ is the density function corresponding to the distribution function $G(\cdot; \lambda)$. The corresponding hazard function is

$$h(t; \lambda, \alpha) = \alpha h_g(t; \lambda),$$

where $h_g(\cdot;\lambda)$ is the hazard function corresponding to $G(\cdot;\lambda)$, i.e., $h_g(t;\lambda) = \frac{g(t;\lambda)}{1-G(t;\lambda)}$.

The likelihood function based on observed LTRC data from this family of distributions, using the notation described in Section 2, is given by

$$L(\lambda,\alpha|DATA) = \prod_{i\in R_1}\left\{f(y_i;\lambda,\alpha)\right\}^{\delta_i}\left\{1-F(y_i;\lambda,\alpha)\right\}^{1-\delta_i} \times \prod_{i\in R_0}\left\{\frac{f(y_i;\lambda,\alpha)}{1-F(\tau_{Li};\lambda,\alpha)}\right\}^{\delta_i}\left\{\frac{1-F(y_i;\lambda,\alpha)}{1-F(\tau_{Li};\lambda,\alpha)}\right\}^{1-\delta_i},$$

which after simplification becomes

$$L(\lambda,\alpha|DATA) = \alpha^m \prod_{i\in S_1}\left\{\frac{g(y_i;\lambda)}{1-G(y_i;\lambda)}\right\}\prod_{i=1}^{n}\left\{[1-G(y_i;\lambda)]^{\alpha}\right\} \times \prod_{i\in R_0}\left\{[1-G(\tau_{Li};\lambda)]^{-\alpha}\right\}. \quad (3)$$

The log-likelihood function can be obtained from (3) to be

$$\log L(\lambda,\alpha|DATA) = m\log\alpha + \sum_{i\in S_1}\log h_g(y_i;\lambda) + \alpha\sum_{i=1}^{n}[\log\{1-G(y_i;\lambda)\} - (1-\nu_i)\log\{1-G(\tau_{Li};\lambda)\}]. \quad (4)$$

Here, $DATA$ is used in a general sense, that implies the entire information at hand in this context including the observed lifetimes, censored lifetimes (up to the point of right censoring), left truncation times, and the truncation and censoring status of units (truncated or not, censored or not).

For a known λ, solving the likelihood equation for α, we get

$$\alpha = -\frac{m}{\sum_{i=1}^{n}[\log\{1-G(y_i;\lambda)\} - (1-\nu_i)\log\{1-G(\tau_{Li};\lambda)\}]} = \alpha(\lambda), \quad (5)$$

i.e., the parameter α can be expressed in terms of λ. Using (5) in (4), we get the profile-loglikelihood for λ as

$$p(\lambda) = -m\log\left\{\sum_{i=1}^{n}[(1-\nu_i)\log\{1-G(\tau_{Li};\lambda)\} - \log\{1-G(y_i;\lambda)\}]\right\} + \sum_{i\in S_1}\log h_g(y_i;\lambda), \quad (6)$$

ignoring constants. The profile-loglikelihood $p(\lambda)$ in (6) can be maximized using any routine one-dimensional maximizer to obtain the MLE $\widehat{\lambda}$ of λ. Then, the MLE $\widehat{\alpha}$ can be easily obtained by plugging in $\widehat{\lambda}$ in (5), i.e., $\widehat{\alpha} = \alpha(\widehat{\lambda})$.

3.1 Weibull model

The log-likelihood function for the Weibull model is

$$\log L(\lambda, \alpha|DATA) = m(\log\alpha+\log\lambda)+(\lambda-1)\sum_{i\in S_1}\log y_i-\alpha\sum_{i=1}^{n}\left[y_i^{\lambda}-(1-\nu_i)\tau_{Li}^{\lambda}\right].$$

Solving the log-likelihood equation for α we have

$$\alpha = \frac{m}{w_W(\lambda)}, \tag{7}$$

where $w_W(\lambda) = \sum_{i=1}^{n}[y_i^{\lambda} - (1-\nu_i)\tau_{Li}^{\lambda}]$. The profile-loglikelihood in λ in this case becomes

$$p_W(\lambda) = -m\log w_W(\lambda) + m\log\lambda + (\lambda-1)\sum_{i\in S_1}\log y_i, \tag{8}$$

ignoring constants. Any routine one-dimensional optimizer can be used to maximize $p_W(\lambda)$ with respect to λ. Alternatively, one can setup a fixed point equation in λ equating the first derivative $p'_W(\lambda)$ to zero, and expressing λ from the resulting equation as

$$\lambda = \frac{m w_W(\lambda)}{m w'_W(\lambda) - w_W(\lambda)\sum_{i\in S_1}\log y_i} = q(\lambda), \quad \text{say}. \tag{9}$$

Then, for a prefixed $\epsilon > 0$, a simple iterative algorithm can be set up as follows. Let $\widehat{\lambda}$ and $\widehat{\alpha}$ be the MLEs of λ and α, respectively.
ALGORITHM 1:
STEP 1: Choose an initial value for λ, say $\lambda^{(0)}$;
STEP 2: Update λ by $\lambda^{(k)} = q(\lambda^{(k-1)})$, $k = 1, 2, ...$;
STEP 3: Stop at the $r-$th stage when $|\lambda^{(r)} - \lambda^{(r-1)}| < \epsilon$;
STEP 4: Set $\widehat{\lambda} = \lambda^{(r)}$;
STEP 5: Obtain $\widehat{\alpha} = \frac{m}{w_W(\widehat{\lambda})}$.

3.2 Gompertz model

For the Gompertz model, the log-likelihood equation becomes

$$\log L(\lambda, \alpha|DATA) = m(\log\alpha + \log\lambda) + \alpha(n - n_1) + \lambda\sum_{i\in S_1} y_i$$
$$-\alpha\sum_{i=1}^{n}\left[e^{\lambda y_i} - (1-\nu_i)e^{\lambda\tau_{Li}}\right].$$

The equation for α turns out to be

$$\alpha = \frac{m}{w_G(\lambda)}, \tag{10}$$

where $w_G(\lambda) = \sum_{i=1}^{n}[e^{\lambda y_i} - (1-\nu_i)e^{\lambda\tau_{Li}}] - (n-n_1)$. The profile-loglikelihood in λ, ignoring constants, is

$$p_G(\lambda) = -m\log w_G(\lambda) + m\log\lambda + \lambda\sum_{i\in S_1} y_i. \tag{11}$$

The fixed point equation in λ, in this case, becomes

$$\lambda = \frac{m w_G(\lambda)}{m w_G'(\lambda) - w_G(\lambda)\sum_{i\in S_1} y_i}. \tag{12}$$

An algorithm similar to Algorithm 1 can be set up for determining the MLEs $\widehat{\lambda}$ and $\widehat{\alpha}$ in this case as well.

3.3 Lomax model

The log-likelihood equation for the Lomax model is

$$\log L(\lambda,\alpha|DATA) = m(\log\alpha + \log\lambda) - \sum_{i\in S_1}\log(1+\lambda y_i)$$
$$-\alpha\sum_{i=1}^{n}\big[\log(1+\lambda y_i) - (1-\nu_i)\log(1+\lambda\tau_{Li})\big].$$

The equation for α is obtained as

$$\alpha = \frac{m}{w_L(\lambda)}, \tag{13}$$

where $w_L(\lambda) = \sum_{i=1}^{n}[\log(1+\lambda y_i) - (1-\nu_i)\log(1+\lambda\tau_{Li})]$. The profile-loglikelihood in λ without the constants is

$$p_L(\lambda) = -m\log w_L(\lambda) + m\log\lambda - \sum_{i\in S_1}\log(1+\lambda y_i). \tag{14}$$

The fixed point equation in λ is obtained as

$$\lambda = \frac{m w_L(\lambda)}{m w_L'(\lambda) + w_L(\lambda) v_L(\lambda)}, \tag{15}$$

where $v_L(\lambda) = \sum_{i\in S_1}\frac{y_i}{1+\lambda y_i}$. Then, following a similar process as above, one can obtain the MLEs $\widehat{\lambda}$ and $\widehat{\alpha}$.

4 Estimation using the stochastic EM algorithm

The expectation maximization (EM) algorithm is a powerful tool for analyzing incomplete data (Dempster, Laird, and Rubin 1977). However, one may encounter several issues with the EM algorithm. For example, in the expectation or E-step, the algorithm requires analytical expressions of conditional expectation of the pseudo-complete likelihood given the observed data. In many applications, such conditional expectations are analytically intractable. Another difficulty of the EM algorithm is that it has a tendency to get trapped in local optima. For a detailed discussion on the EM algorithm, its advantages and shortcomings, one may refer to McLachlan and Krishnan (2008).

The stochastic EM (St-EM) algorithm is one of the alternatives to the traditional EM algorithm. Instead of analytically obtaining the conditional expectation of the pseudo-complete likelihood function in the E-step, the St-EM algorithm imputes randomly generated observations from appropriate conditional distributions in place of the missing data. This random generation of values from appropriate conditional distributions, and the imputation of such values to obtain a pseudo-complete likelihood function is called the stochastic expectation (St-E) step. Then, in the maximization or M-step, the algorithm maximizes the pseudo-complete likelihood; see Celeux and Diebolt (1985) for details. The estimates obtained by using the St-EM algorithm are almost identical in nature with those obtained by using the EM algorithm. Moreover, the St-EM algorithm has some advantages over the EM algorithm. For example, it does not require analytical derivation of conditional expectations, or it does not get trapped in saddle points (Ng and Ye 2014). Many researchers have explored theoretical properties and applications of the St-EM algorithm; see Celeux, Chauveau, and Diebolt (1996), Chauveau (1995), Diebolt and Celeux (1993), and Nielsen (2000), for example. The use of the St-EM algorithm for LTRC data has been suggested by Bordes and Chauveau (2014), and Ng and Ye (2014).

Each stage of the St-EM algorithm consists of two steps. The steps are termed as the stochastic expectation (St-E) step, and the maximization (M) step. A typical implementation of the algorithm is as follows. At the k-th stage, in the St-E step, the conditional distribution of the i-th latent failure time given the observed data is obtained provided the i-th unit is a right censored one. If $\boldsymbol{\theta}^{(k)} = (\lambda^{(k)}, \alpha^{(k)})$ is the value of the parameter at the k-th stage, then the conditional distribution of T_i given the observed data is $f_{T_i|T_i>y_i}(t_i|t_i > y_i, \boldsymbol{\theta}^{(k)})$ provided the i-th unit, $i = 1, 2, ..., n$, is right censored. Then, a random observation from this conditional distribution is drawn to replace the lifetime of the $i-$th unit. After replacing all the right censored lifetimes by random observations generated from appropriate conditional distributions in this manner, the pseudo-complete data are then used to construct the pseudo-complete likelihood. In the M-step, the pseudo-complete likelihood is maximized to obtain the updated estimate $\boldsymbol{\theta}^{(k+1)}$. The St-E and M-steps are then iterated for a large number of cycles.

The number of stages or cycles to be considered for the St-EM algorithm is user-defined. If there are M stages, then finally one obtains a sequence of estimates $\boldsymbol{\theta}^{(k)}, k = 1, 2, ..., M$. After discarding first few, say M_1, iterations for burn-in, an average over the remaining iterations is taken to get the final estimate $\widehat{\boldsymbol{\theta}}$. As mentioned by Ye and Ng (2014), the sequence $\boldsymbol{\theta}^{(k)}$ converges to a random variable whose mean is an asymptotically efficient estimator of $\boldsymbol{\theta}$, when some suitable regularity conditions are satisfied. It may be noted here that the regularity conditions are quite general (Nielsen 2000). Ye and Ng (2014) also mention that the values of M and M_1 are to be determined from experience; usually they are taken as 1000 and 100, respectively, but larger values may be needed for complicated censoring schemes.

Now we describe the St-E and M-steps for a typical stage of the algorithm. When the latent lifetime is assumed to come from the family of distributions in (1), the conditional distribution for the i-th unit which is right censored, is given by

$$f_{T_i|T_i>y_i}(t_i|t_i > y_i, \boldsymbol{\theta}) = \frac{\alpha g(t_i;\lambda)[1 - G(t_i;\lambda)]^{\alpha-1}}{[1 - G(y_i;\lambda)]^{\alpha}}, \quad t_i > y_i. \tag{16}$$

Using conditional distributions of the form (16), one can obtain the pseudo-complete data, and the pseudo-complete likelihood function as

$$L_c(\boldsymbol{\theta}) = \prod_{i\in R_1} \{\alpha g(t_i;\lambda)[1 - G(t_i;\lambda)]^{\alpha-1}\} \times \prod_{i\in R_0} \left\{ \frac{\alpha g(t_i;\lambda)[1 - G(t_i;\lambda)]^{\alpha-1}}{[1 - G(\tau_{Li};\lambda)]^{\alpha}} \right\}. \tag{17}$$

The simplified pseudo-complete log-likelihood function is then

$$\begin{aligned}\log L_c(\boldsymbol{\theta}) = n\log\alpha + \alpha\sum_{i=1}^{n}[\log\{1 - G(t_i;\lambda)\} - (1-\nu_i)\log\{1 - G(\tau_{Li};\lambda)\}] \\ + \sum_{i=1}^{n}[\log g(t_i;\lambda) - \log\{1 - G(t_i;\lambda)\}]. \end{aligned} \tag{18}$$

Solving the log-likelihood equation for α gives

$$\alpha = -\frac{n}{\sum_{i=1}^{n}[\log\{1 - G(t_i;\lambda)\} - (1-\nu_i)\log\{1 - G(\tau_{Li};\lambda)\}]}. \tag{19}$$

Substituting (19) in (18) gives the profile log-likelihood for λ which can then be maximized to obtain updated estimate of λ. The estimate of λ is then plugged into (19) to get the updated estimate of α. With these updated estimates, we again go back to the St-E step. This process is continued for M cycles, and after discarding initial M_1 iterations for burn-in, the remaining estimates are averaged to get the final estimates of α and λ.

4.1 Weibull model

Corresponding to the i-th unit which is right censored, $i = 1, ..., n$, the conditional distribution of T_i given $T_i > y_i$ is given by

$$f_{T_i|T_i>y_i}(t_i|t_i > y_i, \boldsymbol{\theta}^{(k)}) = \frac{\alpha^{(k)}\lambda^{(k)} t_i^{\lambda^{(k)}-1} e^{-\alpha^{(k)} t_i^{\lambda^{(k)}}}}{e^{-\alpha^{(k)}\tau_{Li}^{\lambda^{(k)}}}}, \tag{20}$$

at the k-th stage of the algorithm. Conditional distributions of this type are used to generate values for replacing the censored lifetimes and the construction of the pseudo-complete likelihood function. Solving for α from the pseudo-complete log-likelihood, we get

$$\alpha = \frac{n}{w_{pcW}(\lambda)} = \alpha_{pc}(\lambda), \tag{21}$$

where $w_{pcW}(\lambda) = \sum_{i=1}^{n}[t_i^{\lambda} - (1-\nu_i)\tau_{Li}^{\lambda}]$. Substituting (21) in the pseudo-complete log-likelihood function, and equating the first derivative of the resulting profile-likelihood in λ to zero, we get

$$\lambda = \frac{n w_{pcW}(\lambda)}{n w'_{pcW}(\lambda) - w_{pcW}(\lambda)\sum_{i=1}^{n}\log t_i} = q_{pc}(\lambda), \text{say}. \tag{22}$$

Integrating the St-E and M-steps, the St-EM algorithm at the $k-$th stage is as follows:
Algorithm 2:
Step 1: Generate random values from conditional distributions of the type (20) to obtain pseudo-complete data $\{t_i^{(k)}, i = 1, 2, ..., n\}$;
Step 2: Choose initial value $\lambda_{(0)}^{(k)}$;
Step 3: Using (22), continue to update $\lambda_{(l)}^{(k)} = q_{pc}(\lambda_{(l-1)}^{(k)})$, $l = 1, 2, ...$ until convergence;
Step 4: The final value in Step 3 is the updated estimate $\lambda^{(k+1)}$;
Step 5: Using (21), calculate $\alpha^{(k+1)} = \alpha_{pc}(\lambda^{(k+1)})$;
Step 6: With the updated estimate $\boldsymbol{\theta}^{(k+1)} = (\lambda^{(k+1)}, \alpha^{(k+1)})$, go back to Step 1.

The St-E and M steps are repeated for M times, and after discarding first M_1 iterations for burn-in, the remaining estimates are averaged to get the final estimates $\widehat{\lambda}_{St-EM}$ and $\widehat{\alpha}_{St-EM}$.

4.2 Gompertz model

Proceeding in the same way as explained above, we get

$$\alpha = \frac{n}{w_{pcG}(\lambda)}, \tag{23}$$

where $w_{pcG}(\lambda) = \sum_{i=1}^{n}[e^{\lambda t_i} - (1-\nu_i)e^{\lambda\tau_{Li}}] - (n - n_1)$, and

$$\lambda = \frac{n w_{pcG}(\lambda)}{n w'_{pcG}(\lambda) - w_{pcG}(\lambda)\sum_{i=1}^{n} t_i}. \tag{24}$$

Then, a procedure analogous to the one in Section 4.1 can be used with (23) and (24) to obtain the estimates $\widehat{\alpha}_{St-EM}$ and $\widehat{\lambda}_{St-EM}$.

4.3 Lomax model

The expressions for α and λ in this case, following the same process as above, are

$$\alpha = \frac{n}{w_{pcL}(\lambda)}, \tag{25}$$

where $w_{pcL}(\lambda) = \sum_{i=1}^{n}[\log(1 + \lambda t_i) - (1-\nu_i)\log(1 + \lambda\tau_{Li})]$, and

$$\lambda = \frac{n w_{pcL}(\lambda)}{n w'_{pcL}(\lambda) + w_{pcL}(\lambda)\sum_{i=1}^{n}\frac{t_i}{1+\lambda t_i}}. \tag{26}$$

A procedure analogous to the one in Section 4.1 using (25) and (26) will give the estimates $\widehat{\alpha}_{St-EM}$ and $\widehat{\lambda}_{St-EM}$.

4.4 Asymptotic confidence intervals

4.4.1 Missing information principle

The missing information principle of Louis (1982) is widely used to obtain the asymptotic variance-covariance matrix of the estimates in conjunction with the EM algorithm. Ye and Ng (2014) applied the missing information principle successfully in the context of the St-EM algorithm. We follow their approach here.

Note that the complete lifetime data vector $\boldsymbol{t}$ can be written as

$$\boldsymbol{t} = \boldsymbol{y} \cup \boldsymbol{z},$$

where $\boldsymbol{y}$ and $\boldsymbol{z}$ are the observed data and censored data vectors, respectively. With pseudo-complete data, the pseudo-complete likelihood takes the form (17). Let $S(\boldsymbol{\theta}, \boldsymbol{t})$ denote the matrix of first derivatives of (17) with respect to $\boldsymbol{\theta}$, and $H(\boldsymbol{\theta}, \boldsymbol{t})$ denote the matrix of the negative of the second derivatives of (17) with respect to $\boldsymbol{\theta}$. Then, following Ye and Ng (2014), by the missing information principle of Louis (1982), the observed information matrix is given by

$$I(\boldsymbol{\theta}) = E[H(\boldsymbol{\theta}, \boldsymbol{t})|\boldsymbol{y}] - E[S^2(\boldsymbol{\theta}, \boldsymbol{t})|\boldsymbol{y}] + \{E[S(\boldsymbol{\theta}, \boldsymbol{t})|\boldsymbol{y}]\}^2. \tag{27}$$

To evaluate (27), we first impute R samples $\boldsymbol{z}^{(j)}, j = 1, 2, ..., R$, for the missing data $\boldsymbol{z}$ conditional on the observed data $\boldsymbol{y}$ and $\boldsymbol{\theta}$, i.e., by generating samples from conditional distributions in (16). Then, the estimated value of (27) is

$$\widehat{I(\boldsymbol{\theta})} = \frac{1}{R}\sum_{j=1}^{R} H(\boldsymbol{\theta}, \boldsymbol{t}^{(j)}) - \frac{1}{R}\sum_{j=1}^{R}[S(\boldsymbol{\theta}, \boldsymbol{t}^{(j)})]^2 + \left[\frac{1}{R}\sum_{j=1}^{R} S(\boldsymbol{\theta}, \boldsymbol{t}^{(j)})\right]^2 \Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{St-EM}}. \tag{28}$$

Finally, (28) is evaluated at the estimated value of $\boldsymbol{\theta}$ by the St-EM algorithm, and the resulting matrix is then inverted to obtain the asymptotic variance-covariance matrix of the estimates. The asymptotic confidence intervals for the model parameters can then be easily constructed using the asymptotic normality of the estimates.

4.4.2 Parametric bootstrap

Following Balakrishnan and Mitra (2012) and Kundu, Mitra, and Ganguly (2017), we also construct parametric bootstrap confidence intervals for the model parameters. For each of the specific models within the family, we follow the procedure as described here. First, for a given LTRC data, we obtain the MLE $\widehat{\boldsymbol{\theta}}$. Then, treating $\widehat{\boldsymbol{\theta}}$ as the true value of the parameter, and using the same sampling structure, we generate a data set of the same size as the original one. From the data thus generated, the MLE is obtained again, and let that be denoted by $\widehat{\boldsymbol{\theta}}_1^* = (\widehat{\lambda}_1^*, \widehat{\alpha}_1^*)$. This process is then repeated for B times, to obtain B such bootstrap estimates $\widehat{\boldsymbol{\theta}}_1^*, \widehat{\boldsymbol{\theta}}_2^*, ..., \widehat{\boldsymbol{\theta}}_B^*$. Then, a $100(1-\gamma)\%$ parametric bias-corrected bootstrap confidence interval for a model parameter, say α, is obtained as

$$(\hat{\alpha} - b_\alpha - z_{\gamma/2}\sqrt{v_\alpha}, \hat{\alpha} - b_\alpha + z_{\gamma/2}\sqrt{v_\alpha}),$$

where b_α and v_α are the bootstrap bias and variance, respectively, for the parameter α, and z_γ is the upper $\gamma-$percentile point of the standard normal distribution. The bias corrected bootstrap confidence intervals for λ can be obtained similarly. The number B should be reasonable to get a clear idea about the distribution of the estimates. For practical purposes, B may be taken to be 1000 in general, for example.

One can also obtain percentile bootstrap confidence intervals by choosing appropriate percentile points from the ordered values of the bootstrap estimates of the parameters. For parameter α, a $100(1-\beta)\%$ percentile bootstrap confidence interval can be given as $(\widehat{\alpha}^*_{([B\beta/2])}, \widehat{\alpha}^*_{([B(1-\beta/2)])})$, where $\widehat{\alpha}^*_{(1)}, \widehat{\alpha}^*_{(2)}$, ..., $\widehat{\alpha}^*_{(B)}$ are the ordered bootstrap estimates of the parameter α, with $[x]$ indicating the greatest integer contained in x.

5 Simulation study

The Monte Carlo simulations are carried out using the R software. The process of generating LTRC data from Weibull distribution is described here. Similar processes can be followed for the Gompertz and Lomax distributions.

To simulate LTRC lifetime data, we start with two arbitrary sets of years as the possible starting points of the units, i.e., the installation years: one of the sets corresponds to the truncated group while the other corresponds to the untruncated group. We choose the year 2000 as the left truncation point without loss of generality. Naturally, all the years in the set corresponding to the truncated group precede the left truncation point, while all the years in the other set succeed the left truncation point. For this simulation study, without any loss of generality, the sets of years corresponding to the truncated and untruncated groups are chosen to be {1995,1996,...,1999 }, and {2000,...,2003}, respectively. For each unit in the dataset, a random observation from one of these sets, depending on whether the unit is truncated or not, is generated, and then a randomly drawn observation from Weibull distribution is added to it. Thus, we obtain the failure year of a unit. If the failure year of a unit from the truncated group turns out to be less than 2000 (i.e., before the left truncation point), that observation is completely discarded and a new installation year as well as lifetime are generated for that unit until its failure year becomes larger than 2000. This incorporates left truncation in the data. Different percentages of left truncation for units are used in the simulation study. To change the percentage of left truncated units, the sampling proportions from the truncated and untruncated groups are changed. If the failure year of a unit is after the end of the study (i.e., after the right censoring point), the unit is marked as a right censored one, otherwise it is considered to as an observed failure. The right censoring point is taken as 2004 without loss of generality.

The lifetime distributions studied here are Weibull, Gompertz and Lomax. The model parameters of these distributions are so chosen that the censoring percentages and mean lifetimes of units are almost identical for all the three models. This helps us in comparing the results across these distributions. Three percentages for truncation are used, to observe the effect of moderate and heavy truncation on the results. The ML estimation based on observed likelihood is referred to as Method 1 (M1), and the approach based on St-EM algorithm is referred to as Method 2 (M2). These inferential methods are examined with respect to bias and MSE of the estimates. The asymptotic confidence intervals corresponding to missing information principle, and parametric bootstrap approaches are examined with respect to their coverage probability.

From Tables 2 - 4, we note that the methods (M1 and M2) are in close agreement with respect to bias and MSE of estimates. The performance of the methods is quite satisfactory for Weibull and Gompertz distributions. However, for Lomax distribution, the MSE for the scale parameter is quite high for both methods. This can be attributed to the large variance of the lifetimes generated from Lomax distribution; when the data have both truncation and censoring, this makes the estimation of λ somewhat difficult especially for small samples, sometimes the algorithms may not converge to the desired point. However, with increasing sample size, this issue reduces by a good extent as the effective sample size (i.e, number of units in the sample with neither truncation nor censoring) increases.

Table 2: Performance of point and interval estimates for Weibull distribution; $(\lambda, \alpha) = (0.500, 0.707)$ for different truncation percentages (TP). Nominal confidence level is 95%. Coverage probabilities are for CIs corresponding to missing information principle (MI), bias-corrected parametric bootstrap (BB), and percentile bootstrap (PB) approaches.

$n = 50$								
TP	Parameter	Bias(M1)	MSE(M1)	Bias(M2)	MSE(M2)	Cov. Prob.		
						MI	BB	PB
0.2	λ	0.007	0.006	0.018	0.007	0.958	0.962	0.946
	α	0.007	0.018	0.009	0.018	0.925	0.938	0.940
0.4	λ	0.012	0.007	0.023	0.007	0.954	0.962	0.951
	α	0.017	0.022	0.015	0.027	0.921	0.939	0.925
0.7	λ	0.022	0.012	0.039	0.014	0.926	0.966	0.926
	α	-0.012	0.040	-0.023	0.041	0.868	0.898	0.902
$n = 100$								
TP	Parameter	Bias(M1)	MSE(M1)	Bias(M2)	MSE(M2)	Cov. Prob.		
						MI	BB	PB
0.2	λ	0.007	0.003	0.012	0.003	0.945	0.946	0.949
	α	0.003	0.008	0.004	0.008	0.933	0.944	0.947
0.4	λ	0.005	0.003	0.011	0.003	0.968	0.971	0.961
	α	0.004	0.011	0.004	0.011	0.928	0.939	0.939
0.7	λ	0.011	0.005	0.014	0.005	0.924	0.942	0.920
	α	-0.003	0.021	-0.010	0.021	0.898	0.936	0.940

The coverage probabilities of all three asymptotic confidence intervals are quite close to the nominal level in the case of Weibull and Lomax distributions. However, for Gompertz distribution, the confidence intervals, especially the one corresponding to missing information principle, suffer from under coverage. The problem is rectified to some extent for larger sample size (100), but in general, parametric bootstrap confidence intervals perform better than the missing information principle for this model. Note that the method of constructing confidence intervals based on missing information principle is applicable when EM-based parameter estimation is considered. When model parameters are estimated by MLEs using the observed likelihood, confidence intervals may be obtained by using a bootstrap approach which can also be used along with EM-based estimation techniques.

The truncation percentage does not seem to play an important role as far as the bias, MSE, and coverage probability are concerned, unless it is very heavy (70%). For 70% truncation, while bias, MSE, and coverage probability for all models are somewhat affected, the effect is maximum for the Lomax distribution for small sample ($n = 50$).

5.1 Effect of model misspecification

Suppose the lifetimes of LTRC data are originally from a Weibull distribution, but the Gompertz distribution is used as a model for the data. How would it

Table 3: Performance of point and interval estimates for Gompertz distribution; $(\lambda, \alpha) = (0.5, 0.3)$ for different truncation percentages (TP). Nominal confidence level is 95%. Coverage probabilities are for CIs corresponding to missing information principle (MI), bias-corrected parametric bootstrap (BB), and percentile bootstrap (PB) approaches.

TP	Parameter	Bias(M1)	MSE(M1)	Bias(M2)	MSE(M2)	Cov. Prob. MI	Cov. Prob. BB	Cov. Prob. PB
$n = 50$								
0.2	λ	0.034	0.023	0.053	0.024	0.785	0.952	0.922
	α	0.029	0.054	0.017	0.045	0.770	0.879	0.929
0.4	λ	0.033	0.017	0.045	0.017	0.864	0.962	0.926
	α	0.020	0.043	0.015	0.041	0.831	0.901	0.935
0.7	λ	0.028	0.018	0.034	0.018	0.870	0.960	0.902
	α	0.031	0.054	0.032	0.053	0.800	0.852	0.906
$n = 100$								
0.2	λ	0.017	0.009	0.026	0.009	0.814	0.954	0.925
	α	0.007	0.014	0.002	0.014	0.803	0.902	0.918
0.4	λ	0.021	0.009	0.025	0.009	0.851	0.957	0.933
	α	0.004	0.018	0.002	0.018	0.841	0.908	0.936
0.7	λ	0.014	0.009	0.017	0.009	0.914	0.950	0.944
	α	0.017	0.022	0.016	0.022	0.890	0.924	0.948

Table 4: Performance of point and interval estimates for Lomax distribution; $(\lambda, \alpha) = (4, 0.5)$ for different truncation percentages (TP). Nominal confidence level is 95%. Coverage probabilities are for CIs corresponding to missing information principle (MI), bias-corrected parametric bootstrap (BB), and percentile bootstrap (PB) approaches.

TP	Parameter	Bias(M1)	MSE(M1)	Bias(M2)	MSE(M2)	Cov. Prob. MI	Cov. Prob. BB	Cov. Prob. PB
$n = 50$								
0.2	λ	0.708	13.402	0.568	12.937	0.924	0.920	0.953
	α	0.054	0.046	0.114	0.082	0.978	0.951	0.941
0.4	λ	0.683	10.972	0.591	10.773	0.913	0.913	0.948
	α	0.046	0.031	0.089	0.047	0.975	0.951	0.950
0.7	λ	1.357	29.477	1.285	29.851	0.926	0.932	0.956
	α	0.034	0.035	0.036	0.054	0.936	0.954	0.958
$n = 100$								
0.2	λ	0.484	5.066	0.421	4.992	0.928	0.931	0.955
	α	0.015	0.016	0.038	0.021	0.971	0.945	0.950
0.4	λ	0.406	4.603	0.366	4.619	0.913	0.909	0.946
	α	0.016	0.013	0.033	0.015	0.969	0.948	0.941
0.7	λ	0.622	8.133	0.587	8.150	0.910	0.928	0.946
	α	0.011	0.012	0.024	0.014	0.956	0.940	0.934

Table 5: Performance of estimated probabilities with respect to bias and MSE for all three models when the data is generated from Weibull distribution

t^*	ζ (True Value)	$\widehat{\zeta}$(Weibull)		$\widehat{\zeta}$(Gompertz)		$\widehat{\zeta}$(Lomax)	
		Bias	MSE	Bias	MSE	Bias	MSE
4	0.339	0.006	0.002	0.230	0.056	-0.080	0.010
6	0.519	0.007	0.004	0.301	0.093	-0.139	0.026
8	0.632	0.006	0.005	0.295	0.088	-0.178	0.040

affect the inferences? We now assess the effect of model misspecification in this context.

After we fit a model to a given LTRC data, a quantity of practical interest is the probability of failure for a right censored unit by a future time point. Consider a right censored unit that may or may not be left truncated. The probability that the unit will fail by a future time point t^* is that $\zeta = P(T < t^*|T > c) = \zeta(\lambda, \alpha)$, where c is the right censoring time and $t^* > c$. Clearly,

$$\zeta = P(T < t^*|T > c) = \frac{F(t^*) - F(c)}{1 - F(c)}, \quad t^* > c. \tag{29}$$

We can estimate ζ by $\widehat{\zeta} = \zeta(\widehat{\lambda}, \widehat{\alpha})$. Through a Monte Carlo simulation study, we assess the effect of model misspecification on ζ. It may be mentiond here that this type of prediction problem is known as dynamic prediction in the clinical context. Interested readers may refer to Emura, Matsui, and Rondeau (2019), and Shinohara et al. (2020) for survival prediction of cancer patients.

In this context, we consider the cases in which the data are originally from one of the three models in the Lehmann family. For example, consider the case where the data are generated from Weibull distribution. We fit all three candidate models to this data, and estimate their parameters. Then, we estimate ζ for all three candidate models. Similarly, we also consider the cases in which the data are originally generated from the Gompertz or Lomax models.

In relation to the simulation structure presented above, we consider a unit that is installed in the year 2002, without loss of generality. If the unit is right censored in 2004, we would like to estimate the probability of its failure by 2006, 2008, and 2010. Here, $t^* = 4, 6$, and 8, respectively, and $c = 2$. So, the probabilities we would like to estimate are

$$\zeta_1 = P(T < 4|T > 2),\ \zeta_2 = P(T < 6|T > 2),\ \text{and}\ \ \zeta_3 = P(T < 8|T > 2).$$

We simulate LTRC Weibull data of size 100 and truncation percentage 20 with parameters same as in Table 1, i.e., with $(\lambda, \alpha) = (0.500, 0.707)$. For a unit in this dataset that is installed in 2002 and is right censored, we obtain the true values of ζ_1, ζ_2, and ζ_3. Table 5 gives the bias and MSE of the estimated probabilities for all three models.

Next, we simulate LTRC Gompertz data of size 100 and truncation percentage 20. The parameters used for data generation are $(\lambda, \alpha) = (0.5, 0.3)$,

Table 6: Performance of estimated probabilities with respect to bias and MSE for all three models when the data is generated from Gompertz distribution

t^*	ζ (True Value)	$\widehat{\zeta}$(Weibull)		$\widehat{\zeta}$(Gompertz)		$\widehat{\zeta}$(Lomax)	
		Bias	MSE	Bias	MSE	Bias	MSE
4	0.753	-0.038	0.004	0.003	0.003	-0.259	0.068
6	0.994	-0.047	0.003	-0.003	0.000	-0.262	0.070
8	0.999	-0.008	0.000	-0.000	0.000	-0.147	0.022

Table 7: Performance of estimated probabilities with respect to bias and MSE for all three models when the data is generated from Lomax distribution

t^*	ζ (True Value)	$\widehat{\zeta}$(Weibull)		$\widehat{\zeta}$(Gompertz)		$\widehat{\zeta}$(Lomax)	
		Bias	MSE	Bias	MSE	Bias	MSE
4	0.272	0.098	0.012	0.292	0.088	0.006	0.002
6	0.400	0.161	0.030	0.415	0.174	0.007	0.004
8	0.478	0.198	0.044	0.445	0.199	0.007	0.005

i.e., the same as in Table 2. Table 6 gives the true value of the probabilities ζ_1, ζ_2, and ζ_3, as well as the bias and MSE of the estimated probabilities for all three models.

The simulated LTRC Lomax datasets are each of size 100 and have the truncation percentage as 20. The true parameter values are as in Table 3, i.e., $(\lambda, \alpha) = (4, 0.5)$. The values of ζ_1, ζ_2, and ζ_3, and the bias and MSE of the estimated probabilities are given in Table 7.

As expected, in Tables 5 - 7, we observe that the effect of model misspecification could be quite severe on the inferences. As the distance between the right censoring point and the future time point increases, the effect of model misspecification magnifies. This clearly shows that it is of utmost importance to select an appropriate model for a given LTRC data, in order to achieve accuracy in associated inferences.

6 Model selection within the family

A natural question that arises while analysing a LTRC dataset is: Which model to select? This, in general, is one of the most important problems in parametric inference, and it has been addressed by many researchers in different ways. One may refer to Pereira and Pereira (2016) for a comprehensive coverage of relevant literature. In this paper, following the approach of Marshall, Meza, and Olkin (2001) for complete data, we address the problem of model selection for LTRC data. We restrict our candidate models to the members of the Lehmann family.

For a given LTRC data from one of the three models belonging to the Lehmann family, we obtain Monte Carlo probabilities of selecting each of the candidate models as the best fit. This serves the following main purposes:

First, this tells us how often an LTRC dataset can correctly identify its parent distribution within the Lehmann family. Second, if there are more than one methods for model selection, we can compare their performances. Third, we get some idea about the sample size required for correct identification. Fourth, it provides us with a procedure to follow for selecting an appropriate model for a given LTRC data, at least within the Lehmann family.

6.1 A likelihood-based method

A straightforward approach would be to choose the model that gives the largest maximum likelihood for a given LTRC data. This method was used for complete samples by Marshall, Meza, and Olkin (2001); see also Cox (1961). Thus, when choosing between three candidate models, if $\widehat{L}_W(\widehat{\lambda}, \widehat{\alpha}|DATA)$, $\widehat{L}_G(\widehat{\lambda}, \widehat{\alpha}|DATA)$, and $\widehat{L}_L(\widehat{\lambda}, \widehat{\alpha}|DATA)$ are the maximized likelihoods for Weibull, Gompertz and Lomax models, respectively, then we choose that model as the best fit which has the maximized likelihood equal to

$$\text{MAX}\{\widehat{L}_W(\widehat{\lambda}, \widehat{\alpha}|DATA), \widehat{L}_G(\widehat{\lambda}, \widehat{\alpha}|DATA), \widehat{L}_L(\widehat{\lambda}, \widehat{\alpha}|DATA)\},$$

where the general likelihood function for LTRC data from the Lehmann family is given in (3).

It may be mentioned here that the likelihood-based method as described here can be used when all the candidate models have the same number of parameters. Otherwise, this method would bias the selection procedure towards the model with larger number of parameters compared to others. In such cases, one may consider an information-based criterion such as Akaike's information criterion (AIC), and may select the model with minimum AIC among all candidate models. As all the candidate models considered here have same number of parameters, the likelihood-based selection process would give identical results with an AIC-based selection process.

6.2 A minimum distance method

In developing a minimum distance-based method, in principle we follow Taylor and Jakeman (1985), which was also followed by Marshall, Meza, and Olkin (2001) for complete data. As described below, the method is a hybrid one, using maximum likelihood and Kolmogorov distance.

It is necessary to use a non-parametric estimator of the survival function based on LTRC data. The product-limit estimator of the survival function based on LTRC data was proposed by Tsai, Jewell, and Wang (1987) as follows: Let $y_{(1)}, y_{(2)}, ..., y_{(k)}$ be the distinct, ordered, observed failure times. Then the product-limit estimator of the survival function S of the lifetime variable is

$$\hat{S}_n(y) = \prod_{y_{(j)}<y} \frac{n_j - d_j}{n_j}, \tag{30}$$

for $y > y_{(1)}$, and $\hat{S}_n(y) = 1$, for $y \leq y_{(1)}$. Here, d_j denotes the number of failures at time point $y_{(j)}$, and n_j is the cardinality of the risk set at time point $y_{(j)}$, i.e., $n_j = \sum_{i=1}^{n} I(\tau_{Li} \leq y_{(j)} \leq y_i)$, with $I(\cdot)$ being the usual indicator function. This product-limit estimator for LTRC data reduces to the usual Kaplan-Meier estimator (Kaplan and Meier 1958) in case of no truncation. Also, $\widehat{S}_n(y)$ converges to the true survival function of the lifetime variable (Gijbels and Wang 1993, Tse 2003).

For LTRC data generated from a model in the Lehmann family, we first obtain the non-parametric estimate of the survival function using (30). Then, the Kolmogorov distance for each of the candidate models is calculated as

$$\mathrm{K}_{\mathrm{DIST}} = \sup_{y} |\widehat{F}_n(y) - \widehat{F}(y)|, \tag{31}$$

where $\widehat{F}_n(y) = 1 - \widehat{S}_n(y)$ is the nonparametric estimator of the CDF, and $\widehat{F}(\cdot)$ is the estimated CDF of the candidate model. We obtain $\widehat{F}(\cdot)$ for the candidate model by using the MLEs. Finally, the candidate model with minimum Kolmogorov distance is chosen as the best fit. It may be noted here that in case of heavy truncation, the product-limit estimator in (30) may have consistency issues. In the present work, we have not observed this issue in the simulation study for the two truncation percentages (20 and 40). However, we recommend that in case of heavy truncation in the data, one should use the likelihood-based method for the purpose of model selection.

7 Simulation study for model selection

The numerical study carried out is as follows: A LTRC dataset is first generated from one of the models - Weibull, Gompertz, or Lomax. This is the parent distribution of the data. Then, fitting all the candidate models including the parent one to the dataset, the model with largest maximum likelihood is chosen as the best fit. Finally, repeating this process 10,000 times and calculating the proportion of times each model is chosen as the best fit, we obtain the Monte Carlo probabilities for each of the candidate models to be selected. A similar process is followed for the minimum Kolmogorov distance criterion as well. Through simulations, we first fit two candidate models including the parent one to the data, and then fit all three candidate models. This allows us to explore the relative richness of models while fitting them to LTRC data. We use samples of sizes 20, 30, 40, 50, 60, 100, 150, 200 and 300 for this study. For the study on model selection, we use the same process of data generation as described in Section 5.

7.1 Weibull data

When the parent model is the Weibull, and the candidate models are Weibull and Gompertz, even for very small samples Weibull model has a very high

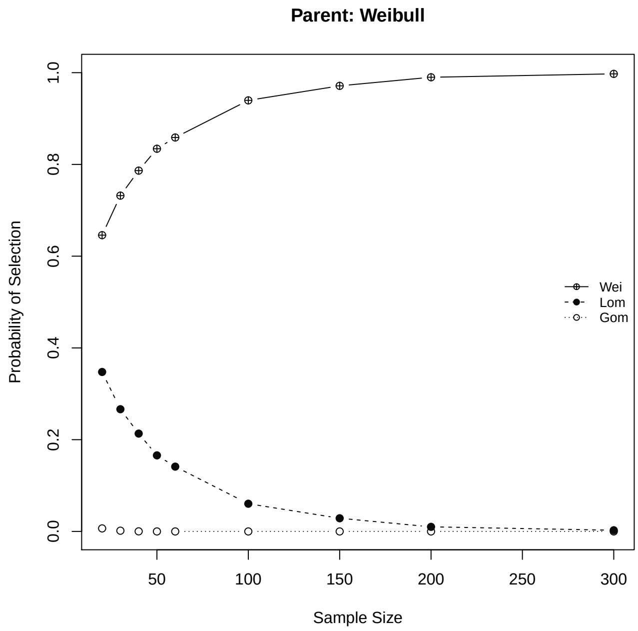

Fig. 1: Probability of selecting a model for the likelihood-based approach when the parent of the data is Weibull distribution

probability of being selected as the best fit. However, when the candidate models are Weibull and Lomax, the probability of selecting Lomax as the best fit is moderate for very small samples. With increase in sample size the chance for Lomax to be selected reduces, and Weibull is almost always chosen as the best model for samples larger than 150.

From Figures 1 and 2, we notice that when all three candidate models are fitted to the LTRC data with Weibull lifetimes, the probability of selecting the Gompertz model is almost negligible for the likelihood-based method while it is small for the distance-based method. Lomax, on the other hand, has a moderate probability of being selected, though it is significantly less than that for the Weibull distribution even for moderate sample sizes.

7.2 Gompertz data

When the Gompertz is the parent model, and the candidates are Gompertz and Weibull, the latter has a moderate probability of being selected, especially for small samples. But when the candidate models are Gompertz and Lomax, the former always gets selected as the latter has insignificant probability of being selected even for very small samples.

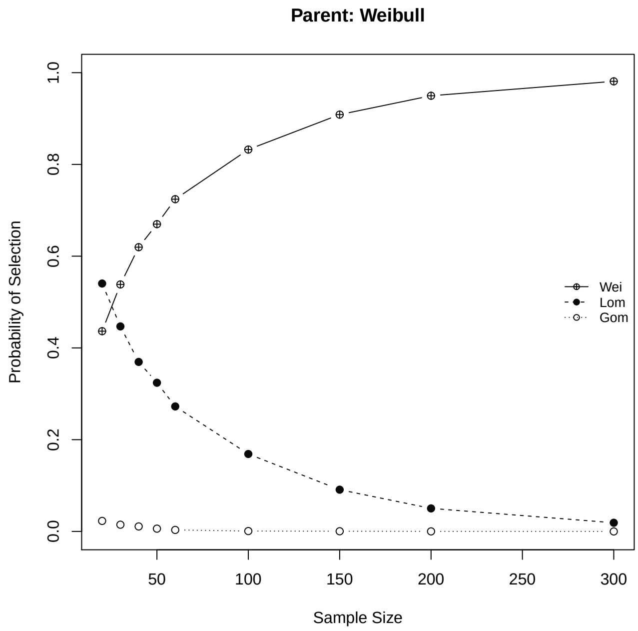

Fig. 2: Probability of selecting a model for the minimum distance-based approach when the parent of the data is Weibull distribution

Figures 3 and 4 show the results when all three candidate models are fitted to Gompertz data. Lomax does not have much chance of being selected as the best model, while Weibull is more often selected by the minimum distance-based method than the likelihood-based method. Clearly, the likelihood-based method chooses Gompertz as the best model much more successfully, while the distance-based method gets confused between choosing Weibull and Gompertz unless the sample size is large.

7.3 Lomax data

When the Lomax is the parent model, in the two-model setup, Gompertz model does not have much chance of being selected, while Weibull model is selected only in a very small number of cases. We see from Figures 5 and 6 that when all three models are candidates, Gompertz model has no chance of being selected for samples of size more than 40, while the chance is negligible for even smaller samples. Weibull model has a small chance of being selected for very small samples, but with increase in sample size, the probability drastically decreases. In this case, both likelihood-based and distance-based methods seem to perform closely.

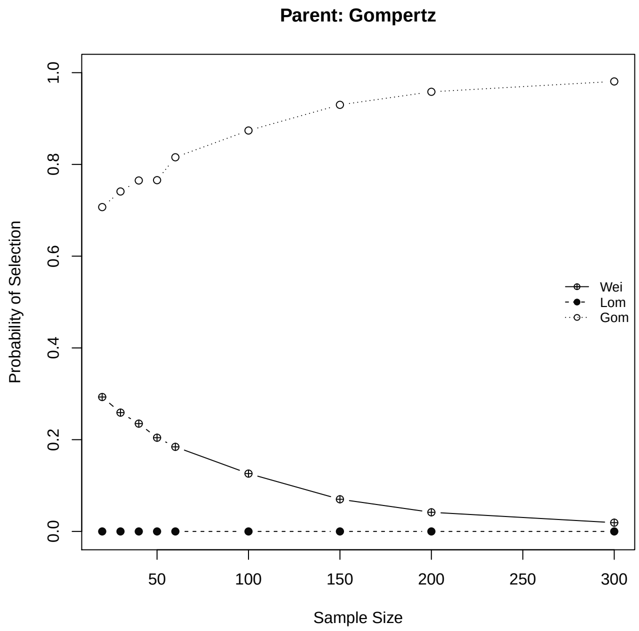

Fig. 3: Probability of selecting a model for the likelihood-based approach when the parent of the data is Gompertz distribution

From the above simulation results, it may be noted that for modeling LTRC data, within the Lehmann family, Weibull seems to be the most versatile model. The Gompertz model, on the other hand, seems to be least versatile in this regard. The Lomax and the Gompertz models seem to be far apart in nature, as both of them have negligible chance of being selected when the parent model of the LTRC data is the other one. Between the two methods, the likelihood-based method performs better for LTRC data compared to the distance-based method; this was observed by Marshall, Meza, and Olkin (2001) also for complete samples.

Thus, we believe it will not be unreasonable to propose the likelihood-based method for selecting an appropriate model for a LTRC data, at least when the underlying lifetime distribution belongs to the Lehmann family. To a given LTRC data, one would fit all the candidate models belonging to the Lehmann family, and would select that model having the largest maximum likelihood value. As we have observed through the Monte Carlo simulations, this method has a very high chance of selecting the correct distribution for the given data. Naturally, the method performs very well for larger samples. But even for small samples, its performance is observed to be satisfactory.

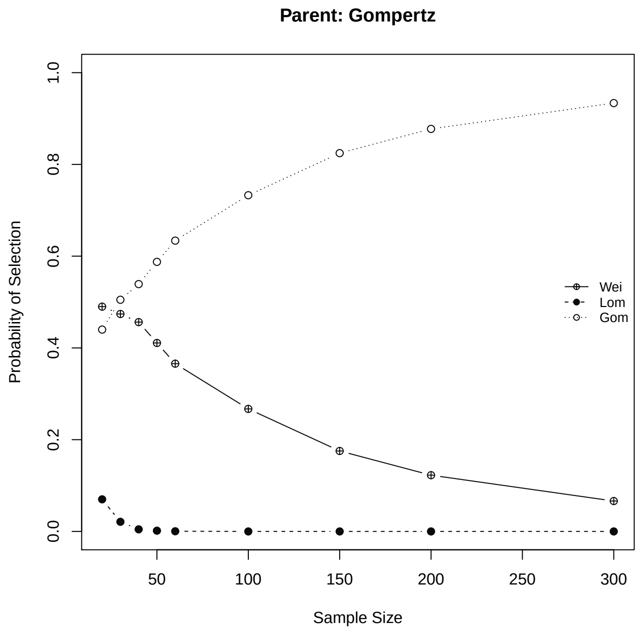

Fig. 4: Probability of selecting a model for the minimum distance-based approach when the parent of the data is Gompertz distribution

8 Illustrative examples

8.1 Channing House Data

The first illustrative example is based on a real dataset which is popularly known as the Channing House data. The Channing House data is regarding lifetimes of residents at a retirement centre in Palo Alto, California. There were 97 men, and 365 women in the retirement centre, and the data were collected from the year 1965 to 1975.5 (the middle of 1975). Some residents died before the data collection ended, while some were still alive at the end of the study. A restriction on the admission to the centre was that the age of an individual had to be at least 60 years. Different residents entered the centre at different ages (after the age of 60). Clearly, all the observations are left truncated here, while some are right censored. Table 8 gives a summary of this data.

We first change the origin and scale of the data, by subtracting 720 (number of months) from each of the lifetimes (and truncation times), and by dividing each of the lifetimes (and truncation times) by 100. The change of origin and scale will not affect the inferential results in any way. All the three candidate models are fitted to the data. The model parameters are estimated by the St-

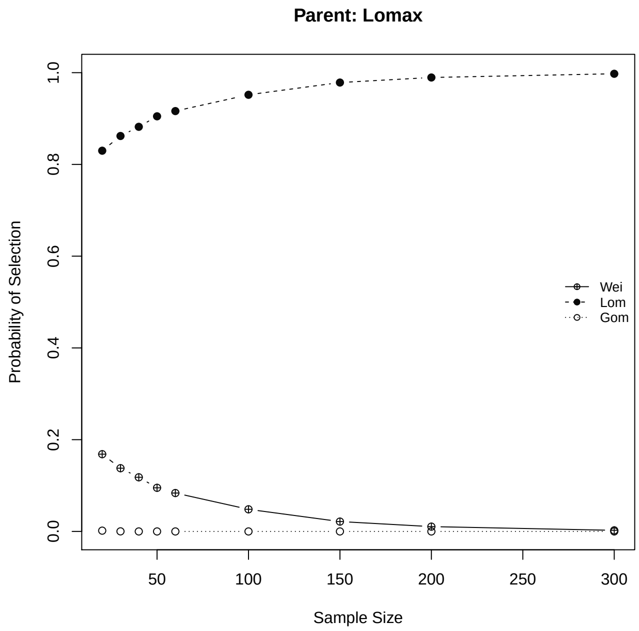

Fig. 5: Probability of selecting a model for the likelihood-based approach when the parent of the data is Lomax distribution

Table 8: A descriptive summary of the Channing House Data

Group	Size	Observed Failures	Average lifetime (Months)	Std. Dev. lifetimes (Months)
Male	97	46	991.58	73.71
Female	365	130	984.51	73.81
Combined	462	176	985.99	73.78

EM algorithm, and by the maximum likelihood method. The numerical results of model fitting are provided in Table 9.

For selecting an appropriate model for this data, we compute the maximized log-likelihoods values and Kolmogorov distances as below:

$$\log \widehat{L}_W(\widehat{\lambda}, \widehat{\alpha}|DATA) = -277.960,$$
$$\log \widehat{L}_G(\widehat{\lambda}, \widehat{\alpha}|DATA) = -274.817,$$
$$\log \widehat{L}_L(\widehat{\lambda}, \widehat{\alpha}|DATA) = -310.977$$

and

$$\mathrm{K}^W_{\mathrm{DIST}} = 0.250, \quad \mathrm{K}^G_{\mathrm{DIST}} = 0.234, \quad \mathrm{K}^L_{\mathrm{DIST}} = 0.279.$$

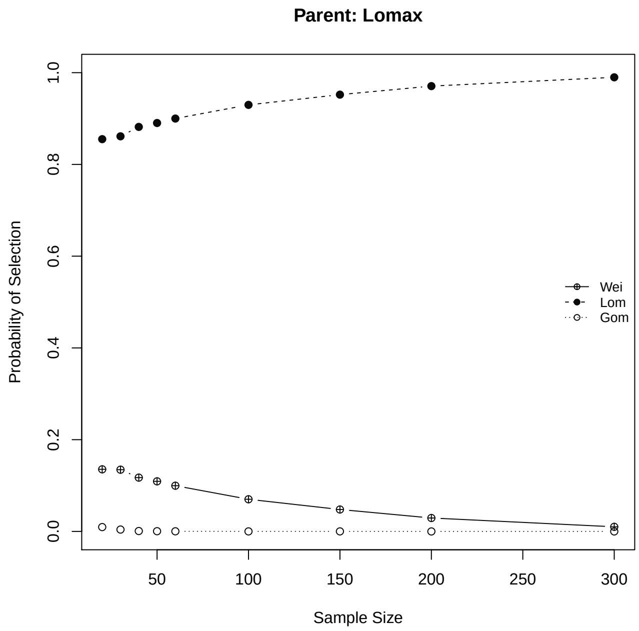

Fig. 6: Probability of selecting a model for the minimum distance-based approach when the parent of the data is Gompertz distribution

Table 9: Point and interval estimates for model parameters based on the Channing House Data; Method 1 is ML estimation based on observed likelihood and Method 2 is the St-EM algorithm

Fitted Model	Method 1	Method 2	95% Asymptotic CI
Weibull	$\widehat{\lambda} = 2.911$ $\widehat{\alpha} = 0.031$	$\widehat{\lambda} = 2.917$ $\widehat{\alpha} = 0.031$	(2.619, 3.214) (0.017, 0.044)
Gompertz	$\widehat{\lambda} = 0.787$ $\widehat{\alpha} = 0.083$	$\widehat{\lambda} = 0.799$ $\widehat{\alpha} = 0.082$	(0.738, 0.860) (0.058, 0.105)
Lomax	$\widehat{\lambda} = 0.050$ $\widehat{\alpha} = 10.527$	$\widehat{\lambda} = 0.050$ $\widehat{\alpha} = 10.599$	(0.041, 0.058) (10.543, 10.655)

As observed from the above results, the Gompertz model turns out to be the most appropriate model among the three candidate models for the Channing House data as it has the largest maximum likelihood as well as the smallest Kolmogorov distance.

Table 10: Summary of Power Transformer Data

Group	Size	Percentage	Average lifetime (Years)	Std. Dev. lifetimes (Years)
Right censored	247	86.36	31.15	18.77
Observed failures	39	13.64	22.00	12.67
Combined	286		29.90	18.31
Left truncated	167	58.39	43.49	9.33
Untruncated	119	41.61	10.83	7.71
Combined	286		29.90	18.31

Table 11: Model Selection for Power Transformer Data

Likelihood-based		Distance-based	
Maximized Log-likelihood	Value	Kolmogorov Distance	Value
$\log \widehat{L}_W(\widehat{\lambda}, \widehat{\alpha}\|DATA)$	-234.133	$\text{K}_{\text{DIST}}^{W}$	0.324
$\log \widehat{L}_G(\widehat{\lambda}, \widehat{\alpha}\|DATA)$	-236.401	$\text{K}_{\text{DIST}}^{G}$	0.333
$\log \widehat{L}_L(\widehat{\lambda}, \widehat{\alpha}\|DATA)$	-233.593	$\text{K}_{\text{DIST}}^{L}$	0.337

8.2 Power Transformer Data

Hong, Meeker, and McCalley (2009) analysed the LTRC data pertaining to the lifetimes of power transformers. Though the full dataset was not available, the authors diagrammatically presented a subset of the data. This subset was later analysed by Emura and Shiu (2016); these authors presented the data in a table, by reading off the values from the diagram presented in Hong, Meeker, and McCalley (2009). In this article, we analyse the LTRC power transformer data presented in Emura and Shiu (2016).

The entire LTRC dataset on lives of power transformers had a total of 710 lifetimes (observed failures and right censored units). In the subset, there are 286 lifetimes out of which 39 are observed failures and the rest are right censored observations. Table 9 gives a summary of the dataset that we have analysed here.

We fit the three candidate models to the power transformer data, and with the estimated values of model parameters, we obtain the numerical values of the model selection criteria. Table 10 provides the details of the model selection criteria.

Note that the likelihood-based approach indicates Lomax to be the best fit, while the distance-based approach selects Weibull to be the best model. From the simulation results of the study on model selection, this confusion is not unexpected. In fact, when the parent distribution is Lomax and the sample size is large (greater than 250, say), the likelihood-based approach would almost surely identify the parent, but the distance-based approach may still choose Weibull as the best fit; refer to Figures 3 and 6 in this regard. Hence, in

Table 12: Point and interval estimates of model parameters for Power Transformer Data; Method 1 is ML estimation based on observed likelihood and Method 2 is the St-EM algorithm

Fitted Model	Method 1	Method 2	95% Asymptotic CI
Weibull	$\widehat{\lambda} = 0.903$ $\widehat{\alpha} = 0.010$	$\widehat{\lambda} = 0.923$ $\widehat{\alpha} = 0.010$	(0.825, 1.021) (0.005, 0.015)
Gompertz	$\widehat{\lambda} = 0.012$ $\widehat{\alpha} = 0.426$	$\widehat{\lambda} = 0.011$ $\widehat{\alpha} = 0.435$	(0.009, 0.013) (0.434, 0.436)
Lomax	$\widehat{\lambda} = 0.040$ $\widehat{\alpha} = 0.311$	$\widehat{\lambda} = 0.039$ $\widehat{\alpha} = 0.318$	(0, 0.084) (0.117, 0.519)

this case, we choose the Lomax model for the LTRC power transformer data. Table 12 provides the estimates of model parameters corresponding to the observed likelihood method and St-EM algorithm. Clearly, both methods are in agreement with respect to the estimation of model parameters. The lower limit of the 95% confidence interval for λ in case of Lomax model turned out to be negative, which was replaced by zero.

9 Concluding Remarks

In this article, we have analysed LTRC data from the Lehmann family of distributions. We have considered three members of the family - Weibull, Gompertz, and Lomax, all of which are well-known lifetime models. MLEs of the parameters for these models are obtained. Model parameters are also estimated by employing the St-EM algorithm. Asymptotic confidence intervals are constructed by using the missing information principle, and parametric bootstrap approaches. Through an extensive simulation study, we have examined the performance of the inferential methods discussed here. It is observed that the estimates obtained by maximum likelihood method and the St-EM algorithm are similar with respect to bias and MSE. It is also noted that the performance of all the asymptotic confidence intervals with respect to coverage probability is satisfactory, but it is perhaps better to use a parametric bootstrap procedure for this purpose.

Through a detailed Monte Carlo simulation study, we have studied the important problem of model selection based on LTRC data. The candidate models are taken as the three distributions belonging to the Lehmann family. Two methods of model selection are studied: a likelihood-based approach, and a distance-based approach. It is observed that the discussed methods of model selection are effective in identifying the parent for a given LTRC data. It is also observed through the simulation study that the likelihood-based method performs better compared to the distance-based method for LTRC data, at least within the Lehmann family.

A scenario where a parametric birth process is imposed to the installation years may be considered following D¨rre (2020). In the context of double truncated data D¨rre (2020) used the Poisson birth process for units. The main idea behind this is that if some parametric birth structure is imposed, the efficiency of estimates might improve, which can be particularly helpful in case of small samples. Also, in some other cases, the estimation of the birth process itself may be of interest; see D¨rre (2020).

Another approach to modeling LTRC data would be to use a semi-parametric model. In this approach, along the lines of Balakrishnan and Liu (2018), one can model the baseline hazard function of the underlying lifetime variable by a piecewise exponential model, for example, to capture the local characteristics of the actual hazard. Effects of covariates on the lifetime variable can also be factored in through this model.

In the parametric setup, it will be of interest to study goodness-of-fit for models based on LTRC data. In this regard, it may be necessary to make some adjustments to the classical Kolmogorov-Smirnov, or Cramer-non Mises statistics, to account for the lack of complete information in the data. While Schey (1977) and Dufour and Maag (1978) focused on the modification of the Kolmogorov-Smirnov statistic in this regard, Chernobay, Rachev, and Fabozzi (2005) have considered some other statistics along with the Kolmorogov-Smirnov for checking goodness-of-fit for different types of truncated and censored data. More recently, Dörre et al.(2020) have considered an adjustment to the Kolmogorov-Smirnov statistic for doubly truncated data. It will be of interest to examine which of these modified statistics is suitable for LTRC data - from both theoretical and experimental perspectives. We are working on these problems, and hope to report the findings in future papers.

Acknowledgements

The authors are thankful to the anonymous reviewers for their many helpful comments and suggestions. Thanks are also due to Professor Emura for extending an invitation to contribute this article to the Special Issue.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Balakrishnan, N., & Mitra, D. (2011). Likelihood inference for log-normal data with left truncation and right censoring with an illustration. *Journal of Statistical Planning and Inference, 141*, 3536–3553.

2. Balakrishnan, N., & Mitra, D. (2012). Left truncated and right censored Weibull data and likelihood inference with an illustration. *Computational Statistics and Data Analysis, 56*, 4011-4025.
3. Balakrishnan, N., & Mitra, D. (2013). Likelihood inference based on left truncated and right censored data from a gamma distribution. *IEEE Transactions on Reliability, 62*, 679-688.
4. Balakrishnan, N., & Mitra, D. (2014a). Some further issues concerning likelihood inference for left truncated and right censored lognormal data. *Communications in Statistics - Simulation and Computation, 43*, 400-416.
5. Balakrishnan N. and Mitra D. (2014b). EM-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination (with discussions). *South African Statistical Journal, 48*, 125-204.
6. Balakrishnan, N., & Liu, K. (2018). Semi-parametric likelihood inference for Birnbaum-Saunders frailty model. *REVSTAT - Statistical Journal, 16*, 231-255.
7. Bordes, L., & Chauveau, D. (2014). Comments: EM-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination. *South African Statistical Journal, 48*, 197-200.
8. Celeux, G., Chauveau, D., & Diebolt, J. (1996). Stochastic versions of the EM algorithm: an experimental study in the mixture case. *Journal of Statistical Computation and Simulation, 55*, 287-314.
9. Celeux, G., & Diebolt, J. (1985). The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly, 2*, 73-82.
10. Chauveau, D. (1995). A stochastic EM algorithm for mixtures with censored data. *Journal of Statistical Planning and Inference, 46*, 1-25.
11. Chernobay, A., Rachev, S., and Fabozzi, F. (2005). Composite goodness-of-fit tests for left-truncated loss samples, *Technical Report, University of California Santa Barbara.*
12. Cox, D. R. (1961). Tests of separate families of hypotheses. *Proceedings of the 4th Berkeley Symposium in Mathematical Statistics and Probability, Vol. 1*, University of California Press.
13. Cox, D. R. (1962). Further results on test of separate families of hypotheses. *Journal of the Royal Statistical Society, Series B, 24*, 406-424.
14. Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1-38.
15. Dey, A. K., & Kundu, D. (2009). Discriminating among the log-normal, Weibull and generalized exponential distributions. *IEEE Transactions on Reliability, 58*, 416-424.
16. Dey, A. K., & Kundu, D. (2010). Discriminating between the log-normal and log-logistic distributions. *Communications in Statistics - Theory and Methods, 39*, 280-292.
17. Dey, A. K., & Kundu, D. (2012). Discriminating between the Weibull and log-normal distribution for Type-II censored data. *Statistics, 46*, 197-214.
18. Diebolt, J., & Celeux, G. (1993). Asymptotic properties of a stochastic EM algorithm for estimating mixing proportions. *Communications in Statistics: Stochastic Models, 9*, 599-613.
19. D¨rre, A. (2020). Bayesian estimation of a lifetime distribution under doublt truncation caused by time-restricted data collection. *Statistical Papers, 61*, 945-965.
20. Dörre, A., Huang, C. Y., Tseng, Y. K., and Emura, T. (2021). Likelihood-based analysis of doubly-truncated data under the location-scale and AFT model. *Computational Statistics, 36*, 375-408.
21. Dufour, R., and Maag, U. R. (1978). Distribution results for modified Kolmogorov-Smirnov statistics for truncated and censored samples. *Technometrics, 20*, 29-32.
22. Emura, T., Matsui, S., and Rondeau, V. (2019). Personalized Dynamic Prediction of Survival in *Survival Analysis with Correlated Endpoints*, Singapore: Springer, pp.77-93.
23. Emura, T., & Shiu, S. (2016). Estimation and model selection for left-truncated and right-censored lifetime data with application to electric power transformers analysis. *Communications in Statistics - Simulation and Computation, 45*, 3171-3189.
24. Gijbels, I., & Wang, J. L. (1993). Strong representations of the survival function estimator for truncated and censored data with applications. *Journal of Multivariate Analysis, 47*, 210-229.

25. Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London, 115*, 513-583.
26. Gotwalt, C., Xu, L., Hong, Y., & Meeker, W. Q. (2018). Applications of the Fractional-Random-Weight Bootstrap. arXiv preprint arXiv:1808.08199.
27. Hong, Y., Meeker, W. Q., McCalley, J. D. (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *The Annals of Applied Statistics, 3*, 857-879.
28. Kaplan, E. L., Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*, 457-481.
29. Klein, J. P., Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data, Second Edition*. New York: Springer-Verlag.
30. Kundu, D. (2005). Discriminating between the normal and Laplace distributions. In N. Balakrishnan, H. N. Nagaraja, & N. Kannan (Eds.), *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*, Boston: Birkhauser, pp.65-79.
31. Kundu, D., & Manglick, A. (2004). Discriminating between the Weibull and log-normal distributions. *Naval Research Logistics, 51*, 893-905.
32. Kundu, D., Mitra, D., & Ganguly, A. (2017). Analysis of left truncated right censored competing risks data. *Computational Statistics and Data Analysis, 108*, 12-26.
33. Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B, 44*, 226-233.
34. Marshall, A. W., Meza, J. C., & Olkin, I. (2001). Can data recognize its parent distribution? *Journal of Computational and Graphical Statistics, 10*, 555-580.
35. Marshall, A. W., & Olkin I. (2007). *Life Distributions; Structure of Nonparametric, Semiparametric, and Parametric Families*. New York: Springer.
36. McLachlan, G., & Krishnan, T. (2008). *The EM Algorithm and Extensions, Second Edition*. Hoboken, New Jersey: John Wiley & Sons.
37. Ng, H. K. T., & Ye, Z. (2014). Comments: EM-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination. *South African Statistical Journal, 48*, 177-180.
38. Nielsen, S. F. (2000). The stochastic EM algorithm: estimation and asymptotic results. *Bernoulli, 6*, 457-489.
39. Pereira, B. B., & Pereira, C. A. B. (2016). *Model Choice in Nonnested Families*. Berlin: Springer-Verlag.
40. Schey, H. M. (1977). The asymptotic distribution of the one-sided Kolmogorov-Smirnov statistic for truncated data. *Communications in Statistics - Theory and Methods, 6*, 1361-1366.
41. Shinohara, S., Lin, Y. H., Michimae, H., and Emura, T. (2020). Dynamic lifetime prediction using a Weibull-based bivariate failure time model. *Communications in Statistics - Simulation and Computation*, DOI:10.1080/03610918.2020.1855449
42. Taylor, J. A., & Jakeman, A. J. (1985). Identification of a distributional model. *Communications in Statistics - Simulation and Computation, 14*, 497-508.
43. Tsai, W. Y., Jewel, N. P., & Wang, M. C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika, 74*, 883-886.
44. Tse, S. (2003). Strong Gaussian approximations in the left truncated and right censored model. *Statistica Sinica, 13*, 275-282.
45. Wang, B. X., Yu, K., & Jones, M. C. (2010). Inference under progressively type-II right-censored sampling for certain lifetime distributions. *Technometrics, 52*, 453-460.
46. Ye, Z., & Ng, H. K. T. (2014). On analysis of incomplete field failure data. *The Annals of Applied Statistics, 8*, 1713-1727.