Computational Statistics and Data Analysis 56 (2012) 4011–4025

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Left truncated and right censored Weibull data and likelihood inference with an illustration

N. Balakrishnan, Debanjan Mitra *

Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1

ARTICLE INFO

Article history:
Received 10 June 2011
Received in revised form 4 May 2012
Accepted 5 May 2012
Available online 14 May 2012

Keywords:
Maximum likelihood estimates
EM algorithm
Lifetime data
Left truncation
Right censoring
Weibull distribution
Missing information principle
Asymptotic variance–covariance matrix
EM gradient algorithm
Newton–Raphson method
Asymptotic confidence intervals

ABSTRACT

The Weibull distribution is a very popular distribution for modeling lifetime data. Left truncation and right censoring are often observed in lifetime data. Here, the EM algorithm is applied to estimate the model parameters of the Weibull distribution fitted to data containing left truncation and right censoring. The maximization part of the EM algorithm is carried out using the EM gradient algorithm (Lange, 1995). The Weibull distribution is also fitted using the Newton–Raphson (NR) method. The two methods of estimation are then compared through an extensive Monte Carlo simulation study. The asymptotic variance–covariance matrix of the MLEs under the EM framework is obtained through the missing information principle (Louis, 1982), and asymptotic confidence intervals for the parameters are then constructed. The asymptotic confidence intervals corresponding to the missing information principle and the observed information matrix are compared in terms of coverage probabilities, through a simulation study. Finally, all the methods of inference discussed here are illustrated through some numerical examples.

1. Introduction

Truncation and censoring occur quite commonly while observing lifetime data; the books by Cohen (1991), Balakrishnan and Cohen (1991), and Meeker and Escobar (1998) provide detailed accounts in this regard. In this paper, we are concerned with lifetime data that are left truncated and right censored. Recently, Hong et al. (2009) carried out an analysis of lifetime data of power transformers from an energy company in the US, and these data were naturally left truncated and right censored. They used the Weibull distribution as their lifetime model and fitted it by a direct maximization approach. Here, we describe in detail the steps of the Expectation Maximization (EM) algorithm for fitting the Weibull distribution to left truncated and right censored data; see McLachlan and Krishnan (2008) for a comprehensive discussion on this topic. For comparative purposes, we also fit the Weibull distribution by the Newton–Raphson (NR) method; the two methods give extremely close results under this setup.

The rest of this paper is organized as follows. In Section 2, we present a description of the form of the data and the corresponding likelihood function. In Section 3, the EM algorithm for the maximum likelihood estimation of the model parameters is described. As the expected complete log-likelihood is a complicated non-linear function of the parameters involved, the maximization part of the EM algorithm is done using the EM gradient algorithm (Lange, 1995). In this section, we also derive the asymptotic variance–covariance matrix of the maximum likelihood estimates (MLEs) within the EM framework by using the missing information principle of Louis (1982). Then, we use it to construct corresponding asymptotic

* Corresponding author.
E-mail address: mitrad2@mcmaster.ca (D. Mitra).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.05.004

confidence intervals for the model parameters. The asymptotic confidence intervals, making use of the observed information matrix, are also constructed, and these two confidence intervals are compared in terms of coverage probabilities. An application to prediction purposes is discussed in Section 4. In Section 5, we present the results of an extensive Monte Carlo simulation study in which we evaluate and compare different methods of point estimation as well as interval estimation. Next, in Section 6, we illustrate through some numerical examples all the methods of inference developed here. Finally, we conclude the paper with some remarks in Section 7.

2. Form of data and the likelihood

In this paper, the efficacy of all the methods of inference developed are assessed through Monte Carlo simulations and then illustrated through some numerical examples. For incorporating left truncation and right censoring in the data, in the simulation study, we mimic the dataset used by Hong et al. (2009). That is, the data can be considered as lifetime data of power transformers in the electrical industry, wherein the lifetime is observed only if the unit failed after 1980, as detailed record keeping on the lifetime of machines started in that year, thus resulting in the data being left truncated. Moreover, the lifetimes of the machines are observed until 2008, making that year the right censoring point. We adopt the same setup here and discuss the fitting of a two-parameter Weibull distribution to the left truncated and right censored data. We perform a logarithmic transformation of the lifetime data, as it yields the model of an extreme value distribution (involving location and scale parameters rather than scale and shape parameters as in the case of Weibull) which is more convenient for carrying out the subsequent derivations and calculations.

Let X be the original lifetime variable, which follows a Weibull distribution with scale parameter α and shape parameter η. The density of X is given by [see Johnson et al. (1994)]

$$f_X(x) = \left(\frac{\eta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\eta-1}\exp\left\{-\left(\frac{x}{\alpha}\right)^{\eta}\right\}, \quad x > 0, \alpha > 0, \eta > 0.$$

Then, the log-transformed variable $Y = \log X$ follows an extreme value distribution with density

$$f_Y(y) = \frac{1}{\sigma}\exp\left[\left(\frac{y-\mu}{\sigma}\right) - \exp\left(\frac{y-\mu}{\sigma}\right)\right], \quad -\infty < y < \infty, -\infty < \mu < \infty, \sigma > 0,$$

where $\mu = \log\alpha$ and $\sigma = 1/\eta$ are the location and scale parameters, respectively. Let C denote the log-transformed censoring time variable, δ_i denote the censoring indicator, i.e., δ_i is 0 if the i-th observation is censored and 1 if it is not censored, and τ_i^L denote the log-transformed left-truncation time. To explain this more clearly, for a machine installed before 1980, τ_i^L is the time between the year of installation and the truncation point of 1980. Let ν_i denote the truncation indicator, i.e., ν_i is 0 if the i-th observation is truncated and 1 if it is not truncated. Further, let S_1 and S_2 be two index sets that correspond to the set of machines installed after 1980 and on or before 1980, respectively.

The likelihood function for the left truncated and right censored data is given by

$$\begin{aligned}
L(\mu,\sigma) = &\prod_{i\in S_1}\left\{\frac{1}{\sigma}\exp\left[\left(\frac{y_i-\mu}{\sigma}\right) - \exp\left(\frac{y_i-\mu}{\sigma}\right)\right]\right\}^{\delta_i}\left\{\exp\left[-\exp\left(\frac{y_i-\mu}{\sigma}\right)\right]\right\}^{1-\delta_i} \\
&\times \prod_{i\in S_2}\left\{\frac{\exp\left[\exp\left\{\frac{\tau_i^L-\mu}{\sigma}\right\}\right]}{\sigma}\exp\left[\left(\frac{y_i-\mu}{\sigma}\right) - \exp\left(\frac{y_i-\mu}{\sigma}\right)\right]\right\}^{\delta_i} \\
&\times \left\{\exp\left[\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right)\right]\exp\left[-\exp\left(\frac{y_i-\mu}{\sigma}\right)\right]\right\}^{1-\delta_i}.
\end{aligned}$$

The log-likelihood function, after some simplification and the use of the truncation indicator ν_i, becomes

$$\log L = \sum_{i=1}^{n}\left[-\delta_i\log\sigma + \delta_i\left(\frac{y_i-\mu}{\sigma}\right) - \exp\left(\frac{y_i-\mu}{\sigma}\right)\right] + \sum_{i=1}^{n}(1-\nu_i)\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right).$$

3. Methods of estimation

3.1. The EM algorithm

The EM algorithm is a very powerful and useful tool for analyzing incomplete data; see McLachlan and Krishnan (2008) for an elaborate discussion on the method. The algorithm consists of two steps — the Expectation step (E-step) and the Maximization step (M-step). In the E-step, the conditional expectation of the complete data likelihood is obtained, given the observed incomplete data and the current value of the parameter, real or vector valued. This expected likelihood is essentially a function of the parameter involved, and the current value of the parameter under which the expectation has been calculated. In the M-step, this expected complete data likelihood is then maximized with respect to the parameter. The E- and M-steps are then iterated till convergence. This algorithm is known to have some desirable and advantageous properties over the direct methods for obtaining the MLEs in the case of incomplete data; see the above-mentioned reference for details.

First, we construct the complete data log-likelihood function. Under the extreme value model and with the parameter vector denoted by $\boldsymbol{\theta} = (\mu, \sigma)'$, had there been no censoring, the complete data likelihood would be

$$\begin{aligned} L_c(\mathbf{t}; \boldsymbol{\theta}) = & \prod_{S_1} \left\{ \frac{1}{\sigma} \exp\left[\left(\frac{t_i - \mu}{\sigma} \right) - \exp\left(\frac{t_i - \mu}{\sigma} \right) \right] \right\} \\ & \times \prod_{S_2} \left\{ \frac{\exp\left\{ \exp\left(\frac{\tau_i^L - \mu}{\sigma} \right) \right\}}{\sigma} \exp\left[\left(\frac{t_i - \mu}{\sigma} \right) - \exp\left(\frac{t_i - \mu}{\sigma} \right) \right] \right\}. \end{aligned}$$

Correspondingly, the complete data log-likelihood function, using the truncation indicator ν_i, is given by

$$\log L_c(\mathbf{t}; \boldsymbol{\theta}) = -n \log \sigma + \sum_{i=1}^{n} \left[\left(\frac{t_i - \mu}{\sigma} \right) - \exp\left(\frac{t_i - \mu}{\sigma} \right) \right] + \sum_{i=1}^{n} (1 - \nu_i) \exp\left(\frac{\tau_i^L - \mu}{\sigma} \right). \tag{1}$$

Let $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)'$ be the vector of censoring indicators, and $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ be the observed data vector, where $y_i = \min(t_i, c_i)$, i.e., the minimum of the two variables as the data are right censored.

The E-Step: In the E-step, we calculate the conditional expectation of the complete data log-likelihood, i.e., we calculate

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = E_{\boldsymbol{\theta}^{(k)}}[\log L_c(\mathbf{t}; \boldsymbol{\theta}) | \mathbf{y}, \boldsymbol{\delta}]. \tag{2}$$

Clearly, the expectations of interest are $E_{\boldsymbol{\theta}^{(k)}}[\frac{T_i - \mu}{\sigma} | T_i > y_i]$ and $E_{\boldsymbol{\theta}^{(k)}}[\exp(\frac{T_i - \mu}{\sigma}) | T_i > y_i]$. To derive these conditional expectations, we first consider the conditional density of T_i, given $T_i > y_i$, given by

$$f_{T_i | Y_i = y_i}(t_i) = \frac{\exp\left[\exp\left(\frac{y_i - \mu}{\sigma}\right)\right]}{\sigma} \exp\left[\left(\frac{t_i - \mu}{\sigma} \right) - \exp\left(\frac{t_i - \mu}{\sigma} \right) \right], \quad t_i > y_i.$$

Based on the above conditional density function, the conditional mgf (moment generating function) of $(\frac{T_i - \mu}{\sigma})$, given $T_i > y_i$, can be easily derived to be

$$M_{\left(\frac{T_i - \mu}{\sigma}\right)}(\theta) = \exp[\exp(\xi_i)] \Gamma(\theta + 1, e^{\xi_i}), \tag{3}$$

where $\xi_i = (\frac{y_i - \mu}{\sigma})$ and $\Gamma(p, x) = \int_x^\infty u^{p-1} e^{-u} du$ is the upper incomplete gamma function.

Using the mgf in (3), the required expectations can be derived to be

$$E_{\boldsymbol{\theta}^{(k)}} \left[\frac{T_i - \mu}{\sigma} \middle| T_i > y_i \right] = \frac{\mu^{(k)} + \sigma^{(k)} E_{1i}^{(k)} - \mu}{\sigma}, \tag{4}$$

$$E_{\boldsymbol{\theta}^{(k)}} \left[\exp\left(\frac{T_i - \mu}{\sigma} \right) \middle| T_i > y_i \right] = e^{\left(\frac{\mu^{(k)} - \mu}{\sigma} \right)} \exp[\exp(\xi_i^{(k)})] \Gamma\left(\frac{\sigma^{(k)}}{\sigma} + 1, e^{\xi_i^{(k)}} \right), \tag{5}$$

where

$$E_{1i}^{(k)} = e^{e^{\xi_i^{(k)}}} \Psi(1) - \sum_{p=0}^{\infty} \{\xi_i^{(k)} + \Psi(1)\} \frac{e^{(p+1)\xi_i^{(k)}}}{\Gamma(p+2)} + \sum_{p=0}^{\infty} \frac{e^{(p+1)\xi_i^{(k)}} \Psi(p+2)}{\Gamma(p+2)}.$$

Substituting (4) and (5) into Eq. (2), we obtain

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = & -n \log \sigma + \sum_{i=1}^{n} (1 - \nu_i) \exp\left(\frac{\tau_i^L - \mu}{\sigma} \right) + \sum_{\delta_i = 1} \left\{ \left(\frac{t_i - \mu}{\sigma} \right) - \exp\left(\frac{t_i - \mu}{\sigma} \right) \right\} \\ & + \sum_{\delta_i = 0} \left\{ \frac{\mu^{(k)} + \sigma^{(k)} E_{1i}^{(k)} - \mu}{\sigma} \right\} - \sum_{\delta_i = 0} \left\{ e^{\left(\frac{\mu^{(k)} - \mu}{\sigma} \right)} M_{\left(\frac{T_i - \mu}{\sigma} \right)} \left(\frac{\sigma^{(k)}}{\sigma} \right) \right\}. \end{aligned} \tag{6}$$

The quantity $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ given in (6) needs to be maximized with respect to $\boldsymbol{\theta}$. Evidently, the maximization poses a challenge as the function involved is a complicated non-linear function of μ and σ, and the process used for this purpose is described next.

The M-Step: In the maximization step, the quantity $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ in (6) is maximized with respect to $\boldsymbol{\theta}$ over the parameter space $\boldsymbol{\Theta}$ to obtain the improved estimate of the paramater as

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}).$$

The E-step and the M-step are then continued iteratively until convergence (to a specified tolerance level) to obtain the MLE of the parameter $\boldsymbol{\theta}$. From the complicated form of the function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, it is obvious that there are no explicit MLEs for the parameters, and one has to depend on a numerical maximization procedure. Here, we make use of the EM gradient algorithm (Lange, 1995). In this algorithm, the function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ is maximized by a one-step Newton–Raphson method, to get the updated estimate $\boldsymbol{\theta}^{(k+1)}$. This algorithm is a special case of the generalized EM algorithm (Dempster et al., 1977), and is closely related to the original EM algorithm. The properties of this algorithm are quite close to that of the EM algorithm. However, unlike the EM algorithm, this algorithm depends on the choice of the initial values.

In the M-step, the one-step Newton–Raphson method is carried out with the following expressions:

$$\begin{aligned}\frac{\partial Q}{\partial \mu} &= -\sum_{i=1}^{n} \frac{1-\nu_i}{\sigma} \exp\left(\frac{\tau_i^L-\mu}{\sigma}\right) - \sum_{\delta_i=1} \frac{\left\{1-\exp\left(\frac{t_i-\mu}{\sigma}\right)\right\}}{\sigma} - \sum_{\delta_i=0}\left\{\frac{1}{\sigma}\right\} \\ &\quad + \frac{1}{\sigma}\sum_{\delta_i=0}\left\{\exp\left(\frac{\mu^{(k)}-\mu}{\sigma}\right)\exp(e^{\xi_i^{(k)}})\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}} e^{-w}dw\right\}, \end{aligned} \tag{7}$$

$$\begin{aligned}\frac{\partial Q}{\partial \sigma} &= -\frac{n}{\sigma} - \sum_{i=1}^{n}(1-\nu_i)\left(\frac{\tau_i^L-\mu}{\sigma^2}\right)\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right) - \sum_{\delta_i=1}\left(\frac{t_i-\mu}{\sigma^2}\right)\left\{1-\exp\left(\frac{t_i-\mu}{\sigma}\right)\right\} \\ &\quad - \sum_{\delta_i=0}\left\{\frac{\mu^{(k)}+\sigma^{(k)}E_{1i}^{(k)}-\mu}{\sigma^2}\right\} + \sum_{\delta_i=0}\exp(e^{\xi_i^{(k)}})\exp\left(\frac{\mu^{(k)}-\mu}{\sigma}\right) \\ &\quad \times\left\{\left(\frac{\mu^{(k)}-\mu}{\sigma^2}\right)\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}} e^{-w}dw + \frac{\sigma^{(k)}}{\sigma^2}\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}} e^{-w}\log w dw\right\}, \end{aligned} \tag{8}$$

$$\begin{aligned}-\frac{\partial^2 Q}{\partial \mu^2} &= -\sum_{i=1}^{n}\frac{1-\nu_i}{\sigma^2}\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right) + \sum_{\delta_i=1}\frac{\exp\left(\frac{t_i-\mu}{\sigma}\right)}{\sigma^2} \\ &\quad + \frac{1}{\sigma^2}\sum_{\delta_i=0}\left\{\exp\left(\frac{\mu^{(k)}-\mu}{\sigma}\right)\exp(e^{\xi_i^{(k)}})\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}} e^{-w}dw\right\}, \end{aligned} \tag{9}$$

$$\begin{aligned}-\frac{\partial^2 Q}{\partial \mu \partial \sigma} &= -\sum_{i=1}^{n}(1-\nu_i)\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right)\left\{\frac{1}{\sigma^2}+\left(\frac{\tau_i^L-\mu}{\sigma^3}\right)\right\} \\ &\quad - \sum_{\delta_i=1}\frac{1}{\sigma^2}\left\{1-\exp\left(\frac{t_i-\mu}{\sigma}\right)-\left(\frac{t_i-\mu}{\sigma}\right)\exp\left(\frac{t_i-\mu}{\sigma}\right)\right\} - \sum_{\delta_i=0}\left\{\frac{1}{\sigma^2}\right\} \\ &\quad + \sum_{\delta_i=0}\exp(e^{\xi_i^{(k)}})\exp\left(\frac{\mu^{(k)}-\mu}{\sigma}\right)\left\{\left(\frac{\mu^{(k)}-\mu}{\sigma^3}+\frac{1}{\sigma^2}\right)\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}} e^{-w}dw \right. \\ &\quad \left. + \frac{\sigma^{(k)}}{\sigma^3}\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}} e^{-w}\log w dw\right\}, \end{aligned} \tag{10}$$

$$\begin{aligned}-\frac{\partial^2 Q}{\partial \sigma^2} &= -\frac{n}{\sigma^2} - \sum_{i=1}^{n}(1-\nu_i)\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right)\left\{\frac{2(\tau_i^L-\mu)}{\sigma^3}+\left(\frac{\tau_i^L-\mu}{\sigma^2}\right)^2\right\} \\ &\quad - \sum_{\delta_i=1}\left(\frac{t_i-\mu}{\sigma^2}\right)\left\{\frac{2}{\sigma}-\frac{2}{\sigma}\exp\left(\frac{t_i-\mu}{\sigma}\right)-\left(\frac{t_i-\mu}{\sigma^2}\right)\exp\left(\frac{t_i-\mu}{\sigma}\right)\right\} \\ &\quad - \sum_{\delta_i=0}\left\{\frac{2(\mu^{(k)}+\sigma^{(k)}E_{1i}^{(k)}-\mu)}{\sigma^3}\right\} + \sum_{\delta_i=0}\exp(e^{\xi_i^{(k)}})\exp\left(\frac{\mu^{(k)}-\mu}{\sigma}\right) \\ &\quad \times\left\{\left\{\left(\frac{\mu^{(k)}-\mu}{\sigma^2}\right)^2+\frac{2(\mu^{(k)}-\mu)}{\sigma^3}\right\}\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}} e^{-w}dw \right. \\ &\quad + \left\{\frac{2(\mu^{(k)}-\mu)\sigma^{(k)}}{\sigma^4}+\frac{2\sigma^{(k)}}{\sigma^3}\right\}\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}} e^{-w}\log w dw \\ &\quad \left. + \left(\frac{\sigma^{(k)}}{\sigma^2}\right)^2\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}} e^{-w}(\log w)^2 dw\right\}. \end{aligned} \tag{11}$$

It can be seen readily that the above expressions contain some integrals that require special techniques for their evaluation. For further simplification of the above expressions and relevant derivations, see the Appendix. It has been observed in our extensive empirical study that the numerical MLEs converge to the true parameter values quite accurately.

3.2. Newton–Raphson method

The NR method is a direct approach for obtaining the MLEs by maximizing the likelihood function. It involves calculation of the first and second derivatives of the observed log-likelihood with respect to the parameters. Herein, we use the NR method for comparative purpose. The NR method, although works well in general, fails to converge in some cases under this setup. In our study, we employed the NR method by a default function of the R software, called the "maxNR" function. We observed in our empirical study that the EM and the NR methods yield close results in most cases.

3.3. Asymptotic variances and covariance of the MLEs

Unlike the NR method, in the EM algorithm, the asymptotic variances and covariance of the MLEs are not directly obtained as a byproduct of the algorithm. Within the EM framework, the missing information principle (Louis, 1982) can be applied to obtain the observed information matrix as

$$\text{Observed information} = \text{Complete information} - \text{Missing information}.$$

Then, inverting the observed information matrix, one can obtain the asymptotic variance–covariance matrix of the MLEs.

Let $I_T(\boldsymbol{\theta})$, $I_Y(\boldsymbol{\theta})$ and $I_{C|Y}(\boldsymbol{\theta})$ denote the complete information matrix, observed information matrix and the missing information matrix, respectively. The complete information matrix is given by

$$I_T(\boldsymbol{\theta}) = -E\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\log L_c(\mathbf{t};\boldsymbol{\theta})\right]. \tag{12}$$

The Fisher information matrix in the i-th observation that is censored is given by

$$I^{(i)}_{C|Y}(\boldsymbol{\theta}) = -E\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\log f_{C_i}(c_i|C_i > y_i, \boldsymbol{\theta})\right]. \tag{13}$$

Hence, the expected missing information can then be easily obtained as

$$I_{C|Y}(\boldsymbol{\theta}) = \sum_{i:\delta_i=0} I^{(i)}_{C|Y}(\boldsymbol{\theta}). \tag{14}$$

Thus, by the missing information principle, the observed information matrix can be obtained as

$$I_Y(\boldsymbol{\theta}) = I_T(\boldsymbol{\theta}) - I_{C|Y}(\boldsymbol{\theta}). \tag{15}$$

The asymptotic variance–covariance matrix of the MLE of $\boldsymbol{\theta}$ can be obtained finally by inverting the observed information matrix $I_Y(\boldsymbol{\theta})$ in (15) and evaluating at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

The required conditional density can be derived as [see Ng et al. (2002)]

$$f_{C_i|Y_i}(c_i|C_i > y_i, \mu, \sigma) = \frac{\exp(e^{\xi_i})}{\sigma}\exp\left[\left(\frac{c_i-\mu}{\sigma}\right) - \exp\left(\frac{c_i-\mu}{\sigma}\right)\right], \quad c_i > y_i.$$

The elements of the complete information matrix $I_T(\boldsymbol{\theta})$ are given by

$$-E\left[\frac{\partial^2}{\partial \mu^2}\log L_c(\mathbf{t};\boldsymbol{\theta})\right] = \frac{n}{\sigma^2},$$

$$\begin{aligned}
-E\left[\frac{\partial^2}{\partial \sigma^2}\log L_c(\mathbf{t};\boldsymbol{\theta})\right] &= -\frac{n}{\sigma^2} - \sum_{i=1}^{n}\frac{\nu_i}{\sigma^2}[2\{\Psi(1)-\Psi(2)\} - \{\Psi^2(2)+\Psi'(2)\}] \\
&\quad - \sum_{i=1}^{n}\frac{(1-\nu_i)}{\sigma^2}\left[2\left\{\lambda_i + \Psi(1) - \lambda_i\exp(e_i^{\lambda}) + \sum_{p=0}^{\infty}\frac{e^{(p+1)\lambda_i}\Psi(p+2)}{\Gamma(p+2)}\right\}\right. \\
&\quad - \left\{\Psi(2)(1+e^{\lambda_i}) + \sum_{p=0}^{\infty}\frac{e^{(p+2)\lambda_i}\Psi(p+3)}{\Gamma(p+3)} - \lambda_i(\exp(e^{\lambda_i}) - 1 - e^{\lambda_i})\right\} \\
&\quad - \left\{(\Psi'(2)+\Psi^2(2))(1+e^{\lambda_i}) - (\lambda_i^2 + 2\lambda_i\Psi(2))(\exp(e^{\lambda_i}) - 1 - e^{\lambda_i})\right. \\
&\quad \left.\left. +\ 2(\lambda_i + \Psi(2))\sum_{p=0}^{\infty}\frac{e^{(p+2)\lambda_i}\Psi(p+3)}{\Gamma(p+3)} + \sum_{p=0}^{\infty}\frac{e^{(p+2)\lambda_i}\{\Psi'(p+3) - \Psi^2(p+3)\}}{\Gamma(p+3)}\right\}\right] \\
&\quad - \sum_{i=1}^{n}\frac{(1-\gamma_i)}{\sigma^2}\left[2\lambda_i e^{\lambda_i} + \lambda_i^2 e^{\lambda_i}\right],
\end{aligned}$$

$$-E\left[\frac{\partial^2}{\partial\sigma\partial\mu}\log L_c(\mathbf{t};\boldsymbol{\theta})\right]=\frac{n\Psi(2)}{\sigma^2}+\sum_{i=1}^{n}\frac{(1-\nu_i)}{\sigma^2}\left[\Psi(2)e^{\lambda_i}+\sum_{p=0}^{\infty}\frac{e^{(p+2)\lambda_i}\Psi(p+3)}{\Gamma(p+3)}-\lambda_i(\exp(e^{\lambda_i})-1)\right]$$

where $\lambda_i=\frac{\tau_i^L-\mu}{\sigma}$.

Next, to obtain the elements of the missing information matrix, let us consider the logarithm of the truncated extreme value density given by

$$\log f_{C_i|Y_i}(c_i|C_i>y_i,\mu,\sigma)=-\log\sigma+\exp(e^{\xi_i})+\left(\frac{c_i-\mu}{\sigma}\right)-\exp\left(\frac{c_i-\mu}{\sigma}\right).$$

From this, we obtain the following expressions:

$$\frac{\partial^2}{\partial\mu^2}\log f_{C_i|Y_i}=\frac{1}{\sigma^2}[e^{\xi_i}-e^{\beta_i}], \tag{16}$$

$$\frac{\partial^2}{\partial\sigma^2}\log f_{C_i|Y_i}=\frac{1}{\sigma^2}[1+2\xi_ie^{\xi_i}+\xi_i^2e^{\xi_i}+2\beta_i-2\beta_ie^{\beta_i}-\beta_i^2e^{\beta_i}], \tag{17}$$

$$\frac{\partial^2}{\partial\mu\partial\sigma}\log f_{C_i|Y_i}=\frac{1}{\sigma^2}[1+e^{\xi_i}+\xi_ie^{\xi_i}-e^{\beta_i}-\beta_ie^{\beta_i}], \tag{18}$$

where $\beta_i=\frac{C_i-\mu}{\sigma}$. To obtain the expected values of these second derivatives, as before, we find the conditional moment generating function of $\beta_i=\frac{C_i-\mu}{\sigma}$, given $C_i>y_i$. The conditional mgf is given by

$$\begin{aligned}M_{\left(\frac{C_i-\mu}{\sigma}\right)}(\theta)&=\exp[\exp(\xi_i)]\Gamma(\theta+1,e^{\xi_i})\\&=\Gamma(\theta+1)\left[e^{e^{\xi_i}}-\sum_{p=0}^{\infty}\frac{e^{(\theta+p+1)\xi_i}}{\Gamma(\theta+p+2)}\right].\end{aligned}$$

Using this expression, the following expectations can be obtained:

$$E[\beta_i|C_i>y_i]=\xi_i+\Psi(1)-\xi_ie^{e^{\xi_i}}+\sum_{p=0}^{\infty}\frac{e^{(p+1)\xi_i}\Psi(p+2)}{\Gamma(p+2)},$$

$$E[\exp(\beta_i)|C_i>y_i]=1+e^{\xi_i},$$

$$E[\beta_i\exp(\beta_i)|C_i>y_i]=\Psi(2)(1+e^{\xi_i})+\sum_{p=0}^{\infty}\frac{e^{(p+2)\xi_i}\Psi(p+3)}{\Gamma(p+3)}-\xi_i(e^{e^{\xi_i}}-1-e^{\xi_i}),$$

$$\begin{aligned}E[\beta_i^2\exp(\beta_i)|C_i>y_i]&=[\Psi^2(2)+\Psi'(2)](1+e^{\xi_i})-[\xi_i^2+2\xi_i\Psi(2)](e^{e^{\xi_i}}-1-e^{\xi_i})\\&\quad+2[\xi_i+\Psi(2)]\sum_{p=0}^{\infty}\frac{e^{(p+2)\xi_i}\Psi(p+3)}{\Gamma(p+3)}+\sum_{p=0}^{\infty}\frac{e^{(p+2)\xi_i}[\Psi'(p+3)-\Psi^2(p+3)]}{\Gamma(p+3)}.\end{aligned}$$

Using Eqs. (16)–(18) and these expectations, the expected missing information matrix $I_{C|Y}(\boldsymbol{\theta})$ can be obtained as in (14), and then the observed Fisher information matrix can be obtained from Eq. (15). Finally, by inverting $I_Y(\hat{\boldsymbol{\theta}})$, the asymptotic variance–covariance matrix of the MLEs can be obtained.

3.4. Confidence intervals

After obtaining the MLEs and their asymptotic variances, the asymptotic confidence intervals for μ and σ can be constructed using the asymptotic normality of the MLEs. Here, we study the asymptotic confidence intervals for μ and σ corresponding to both the EM algorithm and the NR method. Evidently, the confidence intervals corresponding to the EM algorithm and the NR method will be different, due to the different estimates of the parameters and the standard errors obtained from these methods. Then, these confidence intervals are compared in terms of their coverage probabilities through a Monte Carlo simulation study.

One can also construct parametric bootstrap confidence intervals for μ and σ in the following way. First of all, based on a given data of size n, the MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}=(\mu,\sigma)'$ is obtained. Then, using $\hat{\boldsymbol{\theta}}$ as the true value of the parameter, a sample of size n in the same sampling framework with left truncation and right censoring is produced. This process is repeated for 1000 Monte Carlo simulation runs, and the MLEs are obtained for each of these samples. Then, based on these 1000 estimates, the bootstrap bias and variance for the estimates of μ and σ are obtained. In the final step, a $100(1-\alpha)\%$ parametric bootstrap confidence interval for μ is obtained as

$$\text{LCL}:\quad\hat{\mu}-b_\mu-z_{\alpha/2}\sqrt{v_\mu},\qquad\text{UCL}:\quad\hat{\mu}-b_\mu+z_{\alpha/2}\sqrt{v_\mu},$$

where b_μ and v_μ are the bootstrap bias and variance for the estimate of μ, respectively, and z_α is the upper α-percentage point of the standard normal distribution. The confidence interval for σ can be constructed in a similar manner. The parametric bootstrap confidence intervals can then be compared to the other asymptotic confidence intervals in terms of coverage probabilities.

Another alternative could be to construct nonparametric bootstrap confidence intervals for the parameters. However, in the presence of both truncation and censoring, the nonparametric bootstrap is expected to show poor performance compared to the parametric bootstrap; for example, Balakrishnan et al. (2007) observed that parametric bootstrap confidence intervals performed better than their nonparametric counterpart for censored data in the context of step-stress experiments. However, we do not study the performance of the bootstrap confidence intervals here because of some practical constraints that are described in Section 5.

4. An application to prediction

With the estimated parameters μ and σ, one can obtain the probability of a censored unit working to a future year, given that it has worked until Y_{cen} (the right censoring point). Suppose a unit is installed in the year Y_{ins}, before 1980, i.e., the unit is left truncated. Then, the left truncation time for the unit is $\tau^L = 1980 - Y_{ins}$. Then, the probability that this unit will be working till a future year Y_{fut}, given that it is right censored at Y_{cen}, will be given by

$$\pi = P(T > \log(Y_{fut} - Y_{ins}) | T > \log(Y_{cen} - Y_{ins})) = \frac{S^*(\log(Y_{fut} - Y_{ins}))}{S^*(\log(Y_{cen} - Y_{ins}))},$$

where T is the log-transformed lifetime of the unit, and $S^*(\cdot)$ is the survival function of the left truncated log-transformed random variable. Clearly, the above probability reduces to

$$\pi = \frac{S(\log(Y_{fut} - Y_{ins}))}{S(\log(Y_{cen} - Y_{ins}))} = g(\boldsymbol{\theta}),$$

where $S(\cdot)$ is the survival function of the untruncated log-transformed lifetime variable, and $g(\cdot)$ is a function of $\boldsymbol{\theta}$. Incidentally, this is also the probability of the same event for a unit which is not left truncated. One can obtain an estimate $\hat{\pi}$ by using the MLE $\hat{\boldsymbol{\theta}}$ as

$$\hat{\pi} = \exp\left\{\exp\left(\frac{b - \hat{\mu}}{\hat{\sigma}}\right) - \exp\left(\frac{a - \hat{\mu}}{\hat{\sigma}}\right)\right\} = g(\hat{\boldsymbol{\theta}}), \tag{19}$$

where $a = \log(Y_{fut} - Y_{ins})$ and $b = \log(Y_{cen} - Y_{ins})$.

Using the delta-method, and the asymptotic variance–covariance matrix of the MLE $\hat{\boldsymbol{\theta}}$, one can also estimate the variance of the above estimate $\hat{\pi}$. A straightforward application of the delta-method yields

$$\hat{\pi} \sim \mathrm{N}(\pi, \mathrm{Var}(\hat{\pi})),$$

where $\mathrm{Var}(\hat{\pi})$ can be estimated as

$$\widehat{\mathrm{Var}(\hat{\pi})} = \left(\left(\frac{\partial g}{\partial \mu}\right)^2 \mathrm{Var}(\hat{\mu}) + 2\left(\frac{\partial g}{\partial \mu}\right)\left(\frac{\partial g}{\partial \sigma}\right)\mathrm{Cov}(\hat{\mu}, \hat{\sigma}) + \left(\frac{\partial g}{\partial \sigma}\right)^2 \mathrm{Var}(\hat{\sigma})\right)\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{20}$$

Examples of this technique are given in Section 6.

5. Simulation results

The simulations and all required computations to examine all the inferential procedures developed here are performed using the R software. The steps followed for the simulation are as follows. To incorporate a certain percentage of truncation into the data, the truncation percentage is fixed. Then with this fixed percentage of truncation, the installation years are sampled through with-replacement unequal probability sampling from an arbitrary set of years. From a Weibull distribution with specified values of the scale parameter α and shape parameter η, the lifetimes of the machines, in years, are sampled. Addition of these lifetimes to the corresponding installation years gives the years of failure of the machines. Clearly, from the years of failure of the machines, it is decided whether the lifetime of a machine is censored or not. As mentioned earlier, the year of truncation has been fixed as 1980 and the year of censoring has been fixed as 2008 in our study, just as in the work of Hong et al. (2009). As the data are right censored, the lifetime of a censored unit is taken as the minimum of the lifetime and the censoring time. Because the data are left truncated, no information on the lifetime of a machine is available if the year of failure for it is before 1980. Therefore, if the year of failure for a machine is obtained to be a year before 1980, that observation is discarded, and a new installation year and lifetime are simulated for that particular unit. Then, the data are log-transformed, and all the analyses are carried out in the logarithmic scale, i.e., based on the extreme value distribution.

The sample sizes used in this study are 100, 200 and 300. The truncation percentages are fixed at 30 and 40. The two different truncation percentages would demonstrate how the model behaves under small and heavy truncation. The set

Table 1
The number of times the EM algorithm failed to converge, in 1000 Monte Carlo runs.

(μ, σ)	Trunc. (%)	$n = 100$	$n = 200$	$n = 300$
(3.55, 0.33)	30	0	1	0
	40	1	0	0
(3.69, 0.25)	30	8	0	0
	40	10	1	0

of installation years was split into two parts: (1960–1979) and (1980–1995). Unequal probabilities were assigned to the different years as follows: for the period 1980–1995, a probability of 0.1 was attached to each of the first six years, and a probability of 0.04 was attached to each of the remaining years of this period; for the period 1960–1979, a probability of 0.15 was attached to each of the first five years, and the remaining probability was distributed equally over the remaining years of this period. This setup produced, along with the desired level of truncation, sufficiently many censored observations. Two choices of the Weibull parameter vector (α, η) are made: (35, 3) and (40, 4). Thus, the corresponding values for the extreme value parameter vector $\boldsymbol{\theta} = (\mu, \sigma)$ are (3.55, 0.33) and (3.69, 0.25), respectively. All the simulation results are based on 1000 Monte Carlo runs.

We know that when T follows an extreme value distribution with parameters μ and σ, then

$$E(T) = \mu - \gamma\sigma, \qquad \mathrm{Var}(T) = \frac{\pi^2}{6}\sigma^2,$$

where $\gamma = 0.5772$ (approximately) is Euler's constant. From these expressions, the method of moments estimates for μ and σ can be determined easily. For all the values presented in Tables 1–3, the method of moments estimates for μ and σ were used as initial values. Our objectives here are to obtain the estimates of the parameters and to observe the performance of the methods of estimation. The bias and mean square error of the parameter estimates are obtained for the different methods of estimation.

At this stage, some comments on the convergence of the EM gradient algorithm need to be made. This algorithm depends on the choice of initial values and shows a problem of convergence when a good choice is not made. For example, for far-off initial values of the parameters, this algorithm shows a problem with convergence. However, it does converge to the true parameter value quite accurately when the initial values of the parameters are chosen well. For example, in our empirical study, we used the method of moments estimates as the initial values, and the EM gradient algorithm works satisfactorily under this choice. Also, for small sample sizes, this algorithm again faces problems with convergence. We have observed in our simulation study that, for samples of size 100, this method sometimes does not converge. However, for larger sample sizes (such as 200 or 300), we have seen this method to work very well and a problem with convergence is rarely seen. The cases where the EM algorithm failed to converge are excluded from the simulation results. In fact, through a simulation study, we noted the number of times the method failed to converge, and these are reported in Table 1.

It can be noted that, for $n = 100$, the method failed to converge relatively often when the extreme value distribution parameters are 3.69 and 0.25, respectively. We have observed that for these simulation settings, the essential sample size (the observations which are neither truncated nor censored) sometimes could be as low as 4. Thus, divergence can occur due to this very small essential sample size. For larger sample sizes, the essential sample size also increases, thus increasing the chances of convergence of the method. Also, it was observed in the simulation study that in a few cases estimates given by the EM algorithm were highly biased, and substitution of those biased estimates into the missing information principle resulted in negative variance estimates; these cases were discarded from the simulation results. It was also observed that this happened only for smaller sample sizes.

Some comments on the parametric bootstrap confidence intervals should be made here. For samples of size 100, carrying out the necessary calculations for the parametric bootstrap confidence intervals does not seem to be possible since the EM gradient algorithm has problems with convergence for this sample size. For samples of size 200 (or 300), the calculations can be possibly carried out for parametric bootstrap confidence intervals. However, we observed the time required for this computation to be unduly long, which prohibited us from evaluating the performance of this method. We plan to carry this out with a parallel computing facility in the future and hope to report those findings in a future paper.

The bias and mean square error of the parameter estimates corresponding to the EM algorithm and the NR methods are presented in Table 2. The tolerance limit to achieve convergence was set to 0.001. One can choose a smaller tolerance value, which would result in a larger number of iterations with a gradient vector even more close to zero.

It can be observed from Table 2 that the bias and mean square error of the parameter estimates obtained from the two methods are quite close, with some occasional disagreement.

Table 3 gives the coverage probabilities of the asymptotic confidence intervals for μ corresponding to the EM algorithm and the NR method, for different nominal confidence levels and different simulation settings.

Analogous to Table 3, Table 4 presents the coverage probabilities for the two asymptotic confidence intervals for the parameter σ.

From Tables 3 and 4, it can be noticed that the coverage probabilities for μ for the two methods are always close to each other, and are lower than the nominal level. However, for σ, the coverage probabilities corresponding to the EM algorithm are always slightly higher than that of the NR method, and are in general better for both methods. It can also be noticed that

Table 2
Bias (B) and mean square error (MSE) for the EM algorithm and the NR method.

(μ, σ)	Trunc. (%)	Method	B($\hat{\mu}$)	B($\hat{\sigma}$)	MSE($\hat{\mu}$)	MSE($\hat{\sigma}$)
$n = 100$						
(3.55, 0.33)	30	EM	−0.014	−0.002	0.003	0.002
		NR	−0.012	−0.001	0.003	0.002
	40	EM	−0.016	−0.004	0.003	0.002
		NR	−0.014	−0.003	0.003	0.002
(3.69, 0.25)	30	EM	−0.014	−0.006	0.002	0.001
		NR	−0.011	−0.004	0.002	0.001
	40	EM	−0.010	−0.004	0.002	0.001
		NR	−0.008	−0.002	0.002	0.001
$n = 200$						
(3.55, 0.33)	30	EM	−0.015	−0.005	0.002	0.001
		NR	−0.013	−0.004	0.002	0.001
	40	EM	−0.012	−0.003	0.001	0.001
		NR	−0.010	−0.001	0.001	0.001
(3.69, 0.25)	30	EM	−0.014	−0.005	0.001	0.001
		NR	−0.011	−0.002	0.001	0.001
	40	EM	−0.012	−0.005	0.001	0.001
		NR	−0.010	−0.003	0.001	0.001
$n = 300$						
(3.55, 0.33)	30	EM	−0.015	−0.004	0.001	0.001
		NR	−0.013	−0.002	0.001	0.001
	40	EM	−0.013	−0.004	0.001	0.001
		NR	−0.012	−0.002	0.001	0.001
(3.69, 0.25)	30	EM	−0.013	−0.004	0.001	0.000
		NR	−0.010	−0.002	0.001	0.000
	40	EM	−0.011	−0.003	0.001	0.000
		NR	−0.009	−0.001	0.001	0.000

Table 3
Coverage probabilities for the two asymptotic confidence intervals for μ for different nominal confidence levels and different simulation settings.

μ	Truncation (%)	Nominal CL (%)	Coverage probability	
			EM	NR
$n = 100$				
3.55	30	90	0.873*	0.882
		95	0.933*	0.932*
	40	90	0.876*	0.874*
		95	0.923*	0.927*
3.69	30	90	0.862*	0.868*
		95	0.913*	0.920*
	40	90	0.879*	0.878*
		95	0.934*	0.934*
$n = 200$				
3.55	30	90	0.854*	0.860*
		95	0.913*	0.918*
	40	90	0.867*	0.873*
		95	0.936*	0.935*
3.69	30	90	0.838*	0.853*
		95	0.913*	0.922*
	40	90	0.860*	0.862*
		95	0.909*	0.924*
$n = 300$				
3.55	30	90	0.839*	0.848*
		95	0.907*	0.917*
	40	90	0.852*	0.860*
		95	0.914*	0.916*
3.69	30	90	0.848*	0.865*
		95	0.905*	0.913*
	40	90	0.857*	0.871*
		95	0.927*	0.939

* Values are significantly different from the nominal level with 95% confidence.

Table 4
Coverage probabilities for the two asymptotic confidence intervals for σ for different nominal confidence levels and different simulation settings.

σ	Truncation (%)	Nominal CL (%)	Coverage probability	
			EM	NR
$n = 100$				
0.33	30	90	0.928*	0.898
		95	0.954	0.946
	40	90	0.920*	0.877*
		95	0.959	0.934*
0.25	30	90	0.912	0.890
		95	0.949	0.930*
	40	90	0.935*	0.894
		95	0.975*	0.947
$n = 200$				
0.33	30	90	0.918	0.893
		95	0.951	0.937
	40	90	0.917	0.882
		95	0.952	0.921*
0.25	30	90	0.903	0.879*
		95	0.946	0.930*
	40	90	0.927*	0.881*
		95	0.960	0.937
$n = 300$				
0.33	30	90	0.913	0.884
		95	0.954	0.936*
	40	90	0.924*	0.890
		95	0.959	0.932*
0.25	30	90	0.919*	0.888
		95	0.961	0.946
	40	90	0.926*	0.892
		95	0.964*	0.940

* Values are significantly different from the nominal level with 95% confidence.

for higher truncation percentages, the coverage probabilities for σ corresponding to the EM algorithm increase. This might be explained due to the fact that for higher truncation percentages, with the increase of the proportion of incomplete data, the confidence intervals for σ get wider. In general, however, the coverage probabilities for both μ and σ remain reasonably close to the nominal level, for both methods.

6. Illustrative examples

In this section, we give some numerical examples to illustrate the methods of inference developed in the preceding sections. As mentioned earlier, the EM algorithm and the NR method converge to almost the same value. Though the estimates of μ and σ given by the two methods are quite close, the corresponding confidence intervals are not so. In this section, we also obtain the corresponding parametric bootstrap confidence intervals for μ and σ, although it required heavy computational effort even for this one dataset.

For the numerical illustration, we use a sample of size 100, with truncation percentage 40. The true value of the parameter vector $\boldsymbol{\theta} = (\mu, \sigma)$ is taken as (3.55, 0.33). The sample data are given in Table A.1 in the Appendix.

In Tables 5 and 6, the successive steps of iteration of the EM algorithm and the NR method, respectively, are displayed. For this illustration, we use a manually written code in R software for the NR method, rather than using the default "maxNR" function as in the simulation study. The moment estimates are taken as the initial values for μ and σ, which in this case are 3.340 and 0.271, respectively. The tolerance level used here is 0.000001.

Note that the final estimate obtained is (3.537, 0.342), correct up to three decimal places. Here it can be mentioned that if we had set the tolerance level at 0.001, as in the case of the simulation study, we would have terminated in 13 steps, with the estimate (3.536, 0.341), which would have been quite close to those obtained with the lower tolerance value.

It can be noted that the final estimates obtained by the two methods are quite close.

Table 7 gives the asymptotic confidence intervals corresponding to the EM algorithm, the NR method, and the parametric bootstrap technique.

It can be noticed that while the confidence intervals for μ based on the EM and the NR methods are the same, the confidence intervals for σ based on the EM algorithm are wider than those by the NR method. This explains the observed closeness of the coverage probabilities for μ for the two methods in Table 3, as well as the reason behind the higher coverage probability for σ for the EM algorithm, observed in Table 4. The parametric bootstrap confidence intervals for μ are seen

Table 5
Successive steps of iteration of the EM algorithm.

Step	$(\hat{\mu}, \hat{\sigma})$	Tolerance	Step	$(\hat{\mu}, \hat{\sigma})$	Tolerance
1	(3.4559, 0.2554)	0.11740	17	(3.5371, 0.3416)	0.00016
2	(3.4760, 0.2791)	0.03106	18	(3.5372, 0.3417)	0.00011
3	(3.4925, 0.2974)	0.02462	19	(3.5373, 0.3417)	7.67e−05
4	(3.5054, 0.3107)	0.01852	20	(3.5373, 0.3418)	5.27e−05
5	(3.5150, 0.3201)	0.01344	21	(3.5373, 0.3418)	3.62e−05
6	(3.5218, 0.3267)	0.00954	22	(3.5374, 0.3418)	2.49e−05
7	(3.5266, 0.3314)	0.00668	23	(3.5374, 0.3418)	1.71e−05
8	(3.5300, 0.3346)	0.00465	24	(3.5374, 0.3418)	1.17e−05
9	(3.5323, 0.3368)	0.00322	25	(3.5374, 0.3418)	8.06e−06
10	(3.5339, 0.3384)	0.00222	26	(3.5374, 0.3418)	5.54e−06
11	(3.5350, 0.3395)	0.00153	27	(3.5374, 0.3418)	3.80e−06
12	(3.5357, 0.3402)	0.00106	28	(3.5374, 0.3418)	2.61e−06
13	(3.5362, 0.3407)	0.00073	29	(3.5374, 0.3418)	1.80e−06
14	(3.5366, 0.3411)	0.00050	30	(3.5374, 0.3418)	1.23e−06
15	(3.5369, 0.3413)	0.00034	31	(3.5374, 0.3418)	8.47e−07
16	(3.5370, 0.3415)	0.00024			

Table 6
Successive steps of iteration of the NR method.

Step	$(\hat{\mu}, \hat{\sigma})$	Tolerance
1	(3.5044, 0.2532)	0.16587
2	(3.5196, 0.2988)	0.04807
3	(3.5321, 0.3301)	0.03374
4	(3.5369, 0.3409)	0.01177
5	(3.5374, 0.3418)	0.00107
6	(3.5374, 0.3418)	7.96e−06
7	(3.5374, 0.3418)	4.31e−10

Table 7
Confidence intervals obtained by different methods.

Parameter	Nominal CL (%)	EM	NR	Bootstrap
$\mu = 3.55$	90	(3.458, 3.617)	(3.458, 3.617)	(3.465, 3.629)
	95	(3.443, 3.632)	(3.443, 3.632)	(3.449, 3.644)
$\sigma = 0.33$	90	(0.265, 0.419)	(0.275, 0.409)	(0.277, 0.411)
	95	(0.250, 0.433)	(0.262, 0.422)	(0.264, 0.424)

here to be slightly wider than the other two, while for σ it is comparable to the confidence intervals based on the NR method. Thus, we would expect the parametric bootstrap method to have a higher coverage probability for μ.

6.1. Example of the prediction problem

Refer to the 92nd unit in Table A.1. For this unit, Y_{ins} is 1964, i.e., the unit is left truncated; also, it is right censored, with the censoring year being 2008. The probability that this unit will be working till 2016 is obtained, by using (19) and the estimated parameters μ and σ, to be 0.273. The standard error of this probability estimate, obtained by using (20) and the estimated variance–covariance matrix of the MLEs as $\begin{pmatrix} 0.0023 & 0.0001 \\ & 0.0022 \end{pmatrix}$, is given by 0.110. In fact, an approximate 95% confidence interval for this probability is (0.057, 0.489). Similarly, for the 42nd unit, for which the installation year is 1989 (i.e., not left truncated and also right censored), the probability that the unit will be working till 2016 is estimated to be 0.728, with the standard error of 0.033. An approximate 95% confidence interval for this probability is (0.663, 0.793). It may be of interest to note here that the second unit (installed in 1989) has a higher probability to work till 2016 than the first unit (installed in 1964), as one would expect.

7. Concluding remarks

In this paper, the necessary steps of the EM algorithm are developed for fitting the Weibull distribution to left truncated and right censored data. The maximization part of the EM algorithm is carried out by the one-step Newton–Raphson method, thus making use of the EM gradient algorithm. The Weibull distribution is also fitted to the left truncated and right censored data by the NR method. The efficiency of the parameter estimates are examined and the two methods of estimation are compared through an extensive Monte Carlo simulation study. The asymptotic variance–covariance matrix of the MLEs are obtained by the missing information principle, and then the asymptotic confidence intervals are constructed for the parameters based on the EM algorithm. The asymptotic confidence intervals are also obtained based on the NR method by using the observed information matrix, and the two confidence intervals are then compared in terms of coverage probabilities through a simulation study. It is observed that the EM algorithm and the NR method yield quite close results in all respects, even though the coverage probabilities corresponding to these two methods slightly differ in some cases.

Balakrishnan and Mitra (2011) have recently discussed the fitting of a lognormal distribution for left truncated and right censored data via the EM algorithm, and studied the proposed likelihood methods of inference in detail. For a given dataset, it might be of interest to know which of the two distributions – the Weibull or the lognormal – provides the better fit. However, as the two models are not nested, one cannot perform a likelihood-based statistical test for this discrimination problem. The generalized gamma distribution, as considered by Balakrishnan and Peng (2006), constitutes a bigger class of distributions, which include the exponential, gamma, lognormal and Weibull distributions as special cases. Thus, it would be of interest to develop the EM algorithm for fitting the generalized gamma distribution to left truncated and right censored data. This would then facilitate the model discrimination problem satisfactorily using likelihood-based tests.

Another problem of interest would be to incorporate covariates into the Weibull reliability analysis in the presence of truncation and censoring. In fact, Hong et al. (2009) did use covariates for prediction purposes. As a follow up of the present work, we plan to consider the EM algorithm approach in this general case of the model with covariates. We hope to report these findings in a future paper.

Acknowledgments

We thank the referees and the associate editor for insightful comments that helped to improve the earlier version of this paper substantially.

Appendix

In Section 3.1, the M-step of the EM algorithm is carried out by a one-step Newton–Raphson method, using Eqs. (7)–(11). At the k-th step of iteration, the score vector and the observed information matrix, required for the Newton–Raphson method, are evaluated at the currently available parameter value, i.e., at $\boldsymbol{\theta}^{(k)}$. Therefore, at the k-th step of iteration, the integrals to be evaluated are of the form $\int_a^\infty we^{-w}dw$, $\int_a^\infty w(\log w)e^{-w}dw$ and $\int_a^\infty w(\log w)^2e^{-w}dw$, where $a > 0$ (for Eqs. (7)–(11), with $a = e^{\xi_i^{(k)}}$, $i = 1, \ldots, n$). Clearly, the first of the above three integrals is the upper incomplete gamma function $\Gamma(2, a)$. The other integrals can be evaluated using the following lemma.

Lemma. *We have*
(i) $\int_a^\infty w(\log w)e^{-w}dw = e^{-a}(a\log a + 1) + I_0(a)$,
(ii) $\int_a^\infty w(\log w)^2e^{-w}dw = a(\log a)^2e^{-a} + I_1(a) + 2I_0(a)$,
where

$$I_0(a) = e^{-a}\left[\Psi(1) - (e^a - 1)\log a + \sum_{n=0}^{\infty}\frac{a^{n+1}\Psi(n+2)}{\Gamma(n+2)}\right],$$

$$I_1(a) = e^{-a}\Bigg[\Psi^2(1) + \Psi'(1) - 2\Psi(1)\log a(e^a - 1) - (\log a)^2(e^a - 1)$$
$$+ 2(\Psi(1) + \log a)\sum_{n=0}^{\infty}\frac{a^{n+1}\Psi(n+2)}{\Gamma(n+2)} + \sum_{n=0}^{\infty}\frac{a^{n+1}\Psi'(n+2)}{\Gamma(n+2)} - \sum_{n=0}^{\infty}\frac{a^{n+1}\Psi^2(n+2)}{\Gamma(n+2)}\Bigg].$$

Proof. Using integration by parts, we get

$$\int_a^\infty w(\log w)e^{-w}dw = a(\log a)e^{-a} + e^{-a} + \int_a^\infty (\log w)e^{-w}dw,$$
$$\int_a^\infty w(\log w)^2e^{-w}dw = a(\log a)^2e^{-a} + \int_a^\infty (\log w)^2e^{-w}dw + 2\int_a^\infty (\log w)e^{-w}dw.$$

Define

$$I_r(a) = \int_a^\infty (\log w)^{r+1}e^{-w}dw = e^{-a}\int_a^\infty (\log w)^{r+1}e^{-(w-a)}dw = e^{-a}E[Z^{r+1}],$$

where $Z = \log W$ is distributed as a left-truncated standard exponential random variable, with left truncation at a. The mgf of Z is obtained to be

$$M_Z(\theta) = E[e^{\theta Z}] = e^a\Gamma(\theta + 1, a) = e^a\Gamma(\theta + 1)\left[1 - e^{-a}a^{\theta+1}\sum_{n=0}^{\infty}\frac{a^n}{\Gamma(\theta + n + 2)}\right].$$

Table A.1
A simulated dataset, for sample size 100, truncation percentage 40, and the true parameter value of (μ, σ) as (3.55, 0.33). The truncation time, lifetime and censoring time variables are all presented on the log scale.

Serial no.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
1	1984	1	*.	*	0	3.178	3.178
2	1990	1	*.	2001	1	2.398	*
3	1983	1	*.	2002	1	2.944	*
4	1981	1	*.	2000	1	2.944	*
5	1985	1	*.	*	0	3.135	3.135
6	1991	1	*.	*	0	2.833	2.833
7	1982	1	*.	*	0	3.258	3.258
8	1990	1	*.	*	0	2.890	2.890
9	1983	1	*.	1999	1	2.773	*
10	1992	1	*.	*	0	2.773	2.773
11	1983	1	*.	*	0	3.219	3.219
12	1989	1	*.	*	0	2.944	2.944
13	1985	1	*.	*	0	3.135	3.135
14	1982	1	*.	*	0	3.258	3.258
15	1983	1	*.	*	0	3.219	3.219
16	1981	1	*.	*	0	3.296	3.296
17	1985	1	*.	*	0	3.135	3.135
18	1981	1	*.	*	0	3.296	3.296
19	1988	1	*.	2002	1	2.639	*
20	1983	1	*.	*	0	3.219	3.219
21	1984	1	*.	*	0	3.178	3.178
22	1989	1	*.	*	0	2.944	2.944
23	1988	1	*.	*	0	2.996	2.996
24	1982	1	*.	*	0	3.258	3.258
25	1981	1	*.	*	0	3.296	3.296
26	1986	1	*.	*	0	3.091	3.091
27	1987	1	*.	*	0	3.045	3.045
28	1990	1	*.	1997	1	1.946	*
29	1980	1	*.	1996	1	2.773	*
30	1980	1	*.	*	0	3.332	3.332
31	1981	1	*.	*	0	3.296	3.296
32	1983	1	*.	1997	1	2.639	*
33	1980	1	*.	*	0	3.332	3.332
34	1984	1	*.	*	0	3.178	3.178
35	1982	1	*.	*	0	3.258	3.258
36	1980	1	*.	*	0	3.332	3.332
37	1985	1	*.	2007	1	3.091	*
38	1993	1	*.	*	0	2.708	2.708
39	1983	1	*.	*	0	3.219	3.219
40	1980	1	*.	*	0	3.332	3.332
41	1981	1	*.	2001	1	2.996	*
42	1989	1	*.	*	0	2.944	2.944
43	1993	1	*.	*	0	2.708	2.708
44	1983	1	*.	*	0	3.219	3.219
45	1993	1	*.	*	0	2.708	2.708
46	1987	1	*.	*	0	3.045	3.044
47	1994	1	*.	*	0	2.639	2.639
48	1985	1	*.	2007	1	3.091	*
49	1981	1	*.	*	0	3.296	3.296
50	1983	1	*.	2004	1	3.045	*
51	1982	1	*.	*	0	3.258	3.258
52	1981	1	*.	*	0	3.296	3.296
53	1986	1	*.	*	0	3.091	3.091
54	1980	1	*.	1990	1	2.303	*
55	1980	1	*.	1994	1	2.639	*
56	1982	1	*.	*	0	3.258	3.258
57	1990	1	*.	2008	1	2.890	*
58	1985	1	*.	*	0	3.135	3.135
59	1983	1	*.	*	0	3.219	3.219
60	1982	1	*.	*	0	3.258	3.258
61	1963	0	2.833	1996	1	3.497	*
62	1963	0	2.833	2001	1	3.638	*
63	1961	0	2.944	1998	1	3.611	*
64	1961	0	2.944	1992	1	3.434	*
65	1960	0	2.996	1984	1	3.178	*
66	1964	0	2.773	2004	1	3.689	*
67	1961	0	2.944	1994	1	3.497	*
68	1977	0	1.099	1998	1	3.045	*

(*continued on next page*)

Table A.1 (*continued*)

Serial no.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
69	1963	0	2.833	1987	1	3.178	*
70	1960	0	2.996	1991	1	3.434	*
71	1961	0	2.944	1983	1	3.091	*
72	1964	0	2.773	1995	1	3.434	*
73	1963	0	2.833	1998	1	3.555	*
74	1961	0	2.944	2001	1	3.689	*
75	1960	0	2.996	1988	1	3.332	*
76	1974	0	1.792	2006	1	3.466	*
77	1978	0	0.693	1995	1	2.833	*
78	1962	0	2.890	1993	1	3.434	*
79	1963	0	2.833	*	0	3.807	3.807
80	1960	0	2.996	1998	1	3.638	*
81	1962	0	2.890	2007	1	3.807	*
82	1960	0	2.996	1990	1	3.401	*
83	1962	0	2.890	1980	1	2.890	*
84	1961	0	2.944	1981	1	2.996	*
85	1964	0	2.773	1989	1	3.219	*
86	1964	0	2.773	1987	1	3.135	*
87	1960	0	2.996	2006	1	3.829	*
88	1961	0	2.944	1992	1	3.434	*
89	1964	0	2.773	*	0	3.784	3.784
90	1963	0	2.833	1991	1	3.332	*
91	1973	0	1.946	*	0	3.555	3.555
92	1964	0	2.773	*	0	3.784	3.784
93	1972	0	2.079	1984	1	2.485	*
94	1962	0	2.890	2007	1	3.807	*
95	1963	0	2.833	1997	1	3.526	*
96	1964	0	2.773	1987	1	3.135	*
97	1964	0	2.773	2002	1	3.638	*
98	1971	0	2.197	*	0	3.611	3.611
99	1965	0	2.708	1990	1	3.219	*
100	1962	0	2.890	1994	1	3.466	*

From the above expression of the mgf, it can be shown that

$$E[Z] = \Psi(1) - (e^a - 1)\log a + \sum_{n=0}^{\infty} \frac{a^{n+1}\Psi(n+2)}{\Gamma(n+2)},$$

$$\begin{aligned} E[Z^2] = &\ \Psi^2(1) + \Psi'(1) - 2\Psi(1)\log a(e^a - 1) - (\log a)^2(e^a - 1) \\ &+ 2(\Psi(1) + \log a)\sum_{n=0}^{\infty}\frac{a^{n+1}\Psi(n+2)}{\Gamma(n+2)} + \sum_{n=0}^{\infty}\frac{a^{n+1}\Psi'(n+2)}{\Gamma(n+2)} - \sum_{n=0}^{\infty}\frac{a^{n+1}\Psi^2(n+2)}{\Gamma(n+2)}. \end{aligned}$$

From here, it readily follows that

$$I_0(a) = \int_a^{\infty} (\log w)e^{-w}dw = e^{-a}E[Z],$$

and

$$I_1(a) = \int_a^{\infty} (\log w)^2 e^{-w}dw = e^{-a}E[Z^2].$$

Hence, the proof. □

Table A.1 presents an example of left truncated and right censored data wherein * means not applicable. All the results of Section 5 have been obtained based on these data.

References

Balakrishnan, N., Cohen, A.C., 1991. Order Statistics and Inference: Estimation Methods. Academic Press, Boston.

Balakrishnan, N., Kundu, D., Ng, H.K.T., Kannan, N., 2007. Point and interval estimation for a simple step-stress model with Type-II censoring. Journal of Quality Technology 39, 35–47.

Balakrishnan, N., Mitra, D., 2011. Likelihood inference for lognormal data with left truncation and right censoring with an illustration. Journal of Statistical Planning and Inference 141, 3536–3553.
Balakrishnan, N., Peng, Y., 2006. Generalized gamma frailty model. Statistics in Medicine 25, 2797–2816.
Cohen, A.C., 1991. Truncated and Censored Samples. Marcel Dekker, New York.
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.
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.
Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, Vol. 1, second ed. John Wiley & Sons, New York.
Lange, K., 1995. A gradient algorithm locally equivalent to the EM algorithm. Journal of the Royal Statistical Society, Series B 57, 425–437.
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.
McLachlan, G.J., Krishnan, T., 2008. The EM Algorithm and Extensions, Second ed. John Wiley & Sons, Hoboken, New Jersey.
Meeker, W.Q., Escobar, L.A., 1998. Statistical Methods for Reliability Data. John Wiley & Sons, New York.
Ng, H.K.T., Chan, P.S., Balakrishnan, N., 2002. Estimation of parameters from progressively censored data using EM algorithm. Computational Statistics & Data Analysis 39, 371–386.