LIKELIHOOD INFERENCE FOR LEFT TRUNCATED AND RIGHT CENSORED LIFETIME DATA

By

DEBANJAN MITRA, B.Sc., M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

McMaster University

© Copyright by DEBANJAN MITRA, October 1

DOCTOR OF PHILOSOPHY (2012) McMaster University

(Mathematics) Hamilton, Ontario

TITLE: LIKELIHOOD INFERENCE FOR LEFT TRUNCATED AND RIGHT CENSORED LIFETIME DATA

AUTHOR: DEBANJAN MITRA,

B.Sc. (University of Calcutta)

M.Sc. (University of Calcutta)

SUPERVISOR: Professor N. Balakrishnan

NUMBER OF PAGES: 17 , 159

Abstract

Left truncation and right censoring arise naturally in lifetime data. Left truncation arises because in many situations, failure of a unit is observed only if it fails after a certain period. Often, the units under study may not be followed until all of them fail and the experimenter may have to stop at a certain time when some of the units may still be working. This introduces right censoring into the data.

Some commonly used lifetime distributions are lognormal, Weibull and gamma, all of which are special cases of the flexible generalized gamma family. Likelihood inference via the Expectation Maximization (EM) algorithm is developed to estimate the model parameters of lognormal, Weibull, gamma and generalized gamma distributions, based on left truncated and right censored data. The asymptotic variance-covariance matrices of the maximum likelihood estimates (MLEs) are derived using the missing information principle. By using the asymptotic properties of the MLEs, asymptotic confidence intervals for the parameters are constructed. For comparison purpose, Newton-Raphson (NR) method is also used for the parameter estimation, and confidence intervals based on the NR method and parametric bootstrap are also obtained. Through Monte Carlo simulations, the performance of all these methods of inference are studied. With regard to prediction analysis, the probability that a right censored unit will be working until a future year is estimated, and an asymptotic confidence interval for the probability is then derived by the delta method. All the methods of inference developed here are illustrated with some numerical examples.

KEY WORDS: Lifetime data; Left truncation; Right censoring; Lognormal distribution; Weibull distribution; Gamma distribution; Generalized gamma distribution; Maximum likelihood estimates; EM algorithm; Missing information principle; Information matrix; Asymptotic variances and covariances; Asymptotic confidence intervals; Newton-Raphson method; Coverage probabilities; Monte Carlo simulations; Parametric bootstrap; Prediction; Delta method

Acknowledgments

I take this opportunity to mention the names of my parents, Advocate Mr. Mukti Ranjan Mitra and Mrs. Sarbari Mitra, in an effort to indicate their contribution towards my work and life. Words like thankfulness or gratefulness fall short in expressing my gratitude for them, for the unwavering and unconditional support they have given me over the years.

I have no words to express my gratitude for my doctoral supervisor, Professor N. Balakrishnan. Without him being my teacher, this whole process would not have been possible. His care, affection and unfailing support at all times helped me go through this process for last three and a half years. In these years, the bond that has been created between us will last for a lifetime, I am sure.

I am deeply appreciative of my doctoral supervisory committee members, Dr. Roman Viveros-Aguilera and Dr. Aaron Childs, for their careful watch throughout the process. Their incisive and invaluable comments at various stages helped the thesis to improve quite significantly.

I thank all those friends, who have always supported me in different stages of my life.

Last but not least, being a religious person, I am ever thankful to the God Almighty, for I believe that without Him in my life, I would not be the same person I am today.

Publications from the thesis

- Balakrishnan, N. and 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. and Mitra D. (2012). Some further issues concerning likelihood inference for left truncated and right censored lognormal data. *Communications in Statistics - Simulation and Computation*, (to appear).
- Balakrishnan N. and Mitra D. (2012). Left truncated and right censored Weibull data and likelihood inference with an illustration. *Computational Statistics & Data Analysis*, 56: 4011 - 4025.
- Balakrishnan N. and Mitra D. (2012). Likelihood inference based on left truncated and right censored data from gamma distribution. *IEEE Transactions on Reliability* (Submitted for publication, under review).

Materials of Chapter 2 are taken from the first two items of the above list. Materials of Chapter 3 are taken from the third item of the above list, while materials of Chapter 4 are taken from the last item.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Lifetime data

As the name suggests, lifetime data pertain to the lifetimes of units, either industrial or biological. Needless to say, an industrial or a biological unit can not be in operation forever; or, speaking in more general terms, such a unit can not continue to operate in the same condition for ever. The breakdown of its operation, or some predefined change in the mode of operation of such a unit can be termed, in general, as a failure. The time span, starting from the beginning of operation of a unit until its failure, is termed as lifetime of that unit. The study of such lifetime data of industrial units is called reliability analysis, whereas the same corresponding to biological units is known as survival analysis.

A failure does not necessarily mean the end of the actual lifetime of an industrial unit, or the death of a biological unit. For example, for a set of electric bulbs, one can define failure as the breakage of a bulb, and in that case, a failure would mean the termination of actual lifetime of a unit. However, one can also define failure in this case to be the decrease in the illuminating power of a bulb by certain amount. So, using this definition, failure may occur even if the bulb does not break, but its power diminishes. Similarly, for a biological unit, a failure may be defined as the death of that unit; or it may also be defined as the occurence of a certain desease.

The failure, thus can be termed in general, as an event. Lifetime data are often referred to as failure time data, or time-to-event data, since essentially, this type of data are measures of certain variable until the occurence of the event of interest - the span we call the "lifetime".

This general definition of lifetime data suggests that it need not be either the real time or the clock time, nor it need be a continuous variable always. A classic example of this is the time-to-event data of ball bearings, where the lifetime can be defined as the number of revolutions per unit time.

Study of lifetime data is important in many respects. Often, the primary interest is to know the underlying distribution from which the data come from. With this knowledge, one can aim to infer about different aspects of the units under study. For example, once the underlying distribution is known (i.e., an appropriate distributional model is fitted), one may be interested in the probability for a certain unit to be in operation until a future year. In this case, we have a parametric approach to the problem. Another possibility is to estimate certain quantities related to the lifetime variable, such as the median or the quantile function, without any distributional assumption. This nonparametric method of inference on lifetime data is also used in practice, and researchers have given considerable attention to both parametric and nonparametric inferential methods.

1.2 Models for lifetime data

Not all statistical distributions are used as models for lifetime data. There are certain distributions, which are useful in depicting the lifetime data due to some desirable properties that they posses. Here, we briefly describe some statistical distributions that are commonly used for modelling lifetime data. For more details on these distributions, one may refer to Johnson et al (1994), Balakrishnan and Caroni (2013), and Cohen and Whitten (1988).

1.2.1 Exponential distribution

The exponential distribution is the most commonly used distribution for modelling lifetime data. The probability density function (pdf) of a one-parameter exponential distribution is given by

$$f(x) = \frac{1}{\theta}e^{-x/\theta}, \quad x > 0, \theta > 0.$$

The corresponding cumulative density function (cdf) is given by

$$F(x) = 1 - e^{-x/\theta}, \quad x > 0.$$

In the study of lifetime data, it is often customary to deal with the reliability or survival function, which is simply given by

$$S(x) = 1 - F(x) = e^{-x/\theta}, \quad x > 0.$$

An exponential distribution with the above pdf has the mean and variance as follows:

$$E(X) = \theta, V(X) = \theta^2.$$

Another quantity which is often used in reliability or survival analysis is the hazard function, which in this case is

$$h(x) = \frac{f(x)}{S(x)} = \frac{1}{\theta}.$$

The hazard function can be interpreted as the instantaneous rate of failure. It is of interest to note that the hazard function for an exponential distribution is constant with respect to x, which is indeed a characterization of the exponential distribution.

1.2.2 Gamma distribution

Another distribution which is widely used for modeling lifetime data is the gamma distribution, with pdf

$$f_X(x) = \frac{x^{\kappa-1}\exp(-x/\theta)}{\theta^\kappa \Gamma(\kappa)}, \quad x > 0, \theta > 0, \kappa > 0, \tag{1.2.1}$$

where θ and κ are the scale and shape parameters, respectively. The cdf of the gamma distribution is

$$F_X(x) = \frac{\gamma(\kappa, x/\theta)}{\Gamma(\kappa)}, \qquad x > 0, \tag{1.2.2}$$

where $\gamma(p, x) = \int_0^x u^{p-1} e^{-u} du$ is the lower incomplete gamma function. Thus, the survival function of gamma distribution is given by

$$S(x) = \frac{\Gamma(\kappa, x/\theta)}{\Gamma(\kappa)}, \qquad x > 0,$$

where $\Gamma(\kappa, x/\theta) = \Gamma(\kappa) - \gamma(\kappa, x/\theta)$, is the upper incomplete gamma function.

A gamma distribution with the above pdf has the mean and variance as follows:

$$E(X) = \kappa\theta, V(X) = \kappa\theta^2.$$

The hazard function of the gamma distribution is given by

$$h(x) = \frac{f(x)}{S(x)} = \frac{x^{\kappa-1} \exp(-x/\theta)}{\theta^\kappa \Gamma(\kappa, x/\theta)}, \qquad x > 0,$$

which is decreasing for $0 < \kappa < 1$, constant for $\kappa = 1$, and increasing for $\kappa > 1$.

1.2.3 Weibull distribution

Weibull distribution is the most commonly used lifetime data model in both reliability and survival analyses. The pdf of the Weibull distribution, with scale parameter α and shape parameter η, is given by

$$f(x) = \left(\frac{\eta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\eta-1} \exp\left\{-\left(\frac{x}{\alpha}\right)^{\eta}\right\}, \qquad x > 0, \alpha > 0, \eta > 0. \tag{1.2.3}$$

The corrresponding cdf is given by

$$F(x) = 1 - \exp\left\{ -\left(\frac{x}{\alpha}\right)^{\eta} \right\}, \quad x > 0, \tag{1.2.4}$$

and thus the survival function is given by

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

The mean of the Weibull distribution is

$$E(X) = \alpha\Gamma\left(1 + \frac{1}{\eta}\right),$$

and the variance is given by

$$V(X) = \alpha^2\left[\Gamma\left(1 + \frac{2}{\eta}\right) - \left\{\Gamma\left(1 + \frac{1}{\eta}\right)\right\}^2\right].$$

The hazard function of the Weibull distribution is simply

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

which is increasing for $\eta > 1$, constant for $\eta = 1$, and decreasing for $\eta < 1$.

1.2.4 Lognormal distribution

The pdf of the lognormal distribution with scale parameter μ and shape parameter σ is given by

$$f_X(x) = \frac{1}{\sigma x\sqrt{2\pi}}\exp\left\{ -\frac{(\log x - \mu)^2}{2\sigma^2} \right\}, \quad x > 0, \tag{1.2.5}$$

and the corresponding cdf is

$$F(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), \quad x > 0, \tag{1.2.6}$$

where $\Phi(\cdot)$ denote the cdf of the standard normal distribution. Thus, the survival function is simply

$$S(x) = 1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right) = \Phi\left(\frac{\mu - \log(x)}{\sigma}\right), \quad x > 0.$$

Logarithmic transformation of a lognormal random variable clearly makes it a normal random variable, with μ and σ as the location and scale parameters, respectively. The mean and variance of a lognormal distribution are as follows:

$$E(X) = e^{\mu+\sigma^2/2}, V(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

The lognormal hazard function initially increases with x. But after reaching a peak, it starts to decrease, and reaches zero as $x \to \infty$. This property of the lognormal hazard function makes the use of this distribution difficult in some practical situations. But inspite of this, the lognormal distribution provides good fit to many kinds of lifetime data, and is therefore used quite extensively.

1.2.5 Generalized gamma distribution

The generalized gamma distribution has its pdf as

$$f(x) = \frac{\eta}{\alpha\Gamma(\kappa)}\left(\frac{x}{\alpha}\right)^{\kappa\eta-1} \exp\left\{-\left(\frac{x}{\alpha}\right)^{\eta}\right\}, \quad x > 0, \alpha > 0, \eta > 0, \kappa > 0, \qquad (1.2.7)$$

and the cdf is

$$F(x) = 1 - \frac{\Gamma(\kappa, \left(\frac{x}{\alpha}\right)^{\eta})}{\Gamma(\kappa)}, \quad x > 0, \qquad (1.2.8)$$

where α is the scale parameter, and η and κ are both shape parameters. The generalized gamma distribution is a general flexible model, with exponential, gamma, Weibull and lognormal distributions all belonging to it as special cases. The specific choices of the parameter values, for which these special cases are obtained, are as follows:

$\kappa = 1$	Weibull
$\eta = 1$	Gamma
$\kappa = 1,\ \eta = 1$	Exponential
$\eta \to 0,\ \kappa \to \infty$	Lognormal

The mean and variance of this distribution are given by

$$E(X) = \alpha\left[\frac{\Gamma(\kappa + \frac{1}{\eta})}{\Gamma(\kappa)}\right]$$

and

$$V(X) = \alpha^2\left[\frac{\Gamma(\kappa + \frac{2}{\eta})}{\Gamma(\kappa)} - \frac{\Gamma^2(\kappa + \frac{1}{\eta})}{\Gamma^2(\kappa)}\right],$$

respectively.

1.2.6 Extreme value distribution

Logarithmic transformation of a Weibull distribution, as presented in Section 1.2.3, gives an extreme value (Gumbel) distribution with pdf

$$f(x) = \frac{1}{\sigma}\exp\left[\left(\frac{x-\mu}{\sigma}\right) - \exp\left(\frac{x-\mu}{\sigma}\right)\right], \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0. \tag{1.2.9}$$

The corresponding cdf is given by

$$F(x) = 1 - \exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right]\Bigg\}, \quad -\infty < x < \infty. \tag{1.2.10}$$

Here, μ and σ are the location and scale parameters of the distribution, and their relation to the original Weibull parameters α and η can be easily seen to be

$$\mu = \log \alpha, \sigma = 1/\eta.$$

The mean and variance of the extreme value distribution are

$$E(X) = \mu - \gamma\sigma, V(X) = \frac{\pi^2}{6}\sigma^2,$$

where $\gamma = 0.5772$ (approx.) is the Euler's constant. Clearly, the hazard function of the distribution is

$$h(x) = \frac{1}{\sigma} \exp\left(\frac{x-\mu}{\sigma}\right),$$

which is an increasing function. The main importance of the extreme value distribution in reliability analysis lies in its relationship to the Weibull distribution.

Apart from these widely used models, Gompertz distribution, Inverse Gaussian distribution, log-logistic distribution, Birnbaum-Saunders distribution, and some others are also used in lifetime data analyses.

1.3 Censoring in the data

While dealing with lifetime data, it is often impossible to follow the units till the end of their lifetimes. For example, it is practically infeasible to follow a set of electric bulbs till all of them fail. Thus, it is common in the analysis of lifetime data to follow the units until some prespecified time, or until some predefined event occurs. At this point, the experimenter stops collecting the data, and proceeds to the analyses with whatever data that have been observed. Clearly, at this stopping point of the experiment, the data may be a mixture of two different types of observations. For some units, failure may have occurred, while for some others it may not have. The units which have not failed, when the experiment is stopped, are said to be censored.

Thus, censoring is a property of the sample collected, and it arises from the practical constraint of the experimenter being unable to follow the units till the end of their lifetimes. Censoring in a data can arise from many reasons, and depending on the reasons, censoring may be of many different types. Here, we describe some types of censoring. For detailed account of censored data, one may refer to Cohen (1991).

1.3.1 Informative and non-informative censoring

When the cause of censoring is not related to the lifetimes of the units, the censoring is said to be non-informative censoring. All the standard techniques of analysing censored data are based on the assumption that the censoring is non-informative. On the other hand, when the censoring cause is related to the lifetimes, it is said to be informative censoring. To give examples, when an experimenter observes that some of the units are not operating properly and may fail soon, then he/she may withdraw those units from the experiment. In this case, these units become censored, but the cause of censoring is directly related to the lifetimes of those units. The experimenter withdrew those units knowing that had they been allowed to stay under test, those units may have failed, that too soon enough. This is an example of informative censoring.

Also, sometimes, a patient may withdraw from a medical study realizing that the treatment is not satisfactory and may choose to go for another treatment. This is also informative censoring. But if the patient withdraws from the medical study due to some other reasons, for example, if she moves out to another place (i.e., the reason for her withdrawal is not related to her lifetime), then the censoring becomes non-informative.

1.3.2 Time censoring

Sometimes, the experimental units are followed until some prespecified (clock) time at which the experimenter stops collecting the data. The units which remain unfailed at this point automatically get censored. To be more specific, these units become *right censored*, as their actual failure times fall to the right of this censoring point. Suppose the i-th unit has the lifetime variable T_i, and depending on the time until which it is followed, its censoring time is C_i. Sometimes, the censoring time for all units may be fixed, i.e., C. Then, for a right censored unit, the minimum of its actual lifetime T_i and its censoring time C_i will be observed, i.e, the observed lifetime for a right censored unit will be $\min(T_i, C_i)$. In other words, we observe the failure only when $T_i \leq C_i$; the unit becomes right censored when $T_i > C_i$. Here, all that is known about a right censored unit is that its lifetime is greater than C_i, the censoring time. This censoring is also termed as *Type-I censoring* in the literature.

Like right censoring, the censoring in time can come from left also, which is termed as *left censoring*. In this case, it is known that a unit failed before some prespecified time C (or in general, C_i), but the actual time of failure of the left censored unit is unknown. Another form of time censoring is *interval censoring*, wherein only the number of failures in specific intervals are available, without any more specific information about the individual failure times.

Now, we present the general form of the likelihood functions for right, left and interval censored data. Consider a lifetime variable T with pdf $f(t)$ and cdf $F(t)$. Also, let U and C denote the sets that contain the uncensored and censored units, respectively. Then, under right censoring, the likelihood function L will be given by

$$L \propto \prod_{i \in U} f(t_i) \times \prod_{i \in C} \{1 - F(t_i)\}.$$

Note that this form is obtained by using the fact that a right censored observation contributes $1 - F(t)$ to the overall likelihood. With a similar argument, for a left censored data, the likelihood will be

$$L \propto \prod_{i \in U} f(t_i) \times \prod_{i \in C} \{F(t_i)\},$$

and for an interval censored data, the likelihood will be

$$L \propto \prod_{i \in U} f(t_i) \times \prod_{i \in C} \{F(t_{upp_i}) - F(t_{low_i})\}.$$

1.3.3 Failure censoring

Also termed as *Type-II censoring* in statistical literature, this type of censoring fixes the number of failures to be observed. Suppose an experiment starts with n units, and the experimenter decides to continue the experiment until the r-th failure occurs, i.e., until time $T_{r:n}$, which is the r-th smallest order statistic among the lifetimes of these n units. When the experiment stops, a prespecified number of failures (r) is obtained, and all that is known about the censored observations is that their lifetime $T > T_{r:n}$. This type of censoring is used in many reliability experiments. For this type of censoring, the likelihood function is given by

$$L \propto \prod_{i=1}^{r} f(t_i) \times \{1 - F(t_{r:n})\}^{n-r}.$$

1.3.4 Other forms of censoring

Among other forms of censoring are the *hybrid censoring* and the *progressive censoring*. In hybrid censoring, a mixture of Type-I and Type-II censoring schemes is used, in order to draw best information from the data. The progressive censoring is the most generalized case which is used with a view to draw information from the entire range of the underlying distribution. For details on these advanced schemes, see Childs et. al. (2003), Balakrishnan and Aggarwala (2000), Balakrishnan (2007), and Balakrishnan and Kundu (2013).

1.4 Truncation in the data

Truncation is a property of the underlying distribution from which the data arise. Sometimes, due to some practical constraints, or simply due to the nature of the model, the values of a random variable become observable only when they exceed a lowerbound, or stay within an upperbound, though theoretically, the range of the random variable is not restricted. All values of such a random variable that fall outside these bounds are never observable, and consequently their existence is not known to us. This property of a statistical distribution is called truncation. Truncation can be observed not only in lifetime data, but also in other types of data. There are once again different forms of truncation in statistical data as described below.

1.4.1 Left truncation

The values of a random variable may be observable only when they are greater than a specific lowerbound, i.e., only when the values exceed a definite threshold value. All the values that fall short of that threshold are never observed. This gives rise to left truncation in the distribution of that random variable. The pdf of the left truncated random variable can be obtained by normalizing the original density of the random variable over the observable range. To be specific, let $f(\cdot)$ and $F(\cdot)$ denote the pdf and cdf of a random variable X. Then, if the random variable X becomes left truncated at a point τ^L (i.e., realizations of X are observable only when they exceed

τ^L), then the pdf of the left truncated random variable is given by

$$f_{LT}(x) = \frac{f(x)}{1 - F(\tau^L)}, \quad x > \tau^L,$$

and the corresponding left truncated cdf is given by

$$F_{LT}(x) = \frac{F(x) - F(\tau^L)}{1 - F(\tau^L)}, \quad x > \tau^L.$$

Left truncation is the most common form of truncation in lifetime data. For example, all electronic goods are sold after they are tested for some prespecified number of hours. So, when we buy a TV or a refrigerator, their lifetimes have already exceeded the threshold value that the manufacturer decided to be the testing period, say 200 hours. It is clear that some of the units may have failed during the testing period, but neither their lifetimes nor the number of such failed units are known to us. All we know is that the units sold have lifetime $X > 200$.

1.4.2 Right truncation

Though not common in lifetime data, right truncation occurs when the values of a random variable can be observed only when they are smaller than an upperbound (the right truncation point); that is, all values greater than this upperbound are not observable. If the right truncation point is τ^R, then the pdf and cdf of a right truncated random variable are

$$f_{RT}(x) = \frac{f(x)}{F(\tau^R)}, \quad x < \tau^R,$$

$$F_{RT}(x) = \frac{F(x)}{F(\tau^R)}, \quad x < \tau^R,$$

respectively.

1.4.3 Double truncation

Double truncation occurs when the values of a variable are observable only over a finite range, i.e., only when they exceed a lowerbound and are also smaller than some upperbound. This type of truncation is also not common in lifeitme data, but in other forms of data, this may be observed. To obtain the pdf and cdf of a doubly truncated random variable X, suppose the values of X are observable only when $\tau^L < x < \tau^R$. Then, we have the doubly truncated pdf to be

$$f_{DT}(x) = \frac{f(x)}{F(\tau^R) - F(\tau^L)}, \quad \tau^L < x < \tau^R,$$

and the cdf to be

$$F_{DT}(x) = \frac{F(x) - F(\tau^L)}{F(\tau^R) - F(\tau^L)}, \quad \tau^L < x < \tau^R.$$

An example of double truncation can be found in soundwave frequencies. Frequency of a soundwave can be any non-negative number. However, we can not hear all sounds. Only those sounds generated from soundwaves with frequencies greater than 20 Hertz and less than 20000 Hertz are audible to human ears. Here, the truncation point on the left is 20 Hertz and that on the right is 20000 Hertz.

1.4.4 Censoring and truncation

Though the censoring and truncation seem to be close as far as incompleteness of data is concerned, they are actually distinctly different. The difference comes from the fact that censoring is a property of the sample from a population, while truncation is a property of the population itself. To elaborate, let us consider left truncation and left censoring. In left censoring, we know that some units failed before a specified time T, and the number of such units is known to us. All that we know about these failed units is that their lifetimes belong to the interval $[0, T]$, without any more specific knowledge. On the other hand, in left truncation, we have no information at all on all these units that failed before the left truncation point, say T. We know that the existing units have exceeded the threshold T, and some units may have failed before T. But we have no knowledge on the failed units at all, including the fact that if such units even existed.

1.5 Inferential methods for incomplete data

The primary interest of parametric statistical inference is to estimate the parameters of a statistical distribution, based on available data. In doing so, one can employ different methods, depending on their suitability and necessity for a particular problem. Maximum likelihood estimation is one of the methods which is used most extensively in statistical literature due to its simplicity, flexibility and optimality.

This method estimates the parameters of a distribution by maximizing the probability or likelihood of the observed data from that distribution, with respect to the parameters. Maximum likelihood estimates (MLEs) are thus easy to obtain, though in most situations except some standard cases, closed-form expressions (or explicit solutions to the maximum likelihood equations) can not be obtained, and one has to use a numerical technique. Another importance of MLEs lies in their asymptotic properties. The asymptotic distribution of MLEs, under some regularity conditions, can be shown to be normal, with specific mean and variance: a property that can be used to obtain asymptotic confidence intervals for the parameters of the model, thus extending the point estimation of the parameters to interval estimation problem quite naturally.

However, if the available data are incomplete, then one needs to modify the standard procedures of parameter estimation, which assume the available data to be complete, accordingly. In the preceding section, we have discussed censoring and truncation; if the data are censored or truncated, then we call the data to be incomplete. Thus, for such data, we may want to use some special techniques for statistical inference. Under maximum likelihood estimation, the EM algorithm (Dempster, Laird and Rubin, 1977) provides such a technique for estimation based on incomplete data.

1.5.1 The EM algorithm

The Expectation Maximiation (EM) algorithm is a very powerful and useful tool for analyzing incomplete data; see McLachlan and Krishnan (2008) for an elaborate discussion on this method. The algorithm consists of two steps – the Expectation step (E-step) and the Maximization step (M-step). First of all, the complete data likelihood for the given problem is formed. Then, in the E-step, the conditional expectation of the complete data log-likelihood is obtained, given the observed incomplete data and the current value of the parameter, scalar- or vector-valued. This expected log-likelihood is essentially a function of the parameter involved and the current value of the parameter under which the expectation has been calculated. Thus, if the underlying lifetime variable is T, and the observed variable is Y, then in the E-step, our objective is to obtain

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = E_{\boldsymbol{\theta}^{(k)}}\left[\log L_c(\boldsymbol{t}; \boldsymbol{\theta})\middle|\boldsymbol{y}\right],$$

where $L_c(\boldsymbol{t}; \boldsymbol{\theta})$ is the complete data likelihood, $\boldsymbol{\theta}$ is the parameter vector of interest, and $\boldsymbol{\theta}^{(k)}$ is the value of the parameter vector at the k-th step of the algorithm (current value of the parameter).

In the M-step, this expected complete data log-likelihood $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ is maximized with respect to $\boldsymbol{\theta}$ over the parameter space $\boldsymbol{\Theta}$ to obtain the improved estimate of the paramater as

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

Depending on the form of the Q-function, the parameter estimate could be of closed-form, or may have to be determined by the use of some numerical methods.

The E- and M-steps are then iterated till convergence. This algorithm is known to have many 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.

1.5.2 Asymptotic variance-covariance matrix of the MLEs

Asymptotic variance-covariance matrix of the MLEs under the EM framework is not directly obtained. Louis (1982) developed a technique to be used under the EM framework to obtain the asymptotic variance-covariance matrix of the MLEs, and it is based on the missing information principle. According to this principle,

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

Let C and δ denote the censoring time and censoring indicator, respectively; i.e., $\delta = 0$ if a unit is censored, and it is 1 if the unit is not censored. More detailed description of all the symbols used will be given in the following section. Let $I_T(\boldsymbol{\theta})$, $I_Y(\boldsymbol{\theta})$ and $I_{C|Y}(\boldsymbol{\theta})$ denote the complete information matrix, the observed information matrix and the missing information matrix, respectively. The complete information matrix is given by

$$I_T(\boldsymbol{\theta}) = -E\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log L_c(t; \boldsymbol{\theta})\right]. \tag{1.5.1}$$

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

$$I_{C|Y}^{(i)}(\boldsymbol{\theta}) = -E\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log f_{C_i}(c_i | C_i > y_i, \boldsymbol{\theta})\right]. \tag{1.5.2}$$

The expected missing information can thus be obtained as

$$I_{C|Y}(\boldsymbol{\theta}) = \sum_{i:\delta_i=0} I_{C|Y}^{(i)}(\boldsymbol{\theta}). \tag{1.5.3}$$

Therefore, by the missing information principle, the observed information matrix is simply given by

$$I_Y(\boldsymbol{\theta}) = I_T(\boldsymbol{\theta}) - I_{C|Y}(\boldsymbol{\theta}). \tag{1.5.4}$$

The asymptotic variance-covariance matrix of the MLEs of $\boldsymbol{\theta}$, under the EM framework, can then be obtained by inverting the observed information matrix $I_Y(\hat{\boldsymbol{\theta}})$.

1.5.3 Confidence intervals

After obtaining the MLEs of the parameters and the asymptotic variances and covariances of the MLEs, one can construct the asymptotic confidence intervals for the parameters by using the asymptotic normality of the MLEs. For obtaining the asymptotic variance-covariance matrix, one can use the missing information principle as described above, or can simply use the observed information matrix calculated directly from the data.

It is also possible to construct parametric bootstrap confidence intervals for the parameters as follows. First, the MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained from the given data. Then, using $\hat{\boldsymbol{\theta}}$ as the true value of the parameter $\boldsymbol{\theta}$, a sample of size n in the same sampling framework with left truncation and right censoring is produced by simulation. This process is repeated for B Monte Carlo runs and the MLEs are determined for each of these B simulated samples. In the next step, based on these B estimates, the bootstrap bias and variance for the estimates of the parameters are obtained. Finally, the $100(1-\alpha)\%$ parametric bootstrap confidence interval for the parameter μ (a component of $\boldsymbol{\theta}$) is then obtained as

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

where b_μ and v_μ are, respectively, the bootstrap bias and variance for the parameter μ and z_α is the upper α-percentage point of the standard normal distribution.

1.6 Description of the data and the general setup

All the results in this thesis are studied and illustrated through Monte Carlo simulations, and for the simulation, the general setup of left truncated and right censored dataset used by Hong *et al.* (2009) is followed. The data can be considered to be lifetime data of power transformers in an electrical industry, over a particular interval of time. The failure of a machine is observed only if the failure is on or after 1980, as detailed record keeping on the lifetime of machines was started in that year. Complete information on the lifetime of machines installed after 1980 are available, while for those installed before 1980, the installation dates are available but no information is available on machines installed and failed before 1980. This results in the data being left truncated. Also, the lifetime of the machines are followed only until 2008, which incorporates right censoring into the data. The machines which failed on or before 2008 were completely observed, while those that did not fail until 2008 resulted in right censored observations.

Let Y denote the lifetime variable, C denote the censoring time variable, and δ_i denote the censoring indicator, i.e., δ_i is 0 if the i-th unit is censored, and 1 if it is not. Similarly, let ν_i denote the truncation indicator, i.e., ν_i is 0 if the i-th unit is truncated, and 1 if it is not. Let τ_i^L denote the left-truncation time, i.e., for a unit installed before 1980, τ_i^L is the difference of 1980 and its installation year; let S_1 and S_2 denote the index sets corresponding to the units which are not left truncated and are left truncated, respectively. Let $\boldsymbol{\theta}$ denote the model parameter to be estimated.

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

$$L(\boldsymbol{\theta}) \propto \prod_{i \in S_1} \left\{f_Y(y_i)\right\}^{\delta_i} \left\{1 - F_Y(y_i)\right\}^{1-\delta_i} \times \prod_{i \in S_2} \left\{\frac{f_Y(y_i)}{1 - F_Y(\tau_i^L)}\right\}^{\delta_i} \left\{\frac{1 - F_Y(y_i)}{1 - F_Y(\tau_i^L)}\right\}^{1-\delta_i}. \tag{1.6.1}$$

Observe that the likelihood function is constructed based on the contributions of four groups: untruncated and uncensored, untruncated and censored, truncated and uncensored, and finally truncated and censored. Also, this is the observed data likelihood which can be used for direct method of estimation. For employing the EM algorithm, we will need the complete data likelihoodas well which will be discussed in the following chapters for different population distributions.

1.6.1 An example

In Table 1.1, as an example, we present part of a dataset generated from Weibull distribution with left truncation and right censoring. The complete dataset is given in Appendix B. The lifetime, truncation time and censoring time are reported on the log-scale. Thus, the parameters μ and σ of interest here are indeed the extreme value parameters.

Table 1.1: A simulated dataset, for sample size 100, truncation percentage 40, and the true parameter value of (μ, σ) as (3.55, 0.33), wherein $*$ means not applicable. 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	1960	0	2.996	2006	1	3.829	*
7	1961	0	2.944	1992	1	3.434	*
8	1964	0	2.773	*	0	3.784	3.784
9	1963	0	2.833	1991	1	3.332	*
10	1973	0	1.946	*	0	3.555	3.555

The above dataset gives examples of all four types of observations. Unit no. 1 is installed in 1984 (after 1980), and is therefore an untruncated observation; but, it does not fail until 2008, and so becomes a right censored unit. Unit no. 2 is an untruncated unit too, as it is intalled in 1990; but, as it fails in 2001, it becomes a complete failure (or observation). However, unit no. 6 is a left truncated unit, as it is installed in 1960 (before 1980), and its truncation time is thus 2.996, the logarithm of the difference between 1980 and 1960. And as the unit fails in 2006, it is a complete observation. But unit no. 8, which is left truncated because it is installed in 1964, does not fail until 2008, and so becomes a right censored observation.

1.7 Scope of the thesis

Hong, Meeker and McCalley (2009) carried out an analysis of left truncated and right censored data which came from the electrical industry in the US. They fitted a Weibull distribution, which turned out to be the appropriate distribution for their data, using the direct approach of obtaining the MLEs. In this thesis, we develop the likelihood inference via the EM algorithm for fitting a number of important lifetime distributions to left truncated and right censored data.

The distributions we fit to left truncated and right censored data are lognormal, Weibull, gamma and generalized gamma. First, we develop the necessary steps of the EM algorithm for estimating the parameters of these distributions, and then by using the missing information principle as discussed in the preceding section, we obtain the asymptotic variance-covariance matrix of the MLEs. Then, using asymptotic normality of the MLEs, we obtain the asymptotic confidence intervals for the parameters involved. For comparison purpose, we also fit the distributions using the Newton-Raphson (NR) method, which is a direct approach for obtaining the MLEs. All the methods of inference are then studied through extensive Monte Carlo simulation studies. It is observed that the EM algorithm and the NR method perform quite closely.

Once the parameters of the underlying distribution are estimated, we can use them for many purposes. One of them may be to estimate the probability of a unit to work until a future year, given that the unit is right censored. We give the distribution specific estimate for this probability, and also provide a confidence interval for this probability by using the delta method. All the inferential methods developed here are illustrated through some numerical examples.

It may be of interest for a given data to determine which of the above mentioned distributions provides the best fit. One can use a likelihood-based method or an information-based method to determine the best fit. Here, we use information-based techniques, namely, Akaike information criterion (AIC) and Bayesian information criterion (BIC) for determining the best model for a given data. Through a Monte Carlo simulation study, we examine the proportions of selecting the models (both correct and wrong selections) by AIC and BIC when the data come from different distributions, and it is observed that these criteria work well in this case meaning that they do select the correct model with higher probability. Through the empirical study, it is observed that the AIC has a better performance than the BIC. Finally, some numerical examples are given for illustrating the model discrimination problem.

Chapter 2

Inference for lognormal distribution

2.1 Introduction

The lognormal distribution has been used extensively to model lifetime data; see, for example, Johnson, Kotz and Balakrishnan (1994) and Crow and Shimizu (1988), and the references therein. Here, we discuss the fitting of a lognormal distribution to a left truncated and right censored data via the EM algorithm, and study the proposed estimation methods through an extensive Monte Carlo simulation study. The maximization part of the EM algorithm poses a challenge in this situation, as the expected complete log-likelihood involves a complicated non-linear function of the parameters. The required maximization part is carried out through two different methods, and by means of an extensive simulation study, these two methods are compared. The asymptotic variances and covariance of the parameter estimates are derived within the EM framework by using the missing information principle (Louis, 1982), and are used to construct asymptotic confidence intervals for the parameters. The parametric bootstrap confidence intervals are also constructed.

Furthermore, the Newton-Raphson (NR) method is also used to estimate the parameters for comparative purposes. Then, the asymptotic confidence intervals for the parameters are also constructed by using the observed information matrix and all mentioned confidence intervals are evaluated in terms of coverage probabilities, through a detailed Monte Carlo simulation study. A prediction problem is also looked at, regarding the lifetime of a right censored unit. The probability that such a unit will be in operation until a future year is estimated, and an estimate of its standard error is obtained by employing the delta method, using which a confidence interval is provided for this probability. Finally, a numerical example is presented to illustrate all the methods of inference developed here.

2.2 Likelihood function

Let X denote the lifetime variable that follows a lognormal distribution whose density is given by (1.2.5), where μ and σ are the scale and shape parameters, respectively. Then, the variable $Y = \log X$ is normally distributed with density

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{ -\frac{(y-\mu)^2}{2\sigma^2} \right\}, \quad -\infty < y < \infty,$$

where μ and σ are the location and scale parameters, respectively. Here, all the work is done with the log-transformed model, as it reduces the model to a normal distribution which is convenient for the ensuing derivations and algebraic calculations. Let $\phi(\cdot)$ and $\Phi(\cdot)$ denote the probability density function and cumulative distribution function of the standard normal distribution, respectively.

We first give a description of the symbols used throughout the chapter. Let Y denote the log-transformed lifetime variable, 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 the observation is not censored, and τ_i^L denote the left truncation time, log-transformed. To be precise, for a machine installed before 1980, τ_i^L is the time between the year of installation and the truncation point of 1980, on the log-scale. 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, where S_1 is the set of machines installed after 1980 and S_2 is the set installed on or before 1980.

Then, from (1.6.1), we have the likelihood for the left truncated and right censored data as

$$L(\mu,\sigma) = \prod_{i\in S_1}\left\{\frac{1}{\sigma}\phi\left(\frac{y_i-\mu}{\sigma}\right)\right\}^{\delta_i}\left\{1-\Phi\left(\frac{y_i-\mu}{\sigma}\right)\right\}^{1-\delta_i}$$
$$\times\prod_{i\in S_2}\left\{\frac{\frac{1}{\sigma}\phi(\frac{y_i-\mu}{\sigma})}{1-\Phi(\frac{\tau_i^L-\mu}{\sigma})}\right\}^{\delta_i}\left\{\frac{1-\Phi(\frac{y_i-\mu}{\sigma})}{1-\Phi(\frac{\tau_i^L-\mu}{\sigma})}\right\}^{1-\delta_i},$$

where μ and σ are the mean and standard deviation of the log-transformed normal lifetime variable Y. The log-likelihood function (without the constant term), after simplification, and by using the truncation indicator, is given by

$$\log L(\mu,\sigma) = \sum_{i=1}^{n}\left[-\delta_i\log\sigma-\delta_i\frac{1}{2\sigma^2}(y_i-\mu)^2+(1-\delta_i)\log\left(1-\Phi(\frac{y_i-\mu}{\sigma})\right)\right]$$
$$-\sum_{i=1}^{n}(1-\nu_i)\log\left(1-\Phi(\frac{\tau_i^L-\mu}{\sigma})\right). \tag{2.2.1}$$

2.3 Inferential methods

2.3.1 The EM algorithm

The first step to employ the EM algorithm is to obtain the complete data likelihood. Let T denote the log-transformed lifetime of a machine, and $\boldsymbol{\theta}$ denote the parameter vector $(\mu,\sigma)'$. With all the other notation remaining as before, had there been no censoring, the complete data likelihood would be

$$L_c(t;\boldsymbol{\theta}) = \prod_{S_1}\left\{\frac{1}{\sigma}\phi\left(\frac{t_i-\mu}{\sigma}\right)\right\}\times\prod_{S_2}\left\{\frac{\frac{1}{\sigma}\phi(\frac{t_i-\mu}{\sigma})}{1-\Phi(\frac{\tau_i^L-\mu}{\sigma})}\right\}.$$

The log-likelihood function, using the truncation indicator ν_i, is given by

$$\log L_c(t;\boldsymbol{\theta}) = -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(t_i^2 - 2\mu t_i + \mu^2\right) - \sum_{i=1}^{n}(1-\nu_i)\log\left\{1-\Phi\left(\frac{\tau_i^L-\mu}{\sigma}\right)\right\}. \tag{2.3.1}$$

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

The E-Step: In this step, the conditional expectation of the complete data log-likelihood, given the observed data and the current value of the parameter, given by

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = E_{\boldsymbol{\theta}^{(k)}}\left[\log L_c(t;\boldsymbol{\theta})\middle|\boldsymbol{y}, \boldsymbol{\delta}\right] \tag{2.3.2}$$

is calculated. The first and second moments of t_i, given $t_i > y_i$, can be obtained by using the results of Cohen (1991) as

$$E_{\boldsymbol{\theta}^{(k)}}[t_i|t_i > y_i] = \mu^{(k)} + \sigma^{(k)} q_i^{(k)},$$

$$E_{\boldsymbol{\theta}^{(k)}}[t_i^2|t_i > y_i] = {\sigma^{(k)}}^2(1 + \xi_i^{(k)} q_i^{(k)}) + 2\sigma^{(k)}\mu^{(k)} q_i^{(k)} + {\mu^{(k)}}^2,$$

where

$$\xi_i^{(k)} = \frac{y_i - \mu^{(k)}}{\sigma^{(k)}},$$

$$q_i^{(k)} = \frac{\phi(\xi_i^{(k)})}{1-\Phi(\xi_i^{(k)})}.$$

Using these expressions for the conditional expectations, we then obtain

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \; & -n \log \sigma - \sum_{i=1}^{n} (1-\nu_i) \log \left\{ 1 - \Phi\left(\frac{\tau_i^L - \mu}{\sigma} \right) \right\} \\ & -\frac{1}{2\sigma^2} \left[\sum_{i:\delta_i=1} y_i^2 + \sum_{i:\delta_i=0} \{ {\sigma^{(k)}}^2 (1 + \xi_i^{(k)} q_i^{(k)}) + 2\sigma^{(k)} \mu^{(k)} q_i^{(k)} + {\mu^{(k)}}^2 \} \right] \\ & + \frac{\mu}{\sigma^2} \left[\sum_{i:\delta_i=1} y_i + \sum_{i:\delta_i=0} \{ \mu^{(k)} + \sigma^{(k)} q_i^{(k)} \} \right] - \frac{n\mu^2}{2\sigma^2}. \end{aligned} \tag{2.3.3}$$

The M-Step: In the maximization step, the quantity $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ in (2.3.3) is maximized with respect to $\boldsymbol{\theta}$ over the parameter space $\boldsymbol{\Theta}$ to obtain the improved estimate of the parameter 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 obtain the MLE of the parameter $\boldsymbol{\theta}$. However, it may not always be easy to maximize the function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ with respect to $\boldsymbol{\theta}$. A close look at the function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ in (2.3.3) reveals that there is no straightforward method of maximizing it with respect to $\boldsymbol{\theta}$, and that the MLEs of μ and σ do not have any explicit expressions and have to be obtained numerically.

In deriving the MLEs of μ and σ, we employ here two different methods. One of them gives approximate values of the MLEs while the other gives numerical values of the MLEs provided some conditions are satisfied. However, both methods involve some approximations which seems to be inevitable in this case.

Method 1. The first-order derivatives of the function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ with respect to μ and σ involve the hazard function of the normal distribution along with polynomial components in μ and σ, thus making the solution for μ and σ to be quite involved. The equations obtained by equating the first-order derivatives of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ with respect to μ and σ, respectively, to zero are as follows:

$$n\mu - C_1 + \sigma \sum_{i=1}^{n} (1-\nu_i) h_i(\mu, \sigma) = 0, \tag{2.3.4}$$

$$n\sigma^2 + \sigma \sum_{i=1}^{n} (1-\nu_i)(\tau_i^L - \mu) h_i(\mu, \sigma) - C_2 + 2\mu C_1 - n\mu^2 = 0, \tag{2.3.5}$$

where

$$h_i(\mu, \sigma) = \frac{\phi(\frac{\tau_i^L - \mu}{\sigma})}{1 - \Phi(\frac{\tau_i^L - \mu}{\sigma})},$$

$$C_1 = \sum_{i:\delta_i = 1} y_i + \sum_{i:\delta_i = 0} \{\mu^{(k)} + \sigma^{(k)} q_i^{(k)}\},$$

$$C_2 = \sum_{i:\delta_i = 1} y_i^2 + \sum_{i:\delta_i = 0} \{{\sigma^{(k)}}^2 (1 + \xi_i^{(k)} q_i^{(k)}) + 2\sigma^{(k)} \mu^{(k)} q_i^{(k)} + {\mu^{(k)}}^2\}.$$

It is evident that explicit expressions cannot be obtained by directly solving Eqs. (2.3.4) and (2.3.5). However, an approximation can be made by considering the Taylor series expansion of the hazard function $h_i(\mu, \sigma)$. We consider a Taylor expansion of $h_i(\mu, \sigma)$ with respect to μ for fixed σ to obtain an approximate value of the MLE of μ. In the subsequent step, we then consider a Taylor expansion of $h_i(\mu, \sigma)$ with respect to σ for fixed μ to obtain an approximate value for the MLE of σ. This process is then continued iteratively until convergence to obtain the MLEs of μ and σ. Note that these are only approximations to the MLEs of μ and σ.

Thus, carrying out the simplification, we obtain the $(k+1)$-th step estimates of μ and σ as

$$\mu^{(k+1)} \approx \frac{\frac{1}{n}\left[C_1 - \sigma^{(k)} \sum_{i=1}^{n}(1-\nu_i)\left\{h_i(\mu^{(k)}, \sigma^{(k)}) - \mu^{(k)} h_{\mu_i}(\mu^{(k)}, \sigma^{(k)})\right\}\right]}{1 + \frac{\sigma^{(k)}}{n}\sum_{i=1}^{n}(1-\nu_i) h_{\mu_i}(\mu^{(k)}, \sigma^{(k)})} \tag{2.3.6}$$

and

$$\sigma^{(k+1)} \approx \frac{-B + \sqrt{B^2 - 4AK}}{2A}, \tag{2.3.7}$$

where

$$A = n + \sum_{i=1}^{n}(1-\nu_i)(\tau_i^L - \mu^{(k)}) h_{\sigma_i}(\mu^{(k)}, \sigma^{(k)}),$$

$$B = \sum_{i=1}^{n}(1-\nu_i)(\tau_i^L - \mu^{(k)})\left\{h_i(\mu^{(k)}, \sigma^{(k)}) - \sigma^{(k)} h_{\sigma_i}(\mu^{(k)}, \sigma^{(k)})\right\},$$

$$K = -C_2 + 2\mu^{(k)} C_1 - n\{\mu^{(k)}\}^2,$$

h_μ and h_σ being the partial derivatives of the hazard function $h(\mu, \sigma)$ with respect to μ and σ, respectively. It should be noted that while solving the quadratic equation for σ, we obtain two roots of which the larger one is as presented in Eq. (2.3.7). In our simulations, we observed that in most of the cases, this root is positive while the smaller one was negative. However, in some situations, both roots were positive; but, we chose the larger root in (2.3.7) simply because it is used in an iterative process until the occurrence of convergence to the desired tolerance level.

This process is then continued iteratively until convergence to obtain approximate values of the MLEs of μ and σ. The computation is carried out for simulated datasets using R package. We observe that in these simulated datasets, the numerical MLEs converge to the true parameter values quite accurately, and details of the maximization result show that other properties of the EM algorithm are also satisfied. For example, the convergence does not depend on the choice of initial values, and that the numerical value of the log-likelihood increases at each step of the iterative process. Thus, these estimates can be considered as approximate values of the MLEs of μ and σ.

Method 2. An alternative method used here is to maximize $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ by the EM gradient algorithm (Lange, 1995). In this method, the function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ is maximized in the M step by a one-step Newton-Raphson's method. This algorithm is closely related to the original EM algorithm, and qualifies as a special case of the generalized EM algorithm (Dempster, Laird and Rubin, 1977). The properties of this algorithm are almost the same as those of the EM algorithm. However, unlike the EM algorithm, the EM gradient algorithm depends on the choice of the initial values of the parameters.

To use a one-step maximization of the function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ in the M-step by Newton-Raphson method, we need the following expressions:

$$\frac{\partial Q}{\partial \mu} = -\sum_{i=1}^{n} \frac{(1-\nu_i)}{\sigma} \frac{\phi(\frac{\tau_i^L - \mu}{\sigma})}{1 - \Phi(\frac{\tau_i^L - \mu}{\sigma})} + \frac{C_1}{\sigma^2} - \frac{n\mu}{\sigma^2},$$

$$\frac{\partial Q}{\partial \sigma} = -\frac{n}{\sigma} - \sum_{i=1}^{n} (1-\nu_i) \frac{\tau_i^L - \mu}{\sigma^2} \frac{\phi(\frac{\tau_i^L - \mu}{\sigma})}{1 - \Phi(\frac{\tau_i^L - \mu}{\sigma})} + \frac{C_2 - 2\mu C_1 + n\mu^2}{\sigma^3},$$

$$-\frac{\partial^2 Q}{\partial \mu^2} = \sum_{i=1}^{n} \frac{1-\nu_i}{\sigma^2} \frac{\left\{1-\Phi(\frac{\tau_i^L-\mu}{\sigma})\right\}\left(\frac{\tau_i^L-\mu}{\sigma}\right)\phi(\frac{\tau_i^L-\mu}{\sigma}) - \left\{\phi(\frac{\tau_i^L-\mu}{\sigma})\right\}^2}{\left\{1-\Phi(\frac{\tau_i^L-\mu}{\sigma})\right\}^2} + \frac{n}{\sigma^2},$$

$$\begin{aligned} -\tfrac{\partial^2 Q}{\partial \sigma^2} = & -\frac{n}{\sigma^2} + \sum_{i=1}^{n}(1-\nu_i) \frac{\left\{1-\Phi(\frac{\tau_i^L-\mu}{\sigma})\right\}\frac{(\tau_i^L-\mu)^3}{\sigma^5}\phi(\frac{\tau_i^L-\mu}{\sigma}) - \left\{\phi(\frac{\tau_i^L-\mu}{\sigma})\right\}^2\frac{(\tau_i^L-\mu)^2}{\sigma^4}}{\left\{1-\Phi(\frac{\tau_i^L-\mu}{\sigma})\right\}^2} \\ & -2\sum_{i=1}^{n}(1-\nu_i)\frac{\phi(\frac{\tau_i^L-\mu}{\sigma})}{1-\Phi(\frac{\tau_i^L-\mu}{\sigma})}\frac{\tau_i^L-\mu}{\sigma^3} + \frac{3C_2 - 6\mu C_1 + 3n\mu^2}{\sigma^4}, \end{aligned}$$

$$\begin{aligned} -\tfrac{\partial^2 Q}{\partial \mu \partial \sigma} = & \sum_{i=1}^{n} \frac{1-\nu_i}{\sigma} \frac{\left\{1-\Phi(\frac{\tau_i^L-\mu}{\sigma})\right\}\frac{(\tau_i^L-\mu)^2}{\sigma^3}\phi(\frac{\tau_i^L-\mu}{\sigma}) - \left\{\phi(\frac{\tau_i^L-\mu}{\sigma})\right\}^2\left(\frac{\tau_i^L-\mu}{\sigma^2}\right)}{\left\{1-\Phi(\frac{\tau_i^L-\mu}{\sigma})\right\}^2} \\ & -\sum_{i=1}^{n}\frac{1-\nu_i}{\sigma^2}\frac{\phi(\frac{\tau_i^L-\mu}{\sigma})}{1-\Phi(\frac{\tau_i^L-\mu}{\sigma})} + \frac{2(C_1 - n\mu)}{\sigma^3}. \end{aligned}$$

Using these expressions, the Newton-Raphson method of maximization can be carried out. The computation is performed on simulated datasets using the R package, and it was observed in this process that the numerical MLEs converged to the true parameter values quite accurately.

2.3.2 Asymptotic variances and covariance of the MLEs

As described in Section 1.5.2, we can use the missing information principle by Louis (1982) to obtain the asymptotic variance-covariance matrix of the MLEs. The appropriate elements of the complete information matrix $I_T(\boldsymbol{\theta})$, calculating the necessary expectations of the lifetime variable over the two index sets S_1 and S_2, are obtained as follows:

$$-E\left[\tfrac{\partial^2}{\partial \mu^2}\log L_c(t;\boldsymbol{\theta})\right] = \frac{n}{\sigma^2} + \sum_{i=1}^{n}\frac{1-\nu_i}{\sigma^2}\frac{\left\{1-\Phi(\beta_i)\right\}\beta_i\phi(\beta_i) - \left\{\phi(\beta_i)\right\}^2}{\left\{1-\Phi(\beta_i)\right\}^2},$$

$$-E\left[\frac{\partial^2}{\partial\sigma^2}\log L_c(t;\boldsymbol{\theta})\right] = \frac{2n}{\sigma^2}+\sum_{i=1}^{n}\frac{1-\nu_i}{\sigma^2}\big\{\beta_i h_i(1+\beta_i^2-\beta_i h_i)\big\},$$

$$-E\left[\frac{\partial^2}{\partial\mu\partial\sigma}\log L_c(t;\boldsymbol{\theta})\right] = \sum_{i=1}^{n}\frac{1-\nu_i}{\sigma^2}h_i(1+\beta_i^2-\beta_i h_i),$$

where $\beta_i=\frac{\tau_i^L-\mu}{\sigma}$ and $h_i=\frac{\phi(\beta_i)}{1-\Phi(\beta_i)}$.

Now, the conditional distribution required for the calculation of the missing information matrix is given by [see Ng, Chan and Balakrishnan (2002)]

$$f_{C_i|Y_i}(c_i|C_i>y_i,\mu,\sigma)=\frac{\frac{1}{\sigma}\phi(\frac{c_i-\mu}{\sigma})}{1-\Phi(\frac{y_i-\mu}{\sigma})},\quad c_i\geq y_i.$$

The logarithm of the above truncated normal pdf is

$$\begin{aligned}\log f_{C_i|Y_i}(c_i|c_i>y_i,\mu,\sigma) = {} & \text{const}-\log\sigma-\frac{(c_i-\mu)^2}{2\sigma^2}\\ & -\log\left\{1-\Phi\left(\frac{y_i-\mu}{\sigma}\right)\right\}.\end{aligned}\tag{2.3.8}$$

The derivatives of (2.3.8) with respect to μ and σ are given by

$$\frac{\partial}{\partial\mu}\log f_{C_i|Y_i}=\frac{1}{\sigma}\big[\frac{c_i-\mu}{\sigma}-q_i\big],$$

$$\frac{\partial}{\partial\sigma}\log f_{C_i|Y_i}=\frac{1}{\sigma}\big[\big(\frac{c_i-\mu}{\sigma}\big)^2-(1+\xi_i q_i)\big],$$

where $\xi_i=\frac{y_i-\mu}{\sigma}$ and $q_i=\frac{\phi(\xi_i)}{1-\Phi(\xi_i)}$. It can be shown that [see Cohen (1991) and Ng, Chan and Balakrishnan (2002)]

$$E\big[(C_i-\mu)|C_i>y_i,\mu,\sigma\big]=\sigma q_i,$$

$$E\big[(C_i-\mu)^2|C_i>y_i,\mu,\sigma\big]=\sigma^2[1+\xi_i q_i],$$

$$E\left[(C_i-\mu)^3|C_i>y_i,\mu,\sigma\right]=\sigma^3[2+\xi_i^2]q_i,$$

$$E\left[(C_i-\mu)^4|C_i>y_i,\mu,\sigma\right]=\sigma^4[3(1+\xi_i q_i)+\xi_i^3 q_i].$$

Using these expectations, the elements of the Fisher information matrix corresponding to the i-th observation which is censored can be obtained as

$$E\left[\left(\frac{\partial}{\partial\mu}\log f_{C_i|Y_i}\right)^2\right]=\frac{1}{\sigma^2}[1+\xi_i q_i-q_i^2], \tag{2.3.9}$$

$$E\left[\left(\frac{\partial}{\partial\sigma}\log f_{C_i|Y_i}\right)^2\right]=\frac{1}{\sigma^2}[2+\xi_i q_i(1-\xi_i q_i+\xi_i^2)], \tag{2.3.10}$$

$$E\left[\left(\frac{\partial}{\partial\mu}\log f_{C_i|Y_i}\right)\left(\frac{\partial}{\partial\sigma}\log f_{C_i|Y_i}\right)\right]=\frac{1}{\sigma^2}[q_i+\xi_i q_i(\xi_i-q_i)]. \tag{2.3.11}$$

Eqs. (2.3.9) - (2.3.11) give the elements of the matrix (1.5.2), and consequently by using (1.5.3) and (1.5.4), the matrix $I_Y(\boldsymbol{\theta})$ can be obtained. Finally, the asymptotic variance-covariance matrix of the MLEs can be obtained by inverting the matrix $I_Y(\hat{\boldsymbol{\theta}})$.

2.3.3 Newton-Raphson method

For comparative purpose, we also obtain the MLEs of the parameters by the Newton-Raphson (NR) method, which is a common and direct approach for obtaining the MLEs, involving the second derivatives of the likelihood function with respect to the parameters. It is observed that the NR method fails to converge under this setup for some cases. The NR method has been employed here through a default function of the R software, called “maxNR”. It is observed in Section 6 that the EM and the NR methods give quite close results in most cases.

In the NR method, the asymptotic variance-covariance matrix of the MLEs can be estimated by inverting the observed information matrix, and the corresponding asymptotic confidence intervals for the parameters can then be constructed. Here we have constructed such confidence intervals, and have compared them with the other confidence intervals in terms of coverage probabilities through a Monte Carlo simulation study.

2.3.4 Confidence intervals

As described in Section 1.5.3, we can derive asymptotic confidence intervals for the parameters using the asymptotic normality of the MLEs. As the two methods of estimation (Methods 1 and 2) lead to different estimates (however close they are), we obtain confidence intervals corresponding to both these methods. Also, it is always possible to use the observed information matrix, directly obtained by calculating the second derivatives of Eq. (2.2.1), to estimate the asymptotic variances of the MLEs. We use the observed information matrix and the parametric bootstrap technique as well to obtain confidence intervals based on the MLEs.

It would also be possible to construct nonparametric bootstrap confidence intervals, but with both censoring and truncation present in the data, the nonparametric bootstrap method will not be as efficient as a parametric bootstrap method unless the sample size is quite large. A similar observation has been made by Balakrishnan *et al.* (2007) in the context of analysis of censored data from step-stress experiments. In the next section wherein we explain all the relevant simulation details, we present the empirical results pertaining to these confidence intervals.

2.3.5 An application to prediction

With the estimated parameters μ and σ, we can obtain the probability of a censored unit working till a future year, given that it has worked till 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 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

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

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

$$\hat{\pi} = \frac{1 - \Phi(\frac{a-\hat{\mu}}{\hat{\sigma}})}{1 - \Phi(\frac{b-\hat{\mu}}{\hat{\sigma}})} = g(\hat{\boldsymbol{\theta}}), \tag{2.3.12}$$

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 $\hat{\boldsymbol{\theta}}$, we can also estimate the variance of the above estimate $\hat{\pi}$. A straightforward application of the delta-method will yield

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

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

$$\widehat{\mathrm{Var}(\hat{\pi})} = \left((\frac{\partial g}{\partial \mu})^2 \mathrm{Var}(\hat{\mu}) + 2(\frac{\partial g}{\partial \mu})(\frac{\partial g}{\partial \sigma}) \mathrm{Cov}(\hat{\mu}, \hat{\sigma}) + (\frac{\partial g}{\partial \sigma})^2 \mathrm{Var}(\hat{\sigma}) \right) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{2.3.13}$$

Examples of this technique are given in Section 2.6.

2.4 Simulation

2.4.1 Setup

Simulations were performed by using the R software. First, a certain percentage for the truncation was fixed. With this fixed percentage of truncation, the installation years were sampled by unequal probability with-replacement sampling from an arbitrary set of years. For our simulation study, we fixed 1980 and 2008 as the years of truncation and censoring, respectively. The entire set of installation years was divided into two parts: (1960 - 1979) and (1980 - 1995). Unequal probabilities were allocated over the different years as follows: 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; 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. The lifetimes of the machines, in years, were simulated from a lognormal distribution with a specific μ and σ. The lifetimes of the machines were then added to the installation years, to obtain the years of failure. Clearly, the year of failure of a machine determined whether the lifetime of the machine was censored or not. For the right censored units, the lifetime was taken as the minimum of the lifetime and the censoring time. Because of the left truncation of the data, if the year of failure is before 1980 for any particular unit, there is no information available on the lifetime of that unit. So, if the year of failure was obtained to be a year before 1980 for any unit, that observation was discarded and a new installation year and lifetime were then simulated. This setup produced, along with the desired level of truncation, sufficiently many censored observations. The data were then log-transformed and the analyses were performed on the logarithmic scale, i.e., based on the normal distribution.

The total sample sizes used in the simulation study are 75, 100 and 200. We have observed that Method 2 sometimes shows a problem with convergence for smaller sample sizes (such as, 50). Empirically, it has been observed that for samples of size at least 75, the method works fine, and never shows a problem with convergence.

Method 1 and the NR method, however, never show any problem with convergence, even for smaller sample sizes. The truncation percentage was fixed at 30 and 60, respectively. Fixing the truncation percentage at two levels allows us to observe the behaviour of the estimation methods for the model under heavy and moderate truncation amounts. Two different choices of the parameter vector $\boldsymbol{\theta} = (\mu, \sigma)$ was used: (3.5, 0.5) and (3.0, 0.2). All the simulation results were based on 1000 Monte Carlo runs. The method of moment estimates, i.e., the sample mean and sample standard deviation, were used as initial values. The tolerance limit for convergence was fixed as 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.

2.4.2 Results and discussion

It is observed in the simulation study that the two methods under the EM algorithm and the NR method yield very close results. Table 2.1 gives the bias and root mean square errors of the parameter estimates by different methods, obtained from the Monte Carlo simulations.

It can be noticed from Table 2.1 that the three methods give quite close results. The confidence intervals for the parameters are constructed, using the estimated asymptotic variances of the MLEs obtained by the missing information principle. The observed information matrix, by the missing information principle, is derived as explained in Section 2.3.2. For the purpose of comparison, the confidence intervals

Table 2.1: Bias (B) and root mean square error (RMSE) for three different techniques used for parameter estimation.

$n = 75$						
(μ, σ)	Trunc.	Method	B($\hat{\mu}$)	B($\hat{\sigma}$)	RMSE($\hat{\mu}$)	RMSE($\hat{\sigma}$)
	30%	1	-0.012	-0.011	0.084	0.072
(3.5, 0.5)		2	-0.012	-0.011	0.084	0.072
		3	-0.010	-0.009	0.084	0.072
	60%	1	-0.012	-0.006	0.075	0.069
		2	-0.013	-0.006	0.076	0.069
		3	-0.012	-0.005	0.076	0.069
	30%	1	-0.006	-0.005	0.027	0.020
(3.0, 0.2)		2	-0.006	-0.005	0.027	0.020
		3	-0.006	-0.005	0.027	0.020
	60%	1	-0.009	-0.003	0.030	0.017
		2	-0.009	-0.002	0.030	0.017
		3	-0.009	-0.002	0.030	0.017

$n = 100$						
(μ, σ)	Trunc.	Method	B($\hat{\mu}$)	B($\hat{\sigma}$)	RMSE($\hat{\mu}$)	RMSE($\hat{\sigma}$)
	30%	1	-0.014	-0.010	0.073	0.061
(3.5, 0.5)		2	-0.014	-0.010	0.073	0.061
		3	-0.012	-0.009	0.074	0.061
	60%	1	-0.009	-0.004	0.065	0.059
		2	-0.009	-0.004	0.065	0.059
		3	-0.008	-0.003	0.065	0.059
	30%	1	-0.006	-0.003	0.025	0.014
(3.0, 0.2)		2	-0.006	-0.004	0.025	0.017
		3	-0.006	-0.004	0.025	0.017
	60%	1	-0.009	-0.002	0.027	0.017
		2	-0.009	-0.002	0.027	0.017
		3	-0.009	-0.002	0.027	0.017

$n = 200$						
(μ, σ)	Trunc.	Method	B($\hat{\mu}$)	B($\hat{\sigma}$)	RMSE($\hat{\mu}$)	RMSE($\hat{\sigma}$)
	30%	1	-0.020	-0.011	0.056	0.044
(3.5, 0.5)		2	-0.020	-0.012	0.056	0.044
		3	-0.018	-0.010	0.056	0.044
	60%	1	-0.013	-0.004	0.048	0.041
		2	-0.013	-0.005	0.049	0.041
		3	-0.012	-0.003	0.048	0.041
	30%	1	-0.007	-0.003	0.017	0.010
(3.0, 0.2)		2	-0.007	-0.003	0.017	0.010
		3	-0.007	-0.003	0.017	0.010
	60%	1	-0.009	-0.001	0.020	0.010
		2	-0.009	-0.001	0.020	0.010
		3	-0.009	-0.001	0.020	0.010

Note: Methods 1 and 2 stand for the two EM algorithm techniques, while Method 3 stands for the NR method.

based on the observed information matrix and the parametric bootstrap technique are also constructed. Due to the closeness of the estimates obtained by these two methods, for obtaining parametric bootstrap confidence intervals, we employed Method 1 which is the relatively faster process between the two due to its linear nature of the algorithm.

Table 2.2 gives the coverage probabilites of the confidence intervals for μ corresponding to Methods 1 and 2, observed information matrix and parametric bootstrap, for different percentages of truncation and different nominal confidence levels.

Table 2.2: Coverage probabilities for different confidence intervals for μ.

n=75						
μ	Truncation	Nominal CL	Coverage Probability			
			Method 1	Method 2	Observed	Bootstrap
3.5	30%	90%	0.899	0.900	0.901	0.904
		95%	0.939	0.939	0.940	0.942
	60%	90%	0.897	0.896	0.897	0.903
		95%	0.941	0.942	0.941	0.947
3.0	30%	90%	0.889	0.888	0.889	0.902
		95%	0.935	0.935	0.935	0.943
	60%	90%	0.875	0.874	0.874	0.897
		95%	0.935	0.934	0.934	0.953

n=100						
μ	Truncation	Nominal CL	Coverage Probability			
			Method 1	Method 2	Observed	Bootstrap
3.5	30%	90%	0.887	0.886	0.891	0.880
		95%	0.941	0.940	0.946	0.944
	60%	90%	0.901	0.899	0.901	0.895
		95%	0.946	0.945	0.946	0.943
3.0	30%	90%	0.854	0.853	0.855	0.884
		95%	0.925	0.925	0.925	0.929
	60%	90%	0.868	0.863	0.863	0.894
		95%	0.935	0.934	0.934	0.948

n=200						
μ	Truncation	Nominal CL	Coverage Probability			
			Method 1	Method 2	Observed	Bootstrap
3.5	30%	90%	0.848	0.848	0.851	0.871
		95%	0.905	0.904	0.910	0.925
	60%	90%	0.863	0.862	0.864	0.876
		95%	0.923	0.923	0.923	0.932
3.0	30%	90%	0.875	0.875	0.875	0.886
		95%	0.926	0.926	0.926	0.934
	60%	90%	0.849	0.846	0.846	0.907
		95%	0.909	0.906	0.906	0.955

Analogous to Table 2.2, Table 2.3 gives the coverage probabilities for different confidence intervals for σ.

Table 2.3: Coverage probabilities for different confidence intervals for σ.

n=75						
σ	Truncation	Nominal CL	Coverage Probability			
			Method 1	Method 2	Observed	Bootstrap
0.5	30%	90%	0.872	0.871	0.879	0.879
		95%	0.926	0.925	0.929	0.937
	60%	90%	0.881	0.881	0.881	0.877
		95%	0.931	0.930	0.931	0.928
0.2	30%	90%	0.860	0.859	0.859	0.860
		95%	0.907	0.903	0.905	0.918
	60%	90%	0.894	0.893	0.893	0.893
		95%	0.943	0.942	0.943	0.937

n=100						
σ	Truncation	Nominal CL	Coverage Probability			
			Method 1	Method 2	Observed	Bootstrap
0.5	30%	90%	0.876	0.875	0.881	0.882
		95%	0.925	0.923	0.927	0.926
	60%	90%	0.899	0.898	0.900	0.881
		95%	0.939	0.939	0.940	0.938
0.2	30%	90%	0.902	0.897	0.901	0.901
		95%	0.956	0.951	0.951	0.955
	60%	90%	0.873	0.871	0.872	0.873
		95%	0.940	0.939	0.939	0.931

n=200						
σ	Truncation	Nominal CL	Coverage Probability			
			Method 1	Method 2	Observed	Bootstrap
0.5	30%	90%	0.875	0.874	0.881	0.879
		95%	0.921	0.920	0.926	0.937
	60%	90%	0.886	0.886	0.888	0.878
		95%	0.943	0.942	0.946	0.939
0.2	30%	90%	0.866	0.863	0.864	0.869
		95%	0.925	0.918	0.923	0.931
	60%	90%	0.888	0.888	0.888	0.878
		95%	0.940	0.940	0.940	0.929

Table 2.4: Bias (B) and root mean square error (RMSE) for Method 1 and NR method (Method 3) for sample size 50.

$n = 50$						
(μ, σ)	Trunc.	Method	B($\hat{\mu}$)	B($\hat{\sigma}$)	RMSE($\hat{\mu}$)	RMSE($\hat{\sigma}$)
(3.5, 0.5)	30%	1	-0.010	-0.016	0.101	0.088
		3	-0.009	-0.014	0.102	0.088
	60%	1	-0.006	-0.010	0.090	0.087
		3	-0.006	-0.009	0.091	0.087
(3.0, 0.2)	30%	1	-0.005	-0.005	0.033	0.025
		3	-0.005	-0.006	0.033	0.025
	60%	1	-0.008	-0.004	0.036	0.022
		3	-0.008	-0.004	0.036	0.022

Clearly, from Tables 2.2 and 2.3, it can be observed that the coverage probabilities corresponding to Methods 1 and 2 (using the missing information principle) and the observed information matrix are very close, and are close to the nominal level in almost all the cases. For larger sample sizes (such as 200), the coverage probabilities stabilize at a level slightly lower than the nomina level. Same is observed for the parametric bootstrap confidence intervals. In fact, in some cases, the parametric bootstrap confidence intervals provide higher coverage probabilities compared to other confidence intervals. However, computation of parametric bootstrap confidence intervals takes much larger time as well as CPU usage compared to others.

As mentioned before, Method 1 does not show any problem with convergence even for smaller sample sizes. The NR method also works well for smaller samples, and the results obtained by them are quite close. Tables 2.4 and 2.5 present the Bias, RMSE and the coverage probabilities for the parameters corresponding to these two

Table 2.5: Coverage probabilities for different confidence intervals for μ for sample size 50.

n=50					
μ	Truncation	Nominal CL	Coverage Probability		
			Method 1	Observed	Bootstrap
3.5	30%	90%	0.889	0.891	0.902
		95%	0.945	0.946	0.950
	60%	90%	0.890	0.889	0.896
		95%	0.943	0.944	0.949
3.0	30%	90%	0.871	0.871	0.885
		95%	0.940	0.940	0.928
	60%	90%	0.868	0.867	0.873
		95%	0.926	0.924	0.937
σ	Truncation	Nominal CL	Coverage Probability		
			Method 1	Observed	Bootstrap
0.5	30%	90%	0.867	0.873	0.886
		95%	0.914	0.914	0.918
	60%	90%	0.859	0.860	0.857
		95%	0.908	0.909	0.913
0.2	30%	90%	0.882	0.882	0.877
		95%	0.926	0.926	0.926
	60%	90%	0.880	0.880	0.879
		95%	0.928	0.929	0.934

methods, respectively, when the sample size is 50. As can be seen from these two tables, even for sample size 50, the results obtained by these methods are quite close.

2.5 Illustrative example

2.5.1 Point and interval estimation

In this section, an example is given to illustrate the confidence intervals for the parameters developed in the preceding sections. The numerical example is based on a simulated data of size 100. The truncation percentage used in the data was 60, and the true value of the parameter was taken as $(\mu, \sigma) = (3.5,\ 0.5)$.

For this example, the initial value of the parameters, obtained by the method of moments, are (3.29, 0.38). The tolerance level used here is 0.000001. Table 2.6 gives the successive steps of iteration of the EM algorithms and the NR method, and Table 2.7 gives the asymptotic confidence intervals for μ and σ, corresponding to Methods 1 and 2 (using the missing information principle), observed information matrix and the parametric bootstrap technique, for different nominal confidence levels. The data for this example are presented in Table A.1 in Appendix I.

It can be noticed from Table 2.6 that the three methods converge to the same final estimate (3.51, 0.51), correct upto two decimal places. Here, it should 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 fewer number of steps (8 for Method 1, 12 for Mehod 2 and 5 for the NR method), with the same final estimate correct upto two decimal places. It can be noticed from Table 2.7 that the confidence intervals corresponding to Methods 1 and 2 and the NR method are exactly the same, and this explains

Table 2.6: Successive steps of iteration of the EM algorithms and the NR method.

	EM Method 1		EM Method 2		NR Method	
Step	$(\hat{\mu}, \hat{\sigma})$	Tolerance	$(\hat{\mu}, \hat{\sigma})$	Tolerance	$(\hat{\mu}, \hat{\sigma})$	Tolerance
1	(3.441, 0.466)	0.1781	(3.435, 0.383)	0.1498	(3.468, 0.404)	0.1844
2	(3.488, 0.486)	0.0510	(3.458, 0.421)	0.0442	(3.498, 0.468)	0.0705
3	(3.503, 0.496)	0.0185	(3.475, 0.450)	0.0337	(3.511, 0.503)	0.0378
4	(3.509, 0.503)	0.0084	(3.487, 0.471)	0.0241	(3.513, 0.511)	0.0084
5	(3.511, 0.506)	0.0044	(3.496, 0.485)	0.0166	(3.513, 0.512)	0.0003
6	(3.512, 0.509)	0.0024	(3.502, 0.495)	0.0111	(3.513, 0.512)	5.4e-07
7	(3.512, 0.510)	0.0014	(3.506, 0.501)	0.0073		
8	(3.513, 0.511)	0.0008	(3.508, 0.505)	0.0047		
9	(3.513, 0.511)	0.0005	(3.510, 0.507)	0.0030		
10	(3.513, 0.511)	0.0003	(3.511, 0.509)	0.0019		
11	(3.513, 0.511)	0.0002	(3.512, 0.510)	0.0012		
12	(3.513, 0.511)	8.9e-05	(3.512, 0.510)	0.0008		
13	(3.513, 0.512)	5.2e-05	(3.513, 0.511)	0.0005		
14	(3.513, 0.512)	3.0e-05	(3.513, 0.511)	0.0003		
15	(3.513, 0.512)	1.7e-05	(3.513, 0.511)	0.0002		
16	(3.513, 0.512)	1.0e-05	(3.513, 0.511)	0.0001		
17	(3.513, 0.512)	5.8e-06	(3.513, 0.511)	8.1e-05		
18	(3.513, 0.512)	3.4e-06	(3.513, 0.512)	5.1e-05		
19	(3.513, 0.512)	2.0e-06	(3.513, 0.512)	3.2e-05		
20	(3.513, 0.512)	1.1e-06	(3.513, 0.512)	2.1e-05		
21	(3.513, 0.512)	6.5e-07	(3.513, 0.512)	1.3e-05		
22			(3.513, 0.512)	8.3e-06		
23			(3.513, 0.512)	5.3e-06		
24			(3.513, 0.512)	3.3e-06		
25			(3.513, 0.512)	2.1e-06		
26			(3.513, 0.512)	1.3e-06		
27			(3.513, 0.512)	8.5e-07		

the close coverage probabilities seen in Tables 2.2 and 2.3 for these three methods. The confidence intervals corresponding to the parametric bootstrap are slightly wider

Table 2.7: Different confidence intervals for μ and σ.

Parameter	Nominal CL	Method 1	Method 2	Observed	Bootstrap
$\mu = 3.5$	90%	(3.404, 3.622)	(3.404, 3.622)	(3.404, 3.622)	(3.413, 3.642)
	95%	(3.383, 3.643)	(3.383, 3.643)	(3.383, 3.643)	(3.391, 3.664)
$\sigma = 0.5$	90%	(0.412, 0.612)	(0.412, 0.612)	(0.412, 0.612)	(0.417, 0.623)
	95%	(0.392, 0.631)	(0.392, 0.631)	(0.392, 0.631)	(0.397, 0.643)

than these three.

By employing the missing information principle, the complete information matrix for this data is obtained to be $\begin{pmatrix} 326.4489 & 107.6620 \\ & 628.0912 \end{pmatrix}$, while the observed information matrix to be $\begin{pmatrix} 229.4611 & -30.7460 \\ & 274.4936 \end{pmatrix}$. Thus, the ratio of the determinants of the observed information matrix to the complete information matrix is 0.3207, from which we could provide 67.93% as the proportion of information lost due to censoring. Alternatively, we can compute the corresponding variance-covariance matrices as $\begin{pmatrix} 0.0032 & -0.0006 \\ & 0.0017 \end{pmatrix}$ and $\begin{pmatrix} 0.0044 & 0.0005 \\ & 0.0037 \end{pmatrix}$ from which we can compute the trace-efficiency of the estimates based on censored data to complete data to be 60.49% and the determinant-efficiency to be 32.07%.

2.5.2 Prediction

Refer to the 66-th unit in Table A.1. For this unit, Y_{ins} is 1961, i.e., the unit is left truncated; also, it is right censored, with censoring year being 2008. The probability that this unit will be working till 2016 is estimated, by using Eq. (2.3.12) and the estimated parameters μ and σ, to be 0.655. The standard error of this probability estimate, obtained by using Eq. (2.3.13) and the estimated variance-covariance matrix of the MLEs as $\left(\begin{smallmatrix} 0.0044 & 0.0005 \\ & 0.0037 \end{smallmatrix}\right)$, is given by 0.054. In fact, an approximate 95% confidence intervals for this probability is (0.549, 0.761). Similarly, for the 35-th unit in Table A.1 for which the installation year is 1985 (i.e., not left truncated and also right censored), the probability that the unit will be working till 2016 is estimated to be 0.729, with a standard error of 0.035. An approximate 95% confidence interval for this probability is then (0.660, 0.798). It may be of interest to note here that the second unit (installed in 1985) has a higher probability to work till 2016 than the first unit (installed in 1961), as one would expect.

Chapter 3

Inference for Weibull distributions

3.1 Introduction

Weibull distribution is one of most widely used distributions for modelling lifetime data. Recently, Hong, Meeker and McCalley (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 discuss the fitting of a two-parameter Weibull distribution to a left truncated and right censored data via the EM algorithm. In this chapter, we describe in detail the steps of the EM algorithm for fitting the Weibull distribution to left truncated and right censored data. For comparative purposes, we also fit the Weibull distribution by the

Newton-Raphson (NR) method; the two methods give close results under this setup. As in the previous chapter, the asymptotic variance-covariance matrix of the MLEs are obtained by using the missing information principle (Louis, 1982), and asymptotic confidence intervals for the parameters corresponding to the EM algorithm are constructed. The asymptotic confidence confidence intervals corresponding to the NR method are also constructed by using the observed information matrix. Through a detailed Monte Carlo simulation study, these inferential methods are then studied and compared. A prediction problem regarding the future lifetime of a unit is also addressed. Finally, a numerical example is presented.

3.2 Likelihood function

Let X be the original lifetime variable, which follows a Weibull distribution with scale parameter α and shape parameter η, whose pdf is given by (1.2.3) [see Johnson et al. (1994)]. Then, the log-transformed variable $Y = \log X$ follows an extreme value distribution with density given by (1.2.9), where $\mu = \log \alpha$ and $\sigma = 1/\eta$ are the location and scale parameters, respectively. 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 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. 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 then obtained from (1.6.1) to be

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

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

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

3.3 Inferential methods

3.3.1 The EM algorithm

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

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

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

$$\begin{aligned} \log L_c(\boldsymbol{t};\boldsymbol{\theta}) = & -n\log\sigma+\sum_{i=1}^{n}\left[\left(\frac{t_i-\mu}{\sigma}\right)-\exp\left(\frac{t_i-\mu}{\sigma}\right)\right] \\ & +\sum_{i=1}^{n}(1-\nu_i)\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right). \end{aligned} \tag{3.3.1}$$

Let $\boldsymbol{\delta}=(\delta_1,\delta_2,...,\delta_n)'$ be the vector of censoring indicators, and $\boldsymbol{y}=(y_1,y_2,...,y_n)'$ 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

$$Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})=E_{\boldsymbol{\theta}^{(k)}}\big[\log L_c(\boldsymbol{t};\boldsymbol{\theta})\big|\boldsymbol{y},\boldsymbol{\delta}\big]. \tag{3.3.2}$$

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

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

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

$$M_{(\frac{T_i-\mu}{\sigma})}(\theta) = \exp\big[\exp(\xi_i)\big]\Gamma(\theta+1, e^{\xi_i}), \tag{3.3.3}$$

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 (3.3.3), the required expectations can be derived to be

$$E_{\boldsymbol{\theta}^{(k)}}\left[\frac{T_i-\mu}{\sigma}\big|T_i > y_i\right] = \frac{\mu^{(k)}+\sigma^{(k)}E_{1i}^{(k)}-\mu}{\sigma}, \tag{3.3.4}$$

$$E_{\boldsymbol{\theta}^{(k)}}\left[\exp\left(\frac{T_i-\mu}{\sigma}\right)\big|T_i > y_i\right] = e^{(\frac{\mu^{(k)}-\mu}{\sigma})}\exp\big[\exp\big(\xi_i^{(k)}\big)\big]\Gamma\left(\frac{\sigma^{(k)}}{\sigma}+1, e^{\xi_i^{(k)}}\right), \tag{3.3.5}$$

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 (3.3.4) and (3.3.5) into Eq. (3.3.2), we obtain

$$\begin{aligned} Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = \quad & -n\log\sigma + \sum_{i=1}^{n}(1-\nu_i)\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right) \\ & + \sum_{\delta_i=1}\left\{\left(\frac{t_i-\mu}{\sigma}\right) - \exp\left(\frac{t_i-\mu}{\sigma}\right)\right\} + \sum_{\delta_i=0}\left\{\frac{\mu^{(k)}+\sigma^{(k)}E_{1i}^{(k)}-\mu}{\sigma}\right\} \\ & - \sum_{\delta_i=0}\left\{e^{(\frac{\mu^{(k)}-\mu}{\sigma})}M_{(\frac{T_i-\mu}{\sigma})}\left(\frac{\sigma^{(k)}}{\sigma}\right)\right\}. \end{aligned} \tag{3.3.6}$$

The quantity $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})$ given in (3.3.6) 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 (3.3.6) 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 $\boldsymbol{\theta}^{(k+1)}$. As mentioned in the preceding chapter, this algorithm is a special case of the generalized EM algorithm (Dempster, Laird and Rubin, 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:

$$\begin{aligned}\frac{\partial Q}{\partial \mu} = \;& -\sum_{i=1}^{n} \frac{1-\nu_i}{\sigma} \exp\left(\frac{\tau_i^L - \mu}{\sigma}\right) - \sum_{\delta_i=1} \frac{\{1-\exp(\frac{t_i-\mu}{\sigma})\}}{\sigma} - \sum_{\delta_i=0}\{\frac{1}{\sigma}\} \\ &+ \frac{1}{\sigma}\sum_{\delta_i=0}\left\{\exp\left(\frac{\mu^{(k)}-\mu}{\sigma}\right)\exp(e^{\xi_i^{(k)}})\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}} e^{-w} dw\right\}, \qquad (3.3.7)\end{aligned}$$

$$\begin{aligned}
\frac{\partial Q}{\partial \sigma} = \quad & -\frac{n}{\sigma} - \sum_{i=1}^{n}(1-\nu_i)\left(\frac{\tau_i^L-\mu}{\sigma^2}\right)\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right) - \sum_{\delta_i=1}\left(\frac{t_i-\mu}{\sigma^2}\right) \\
& \times\left\{1-\exp\left(\frac{t_i-\mu}{\sigma}\right)\right\} - \sum_{\delta_i=0}\left\{\frac{\mu^{(k)}+\sigma^{(k)}E_{1i}^{(k)}-\mu}{\sigma^2}\right\} \\
& + \sum_{\delta_i=0}\exp(e^{\xi_i^{(k)}})\exp\left(\frac{\mu^{(k)}-\mu}{\sigma}\right)\left\{\left(\frac{\mu^{(k)}-\mu}{\sigma^2}\right)\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}}e^{-w}dw\right. \\
& \left. + \frac{\sigma^{(k)}}{\sigma^2}\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}}e^{-w}\log w dw\right\}, \qquad (3.3.8)
\end{aligned}$$

$$\begin{aligned}
-\frac{\partial^2 Q}{\partial \mu^2} = \quad & -\sum_{i=1}^{n}\frac{1-\nu_i}{\sigma^2}\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right) + \sum_{\delta_i=1}\frac{\exp(\frac{t_i-\mu}{\sigma})}{\sigma^2} \\
& + \frac{1}{\sigma^2}\sum_{\delta_i=0}\left\{\exp\left(\frac{\mu^{(k)}-\mu}{\sigma}\right)\exp(e^{\xi_i^{(k)}})\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}}e^{-w}dw\right\}, \qquad (3.3.9)
\end{aligned}$$

$$\begin{aligned}
-\frac{\partial^2 Q}{\partial \mu \partial \sigma} = \quad & -\sum_{i=1}^{n}(1-\nu_i)\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right)\left\{\frac{1}{\sigma^2}+\left(\frac{\tau_i^L-\mu}{\sigma^3}\right)\right\} \\
& -\sum_{\delta_i=1}\frac{1}{\sigma^2}\left\{1-\exp\left(\frac{t_i-\mu}{\sigma}\right)-\left(\frac{t_i-\mu}{\sigma}\right)\exp\left(\frac{t_i-\mu}{\sigma}\right)\right\} - \sum_{\delta_i=0}\left\{\frac{1}{\sigma^2}\right\} \\
& + \sum_{\delta_i=0}\exp(e^{\xi_i^{(k)}})\exp\left(\frac{\mu^{(k)}-\mu}{\sigma}\right)\left\{\left(\frac{\mu^{(k)}-\mu}{\sigma^3}+\frac{1}{\sigma^2}\right)\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}}e^{-w}dw\right. \\
& \left. + \frac{\sigma^{(k)}}{\sigma^3}\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}}e^{-w}\log w dw\right\}, \qquad (3.3.10)
\end{aligned}$$

$$
\begin{aligned}
-\tfrac{\partial^2 Q}{\partial \sigma^2} = \quad & -\frac{n}{\sigma^2} - \sum_{i=1}^{n}(1-\nu_i)\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right)\left\{\frac{2(\tau_i^L-\mu)}{\sigma^3}+\left(\frac{\tau_i^L-\mu}{\sigma^2}\right)^2\right\} \\
& -\sum_{\delta_i=1}\left(\frac{t_i-\mu}{\sigma^2}\right)\left\{\frac{2}{\sigma}-\frac{2}{\sigma}\exp\left(\frac{t_i-\mu}{\sigma}\right)-\left(\frac{t_i-\mu}{\sigma^2}\right)\exp\left(\frac{t_i-\mu}{\sigma}\right)\right\} \\
& -\sum_{\delta_i=0}\left\{\frac{2(\mu^{(k)}+\sigma^{(k)}E_{1i}^{(k)}-\mu)}{\sigma^3}\right\}+\sum_{\delta_i=0}\exp(e^{\xi_i^{(k)}})\exp\left(\frac{\mu^{(k)}-\mu}{\sigma}\right) \\
& \times\left\{\left\{\left(\frac{\mu^{(k)}-\mu}{\sigma^2}\right)^2+\frac{2(\mu^{(k)}-\mu)}{\sigma^3}\right\}\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}}e^{-w}dw\right. \\
& +\left\{\frac{2(\mu^{(k)}-\mu)\sigma^{(k)}}{\sigma^4}+\frac{2\sigma^{(k)}}{\sigma^3}\right\}\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}}e^{-w}\log w dw \\
& \left.+\left(\frac{\sigma^{(k)}}{\sigma^2}\right)^2\int_{e^{\xi_i^{(k)}}}^{\infty} w^{\frac{\sigma^{(k)}}{\sigma}}e^{-w}(\log w)^2 dw\right\}. \qquad (3.3.11)
\end{aligned}
$$

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 Appendix II. It has been observed in our extensive empirical study that the numerical MLEs converge to the true parameter values quite accurately.

3.3.2 Asymptotic variances and covariance of the MLEs

As described in Section 1.5.2, using the missing information principle of Louis (1982), we can derive the asymptotic variance-covariance matrix of the MLEs within the EM framework.

The elements of the complete information matrix $I_T(\boldsymbol{\theta})$, given by (1.5.1), are given by

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

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

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

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

Next, to obtain the elements of the missing information matrix, the required conditional density can be derived as [see Ng, Chan and Balakrishnan (2002)]

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

Let us consider the logarithm of the truncated extreme value density given by

$$\log f_{C_i|Y_i}(c_i|C_i > y_i, \mu, \sigma) = -\log\sigma + \exp(e^{\xi_i}) + \left(\frac{c_i-\mu}{\sigma}\right) - \exp\left(\frac{c_i-\mu}{\sigma}\right), \quad (3.3.12)$$

from which we obtain the following expressions:

$$\tfrac{\partial^2}{\partial\mu^2}\log f_{C_i|Y_i} = \frac{1}{\sigma^2}[e^{\xi_i} - e^{\beta_i}], \quad (3.3.13)$$

$$\tfrac{\partial^2}{\partial\sigma^2}\log f_{C_i|Y_i} = \frac{1}{\sigma^2}[1 + 2\xi_i e^{\xi_i} + \xi_i^2 e^{\xi_i} + 2\beta_i - 2\beta_i e^{\beta_i} - \beta_i^2 e^{\beta_i}], \quad (3.3.14)$$

$$\frac{\partial^2}{\partial\mu\partial\sigma}\log f_{C_i|Y_i} = \frac{1}{\sigma^2}[1+e^{\xi_i}+\xi_i e^{\xi_i}-e^{\beta_i}-\beta_i e^{\beta_i}], \tag{3.3.15}$$

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

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

Using this expression, the following expectations can be obtained:

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

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

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

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

Using Eqs. (3.3.12) - (3.3.14) and these expectations, the expected missing information matrix $I_{C|Y}(\boldsymbol{\theta})$ can be obtained as in (1.5.3), and then the observed Fisher information matrix can be obtained from Eq. (1.5.4). Finally, by inverting $I_Y(\hat{\boldsymbol{\theta}})$, the asymptotic variance-covariance matrix of the MLEs can be obtained.

3.3.3 Newton-Raphson Method

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. As in the last chapter, 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. The asymptotic variance-covariance matrix of the MLEs can be obtained by inverting the observed information matrix in the NR method. Here, we have constructed the asymptotic confidence intervals based on these asymptotic variances, and have compared them to the asymptotic confidence intervals corresponding to the EM algorithm.

3.3.4 Confidence intervals

As described in Section 1.5.3, we can derive the asymptotic confidence intervals for the parameters by using the asymptotic normality of the MLEs. It is also possible to derive asymptotic confidence intervals for the parameters by using observed information matrix or the parametric bootstrap technique.

Another alternative could be to construct nonparametric bootstrap confidence intervals for the parameters. However, we do not study the performance of the bootstrap confidence intervals here because of some practical difficulties mentioned in Section 3.4.

3.3.5 An application to prediction

With the estimated parameters μ and σ, we 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 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

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

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. We can obtain an estimate $\hat{\pi}$ by using the MLE $\hat{\boldsymbol{\theta}}$ as

$$\hat{\pi} = \exp\left\{\exp\left(\frac{b - \hat{\mu}}{\hat{\sigma}}\right) - \exp\left(\frac{a - \hat{\mu}}{\hat{\sigma}}\right)\right\} = g(\hat{\boldsymbol{\theta}}), \tag{3.3.16}$$

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 $\hat{\boldsymbol{\theta}}$, we 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

$$\widehat{\mathrm{Var}(\hat{\pi})} = \left(\left(\frac{\partial g}{\partial \mu}\right)^2 \mathrm{Var}(\hat{\mu}) + 2\left(\frac{\partial g}{\partial \mu}\right)\left(\frac{\partial g}{\partial \sigma}\right)\mathrm{Cov}(\hat{\mu}, \hat{\sigma}) + \left(\frac{\partial g}{\partial \sigma}\right)^2 \mathrm{Var}(\hat{\sigma}) \right)\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \quad (3.3.17)$$

Examples of this technique are given in Section 3.5.

3.4 Simulation

3.4.1 Setup

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, Meeker and McCalley (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 estimation methods for the model behave under small and heavy truncation. The set 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) were 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.

3.4.2 Results and discussion

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

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

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 3.1 - 3.3, 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

Table 3.1: 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

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 3.1. 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.

The bias and mean square error of the parameter estimates corresponding to the EM algorithm and the NR methods are presented in Table 3.2. 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 3.2 that the bias and mean square error of the parameter estimates obtained from the two methods are quite close, with some occasional disagreement.

Table 3.2: Bias (B) and mean square error (MSE) for the EM algorithm and the NR method.

$n = 100$						
(μ, σ)	Trunc.	Method	B($\hat{\mu}$)	B($\hat{\sigma}$)	MSE($\hat{\mu}$)	MSE($\hat{\sigma}$)
(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
$n = 200$						
(μ, σ)	Trunc.	Method	B($\hat{\mu}$)	B($\hat{\sigma}$)	MSE($\hat{\mu}$)	MSE($\hat{\sigma}$)
(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
$n = 300$						
(μ, σ)	Trunc.	Method	B($\hat{\mu}$)	B($\hat{\sigma}$)	MSE($\hat{\mu}$)	MSE($\hat{\sigma}$)
(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 3.3: Coverage probabilities for the two asymptotic confidence intervals for μ for different nominal confidence levels and different simulation settings.

n=100				
μ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
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				
μ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
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				
μ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
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

*Note: The * values are significantly different from the nominal level with 95% confidence.*

Table 3.4: Coverage probabilities for the two asymptotic confidence intervals for σ for different nominal confidence levels and different simulation settings.

n=100				
σ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
0.33	30%	90%	0.895	0.896
		95%	0.947	0.943
	40%	90%	0.875*	0.881*
		95%	0.918*	0.923*
0.25	30%	90%	0.898	0.906
		95%	0.933*	0.936*
	40%	90%	0.883	0.897
		95%	0.934*	0.940

n=200				
σ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
0.33	30%	90%	0.887	0.891
		95%	0.932*	0.938
	40%	90%	0.888	0.887
		95%	0.935*	0.943
0.25	30%	90%	0.890	0.910
		95%	0.937	0.948
	40%	90%	0.873*	0.883
		95%	0.928*	0.936*

n=300				
σ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
0.33	30%	90%	0.898	0.907
		95%	0.947	0.952
	40%	90%	0.905	0.908
		95%	0.947	0.950
0.25	30%	90%	0.879*	0.901
		95%	0.933*	0.943
	40%	90%	0.898	0.908
		95%	0.944	0.949

*Note: The * values are significantly different from the nominal level with 95% confidence.*

Table 3.3 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 3.3, Table 3.4 presents the coverage probabilities for the two asymptotic confidence intervals for the parameter σ.

From Tables 3.3 and 3.4, it can be noticed that the coverage probabilities for μ for the two methods are always close to each other, and are significantly lower than the nominal level. However, for σ, both methods give coverage probabilities close to the nominal level in most cases, and also the coverage probabilities corresponding to the EM are lower than those of the NR method in almost all the cases. In general, however, the coverage probabilities for both μ and σ remain reasonably close to the nominal level, for both methods.

3.5 Illustrative example

3.5.1 Point and interval estimation

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 presented in Table B.1 in Appendix B.

In Tables 3.5 and 3.6, the successive steps of the iteration process 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 should 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 is 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 3.7 gives the asymptotic confidence intervals corresponding to the EM algorithm, the NR method, and the parametric bootstrap technique.

It can be noticed that in both cases, the confidence intervals for μ and σ obtained

Table 3.5: 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 3.6: 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

by the EM algorithm are very close to those obtained by the NR and bootstrap methods. The parametric bootstrap confidence intervals for μ are seen here to be slightly wider than the other two, while for σ, they are comparable to those obtained

Table 3.7: 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%	(0.276, 0.407)	(0.275, 0.409)	(0.277, 0.411)
	95%	(0.264, 0.420)	(0.262, 0.422)	(0.264, 0.424)

my the EM and NR methods. Thus, we would expect the parametric bootstrap method to have a higher coverage probability for μ.

By employing the missing information principle, the complete information matrix for this data is obtained to be $\begin{pmatrix} 855.7993 & 473.2701 \\ & 1415.6032 \end{pmatrix}$, while the observed information matrix to be $\begin{pmatrix} 427.8996 & -11.6697 \\ & 630.8986 \end{pmatrix}$. Thus, the ratio of the determinants of the observed information matrix to the complete information matrix is 0.2732, from which we could provide 72.68% as the proportion of information lost due to censoring. Alternatively, we can compute corresponding variance-covariance matrices as $\begin{pmatrix} 0.0014 & -0.0005 \\ & 0.0009 \end{pmatrix}$ and $\begin{pmatrix} 0.0023 & 0.0000 \\ & 0.0016 \end{pmatrix}$ from which we can compute the trace-efficiency of the estimates based on censored data to complete data to be 58.97% and the determinant-efficiency to be 27.32%.

3.5.2 Prediction

Refer to the 92nd unit in Table B.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 estimated, by using (3.3.15) and the estimated parameters μ and σ, to be 0.273. The standard error of this probability estimate, obtained by using (3.3.16) and the estimated variance-covariance matrix of the MLEs as $\begin{pmatrix} 0.0023 & 0.0000 \\ & 0.0016 \end{pmatrix}$, is given by 0.097. In fact, an approximate 95% confidence interval for this probability is (0.083, 0.463). Similarly, for the 42nd unit in Table B.1, 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.

Chapter 4

Inference for gamma distribution

4.1 Introduction

The gamma distribution is another important distribution for modelling lifetime data in reliability and survival analyses; see, for example, Johnson, Kotz and Balakrishnan (1994), Bowman and Shenton (1988), and the references therein. In this chapter, we discuss the EM algorithm for estimating the parameters of gamma distribution using the maximum likelihood method, based on a left truncated and right censored data. The asymptotic variance-covariance matrix of the MLEs within the EM framework are obtained by using the missing information principle (Louis, 1982). Then, the asymptotic confidence intervals of the parameters are constructed. The Newton-Raphson (NR) method is also used to obtain the MLEs for comparative purposes, and the corresponding asymptotic variances and confidence intervals based on the NR method are also constructed. Then, these inferential methods are compared by means of a detailed Monte Carlo simulation study. Finally, a numerical example is presented to illustrate all the results developed here.

4.2 Likelihood function

Let Y be a gamma random variable with scale parameter θ and shape parameter κ, whose density is given by (1.2.1) and cdf is given by (1.2.2). Let Y denote the lifetime variable, C denote the censoring time variable, δ_i denote the censoring indicator, i.e., δ_i is 0 if the i-th unit is censored, and 1 if it is not. Similarly, let ν_i denote the truncation indicator, i.e., ν_i is 0 if the i-th unit is truncated, and 1 if it is not. Let τ_i^L denote the left-truncation time, and S_1 and S_2 denote the index sets corresponding to the units which are not left truncated and left truncated, respectively.

The likelihood function for the left truncated and right censored data is then obtained from (1.6.1) as

$$L(\kappa, \theta) = \prod_{i \in S_1} \left\{ f_Y(y_i) \right\}^{\delta_i} \left\{ 1 - F_Y(y_i) \right\}^{1-\delta_i} \times \prod_{i \in S_2} \left\{ \frac{f_Y(y_i)}{1 - F_Y(\tau_i^L)} \right\}^{\delta_i} \left\{ \frac{1 - F_Y(y_i)}{1 - F_Y(\tau_i^L)} \right\}^{1-\delta_i}.$$

When the lifetime distribution is gamma with scale parameter θ and shape parameter κ, upon using the truncation indicator ν_i, the likelihood function can be simplified to the form

$$\log L(\kappa,\theta) = \sum_{i=1}^{n}\left[\delta_i\{(\kappa-1)\log y_i - \frac{y_i}{\theta} - \kappa\log\theta\} + (1-\delta_i)\log\Gamma(\kappa,\tau_i^L/\theta)\right] - \sum_{i=1}^{n}\left[\nu_i\log\Gamma(\kappa) + (1-\nu_i)\log\Gamma(\kappa,\tau_i^L/\theta)\right].$$

4.3 Inferential methods

4.3.1 The EM algorithm

The first step to employ the EM algorithm would be to obtain the complete data likelihood. When $\boldsymbol{t}$ denotes the lifetime data, and $\boldsymbol{\lambda} = (\kappa, \theta)$ denotes the parameter vector, with all other notation remaining the same as before, had there been no censoring, the complete data likelihood function would be

$$L_c(\boldsymbol{t};\boldsymbol{\lambda}) = \prod_{i\in S_1}\left\{\frac{t_i^{\kappa-1}e^{-t_i/\theta}}{\theta^\kappa\Gamma(\kappa)}\right\} \times \prod_{i\in S_2}\left\{\frac{t_i^{\kappa-1}e^{-t_i/\theta}}{\theta^\kappa\Gamma(\kappa,\tau_i^L/\theta)}\right\},$$

where $\Gamma(p,x) = \int_x^\infty u^{p-1}e^{-u}du$ is the upper incomplete gamma function. The loglikelihood function, using the truncation indicator ν_i, then becomes

$$\log L_c(\boldsymbol{t};\boldsymbol{\lambda}) = \sum_{i=1}^{n}\left[(\kappa-1)\log t_i - \frac{t_i}{\theta} - \kappa\log\theta\right] - \sum_{i=1}^{n}\left[\nu_i\log\Gamma(\kappa) + (1-\nu_i)\log\Gamma(\kappa,\tau_i^L/\theta)\right].$$

The E-step: Let $\boldsymbol{y} = (y_1, y_2, ..., y_n)'$ denote the observed data vector, where $y_i = \min(t_i,\ c_i)$, as the data are right censored; $\boldsymbol{\delta} = (\delta_1, \delta_2, ..., \delta_n)'$ denotes the vector of censoring indicators. Let the current value of the parameter vector, at the r-th step, be $\boldsymbol{\lambda}^{(r)}$. Our objective is to obtain

$$Q(\boldsymbol{\lambda},\boldsymbol{\lambda}^{(r)}) = E_{\boldsymbol{\lambda}^{(r)}}[\log L_c(\boldsymbol{t};\boldsymbol{\lambda})|\boldsymbol{y},\boldsymbol{\delta}]. \quad (4.3.1)$$

It can be clearly seen that the required expectations are $E_{1i}^{(r)} = E_{\boldsymbol{\lambda}^{(r)}}[\log T_i | T_i > y_i]$ and $E_{2i}^{(r)} = E_{\boldsymbol{\lambda}^{(r)}}[T_i | T_i > y_i]$. These expectations can be obtained from the conditional distribution of T_i, given $T_i > y_i$. The conditional density function is given by

$$f_{T_i|Y_i=y_i}(t_i) = \frac{t_i^{\kappa-1}\exp(-t_i/\theta)}{\theta^{\kappa}\Gamma(\kappa, y_i/\theta)}, \quad t_i > y_i, \theta > 0, \kappa > 0. \tag{4.3.2}$$

Based on this conditional density, we easily find

$$E_{\boldsymbol{\lambda}^{(r)}}[T_i | T_i > y_i] = \frac{\theta^{(r)}\Gamma(\kappa^{(r)} + 1, y_i/\theta^{(r)})}{\Gamma(\kappa^{(r)}, y_i/\theta^{(r)})}.$$

For the other expectation, it can be seen that

$$E_{\boldsymbol{\lambda}^{(r)}}[\log T_i | T_i > y_i] = \log\theta^{(r)} + \frac{1}{\Gamma(\kappa^{(r)}, y_i/\theta^{(r)})}\int_{y_i/\theta^{(r)}}^{\infty} u_i^{\kappa^{(r)}-1}\log u_i e^{-u_i} du_i. \tag{4.3.3}$$

Following Geddes, Glasser, Moore and Scott (1990), we have

$$\int_{y_i/\theta^{(r)}}^{\infty} u_i^{\kappa^{(r)}-1}\log u_i e^{-u_i} du_i = \left[\frac{d}{da}\Gamma(a, y_i/\theta^{(r)})\right]_{a=\kappa^{(r)}}. \tag{4.3.4}$$

Upon using (4.3.4) in Eq. (4.3.3), we get

$$\begin{aligned} E_{\boldsymbol{\lambda}^{(r)}}[\log T_i | T_i > y_i] &= \log\theta^{(r)} + \frac{1}{\Gamma(\kappa^{(r)}, y_i/\theta^{(r)})}\Bigg[\Psi(\kappa^{(r)})\Gamma(\kappa^{(r)}) \\ &\times\left\{1 - e^{-y_i/\theta^{(r)}}\sum_{p=0}^{\infty}\frac{(y_i/\theta^{(r)})^{\kappa^{(r)}+p}}{\Gamma(\kappa^{(r)}+p+1)}\right\} \\ &-\Gamma(\kappa^{(r)})e^{-y_i/\theta^{(r)}}\log(y_i/\theta^{(r)})\sum_{p=0}^{\infty}\frac{(y_i/\theta^{(r)})^{\kappa^{(r)}+p}}{\Gamma(\kappa^{(r)}+p+1)} \\ &+\Gamma(\kappa^{(r)})e^{-y_i/\theta^{(r)}}\sum_{p=0}^{\infty}\frac{(y_i/\theta^{(r)})^{\kappa^{(r)}+p}\Psi(\kappa^{(r)}+p+1)}{\Gamma(\kappa^{(r)}+p+1)}\Bigg]. \end{aligned}$$

Substituting these two expectations in Eq. (4.3.1), we obtain

$$\begin{aligned} Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(r)}) = &\left\{\sum_{i:\delta_i=1}(\kappa-1)\log t_i + \sum_{i:\delta_i=0}(\kappa-1)E_{1i}^{(r)}\right\} - \left\{\sum_{i:\delta_i=1}\frac{t_i}{\theta} + \sum_{i:\delta_i=0}\frac{E_{2i}^{(r)}}{\theta}\right\} \\ &-n\kappa\log\theta - \sum_{i=1}^{n}\left\{\nu_i\log\Gamma(\kappa) + (1-\nu_i)\log\Gamma(\kappa, \tau_i^L/\theta)\right\}. \end{aligned}$$

The M-step: The objective now is to maximize $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(r)})$ with respect to $\boldsymbol{\lambda}$ over the parametric space Λ to determine

$$\boldsymbol{\lambda}^{(r+1)} = \arg\max_{\boldsymbol{\lambda}\in\Lambda} Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(r)}).$$

The first-order derivatives of the Q-function with respect to θ and κ are, repectively,

$$\frac{\partial Q}{\partial \theta} = \sum_{i:\delta_i=1} \frac{t_i}{\theta^2} + \sum_{i:\delta_i=0} \frac{E_{2i}^{(r)}}{\theta^2} - \frac{n\kappa}{\theta} - \sum_{i=1}^{n}(1-\nu_i)\frac{\frac{\partial}{\partial\theta}\Gamma(\kappa, \tau_i^L/\theta)}{\Gamma(\kappa, \tau_i^L/\theta)}, \tag{4.3.5}$$

$$\frac{\partial Q}{\partial \kappa} = \sum_{i:\delta_i=1} \log t_i + \sum_{i:\delta_i=0} E_{1i}^{(r)} - n\log\theta - \sum_{i=1}^{n}\left\{\nu_i\Psi(\kappa) + (1-\nu_i)\frac{\frac{\partial}{\partial\kappa}\Gamma(\kappa, \tau_i^L/\theta)}{\Gamma(\kappa, \tau_i^L/\theta)}\right\}. \tag{4.3.6}$$

Now, the upper incomplete gamma function can be easily differentiated with respect to θ using Leibnitz's rule, and we then obtain from (4.3.5) that

$$\theta = \frac{1}{n\kappa}\left[\sum_{i:\delta_i=1} t_i + \sum_{i:\delta_i=0} E_{2i}^{(r)} - \sum_{i=1}^{n}(1-\nu_i)\frac{(\tau_i^L)^\kappa e^{-\tau_i^L/\theta}}{\theta^{\kappa-1}\Gamma(\kappa, \tau_i^L/\theta)}\right]. \tag{4.3.7}$$

The RHS of Eq. (4.3.7) can be evaluated at the current parameter value $\boldsymbol{\lambda}^{(r)}$ to obtain the updated parameter estimate $\theta^{(r+1)}$.

Upon expanding the incomplete gamma function as an infinite series, then differentiating and simplifying the expression, (4.3.6) can be expressed as

$$\begin{aligned}\tfrac{\partial Q}{\partial \kappa} = \; & \sum_{i:\delta_i=1} \log t_i + \sum_{i:\delta_i=0} E_{1i}^{(r)} - n\log\theta - n\Psi(\kappa) \\ & - \sum_{i=1}^{n}(1-\nu_i)\left[\log(\tau_i^L/\theta) - \log(\tau_i^L/\theta)\Big/\left\{1 - e^{-\tau_i^L/\theta}\sum_{p=0}^{\infty}\frac{(\tau_i^L/\theta)^{\kappa+p}}{\Gamma(\kappa+p+1)}\right\}\right. \\ & \left. + e^{-\tau_i^L/\theta}\sum_{p=0}^{\infty}\frac{(\tau_i^L/\theta)^{\kappa+p}\Psi(\kappa+p+1)}{\Gamma(\kappa+p+1)}\Big/\left\{1 - e^{-\tau_i^L/\theta}\sum_{p=0}^{\infty}\frac{(\tau_i^L/\theta)^{\kappa+p}}{\Gamma(\kappa+p+1)}\right\}\right].\end{aligned}$$

Equating $\frac{\partial Q}{\partial \kappa}$ to zero, the equation can be solved numerically for κ to obtain the current estimate $\kappa^{(r+1)}$ by using $\theta^{(r+1)}$ for θ. The E-step and the M-step are then repeated till convergence is achieved to the desired level of accuracy.

4.3.2 Asymptotic variances and covariance of the MLEs

By using the missing information principle as described in Section 1.5.2, we can derive the asymptotic variance-covariance matrix of the MLEs.

The elements of the complete information matrix, given by (1.5.1), are given by

$$-E\left[\tfrac{\partial^2}{\partial\theta^2}\log L_c(\boldsymbol{t};\boldsymbol{\lambda})\right] = \frac{n\kappa}{\theta^2} + \sum_{i=1}^{n}(1-\nu_i)\left[\frac{{\tau_i^L}^{\kappa} e^{-\tau_i^L/\theta}}{\theta^{\kappa+2}\Gamma(\kappa,\tau_i^L/\theta)}\left\{\frac{\tau_i^L}{\theta}-\kappa+1\right\} - \left\{\frac{\left(\frac{{\tau_i^L}^{\kappa}}{\theta^{\kappa+1}}\right)e^{-\tau_i^L/\theta}}{\Gamma(\kappa,\tau_i^L/\theta)}\right\}^2\right],$$

$$-E\left[\tfrac{\partial^2}{\partial\kappa^2}\log L_c(\boldsymbol{t};\boldsymbol{\lambda})\right] = \sum_{i=1}^{n}\left[\nu_i\Psi'(\kappa) + (1-\nu_i)\left\{\frac{\frac{\partial^2}{\partial\kappa^2}\Gamma(\kappa,\tau_i^L/\theta)}{\Gamma(\kappa,\tau_i^L/\theta)} - \left\{\frac{\frac{\partial}{\partial\kappa}\Gamma(\kappa,\tau_i^L/\theta)}{\Gamma(\kappa,\tau_i^L/\theta)}\right\}^2\right\}\right],$$

$$-E\left[\tfrac{\partial^2}{\partial\kappa\partial\theta}\log L_c(\boldsymbol{t};\boldsymbol{\lambda})\right] = \frac{n}{\theta} + \sum_{i=1}^{n}(1-\nu_i)\left(\frac{{\tau_i^L}^{\kappa}}{\theta^{\kappa+1}}\right)e^{-\tau_i^L/\theta}\left[\frac{\log(\tau_i^L/\theta)}{\Gamma(\kappa,\tau_i^L/\theta)} - \frac{\frac{\partial}{\partial\kappa}\Gamma(\kappa,\tau_i^L/\theta)}{\Gamma(\kappa,\tau_i^L/\theta)^2}\right].$$

The conditional density, required for the derivation of the missing information matrix, can be obtained to be [see Ng, Chan and Balakrishnan (2002)]

$$f_{C_i|Y_i}(c_i|C_i>y_i,\theta,\kappa) = \frac{c_i^{\kappa-1}e^{-c_i/\theta}}{\theta^{\kappa}\Gamma(\kappa,y_i/\theta)}, \qquad c_i>y_i,\theta>0,\kappa>0.$$

Using this conditional density, we can easily derive the following expressions:

$$-\frac{\partial^2}{\partial\theta^2}\log f_{C_i|Y_i} = \frac{2c_i}{\theta^3} - \frac{\kappa}{\theta^2} + \frac{e^{-y_i/\theta}y_i^{\kappa}}{\Gamma(\kappa,y_i/\theta)\theta^{\kappa+2}}\left\{\frac{y_i}{\theta}-\kappa-1\right\} - \left\{\frac{e^{-y_i/\theta}(y_i^{\kappa}/\theta^{\kappa+1})}{\Gamma(\kappa,y_i/\theta)}\right\}^2, \quad (4.3.8)$$

$$-\frac{\partial^2}{\partial\kappa^2}\log f_{C_i|Y_i} = \frac{\frac{\partial^2}{\partial\kappa^2}\Gamma(\kappa,y_i/\theta)}{\Gamma(\kappa,y_i/\theta)} - \left\{\frac{\frac{\partial}{\partial\kappa}\Gamma(\kappa,y_i/\theta)}{\Gamma(\kappa,y_i/\theta)}\right\}^2, \qquad (4.3.9)$$

$$-\frac{\partial^2}{\partial\kappa\partial\theta}\log f_{C_i|Y_i} = \frac{1}{\theta} + \frac{e^{-y_i/\theta}y_i^{\kappa}\log(y_i/\theta)}{\Gamma(\kappa, y_i/\theta)\theta^{\kappa+1}} - \frac{e^{-y_i/\theta}(y_i^{\kappa}/\theta^{\kappa+1})\frac{\partial}{\partial\kappa}\Gamma(\kappa, y_i/\theta)}{\Gamma(\kappa, y_i/\theta)^2}. \quad (4.3.10)$$

Clearly, the only required expectation here is $E[C_i|C_i > y_i]$ which can be easily derived from the above conditional density to be

$$E[C_i|C_i > y_i] = \frac{\theta\Gamma(\kappa+1, y_i/\theta)}{\Gamma(\kappa, y_i/\theta)}. \quad (4.3.11)$$

Using Eqs. (1.5.2), (1.5.3), and (4.3.8) - (4.3.11), the expected missing information matrix $I_{C|Y}(\boldsymbol{\lambda})$ can be easily obtained using which the observed information matrix can be obtained without any difficulty from (1.5.4). Finally, the asymptotic variance-covariance matrix of the MLEs can be obtained by inverting $I_Y(\hat{\boldsymbol{\lambda}})$.

4.3.3 Newton-Raphson method

The NR method is employed for obtaining the MLEs for comparison purposes through the R-library function, maxNR. Using the observed information matrix, the asymptotic variances of the MLEs are estimated and the corresponding asymptotic confidence intervals are also obtained. It is seen in the simulation study that the two methods perform quite closely, as shown in Sections 4.4 and 4.5.

4.3.4 Confidence intervals

As described in Section 1.5.3, the asymptotic confidence intervals for the parameters can be easily obtained by using the asymptotic normality of the MLEs, when the asymptotic variance-covariance matrix of the MLEs is obtained by the missing information principle. It is also possible to use the observed information matrix to obtain asymptotic variances of the MLEs, and hence the asymptotic confidence intervals for the parameters. Parametric bootstrap technique can also be used for the same purpose.

4.3.5 An application to prediction

With the estimated parameters θ and κ, we can obtain the probability of a censored unit working till a future year, given that it has worked till 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

$$\pi = \frac{S^*(Y_{fut} - Y_{ins})}{S^*(Y_{cen} - Y_{ins})},$$

where $S^*(\cdot)$ is the survival function of the left truncated random variable. Clearly, the above probability reduces to

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

where $S(\cdot)$ is the survival function of the untruncated lifetime variable, and $g(\cdot)$ is a function of $\boldsymbol{\lambda}$. 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 $\hat{\boldsymbol{\lambda}}$ as

$$\hat{\pi} = \frac{\Gamma(\hat{\kappa}, a/\hat{\theta})}{\Gamma(\hat{\kappa}, b/\hat{\theta})} = g(\hat{\boldsymbol{\lambda}}), \tag{4.3.12}$$

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

Using the delta method, and the asymptotic variance-covariance matrix of the MLE $\hat{\boldsymbol{\lambda}}$, we 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

$$\widehat{\mathrm{Var}(\hat{\pi})} = \left(\Big(\frac{\partial g}{\partial \kappa}\Big)^2 \mathrm{Var}(\hat{\kappa}) + 2\Big(\frac{\partial g}{\partial \kappa}\Big)\Big(\frac{\partial g}{\partial \theta}\Big)\mathrm{Cov}(\hat{\kappa}, \hat{\theta}) + \Big(\frac{\partial g}{\partial \theta}\Big)^2 \mathrm{Var}(\hat{\theta}) \right) \Bigg|_{\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}}. \quad (4.3.13)$$

Examples of this technique are demonstrated with the illustrative example in Section 4.5.

4.4 Simulation

4.4.1 Setup

Entire computational work in this chapter are performed using the R software. The simulation steps are as follows. First, a truncation percentage was fixed. The set 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 rest of the 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 rest of the years of this period. The lifetimes of the machines, in years, are sampled from a gamma distribution with scale parameter θ and shape parameter κ. The years of failure of the machines are obtained by adding these lifetimes with the corresponding installation years. The year of failure of a machine decides whether it is a censored observation 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 along the lines 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 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 unit. This setup produced, along with the desired proportion of truncated observations, sufficiently many censored observations.

The sample sizes used in this study are 100, 200 and 300. The truncation percentages are fixed at 10 and 15. As the gamma distribution is right skewed, if the left truncation percentage is at a higher level, then we would expect the estimates to be highly biased. The two choices of the gamma parameter vector (κ, θ) are made as (4, 8) and (5, 5). In this regard, it is useful to mention that Balakrishnan and Cohen (1991) observed that the standard errors of the gamma parameter estimates based on complete data were not reliable when the shape parameter value was at most 2.5 as the corresponding information matrix becomes nearly singular. Here, for left truncated and right censored data, empirically we observed the same as well. All the simulation results are based on 1000 Monte Carlo runs.

4.4.2 Results and discussion

We know that, when T follows Gamma(κ, θ),

$$E(T) = \theta\kappa, \quad \mathrm{Var}(T) = \kappa\theta^2.$$

From these expressions, the moment estimates of θ and κ can be easily derived. However, these estimates are crude as the data are incomplete. Using these crude moment estimates, we obtain the expected lifetimes of the censored observations. These expected lifetimes of the censored observations, along with the observed lifetimes, can be considered to form a pseudo-complete data. Then, the moment estimates for the parameters are obtained from this pseudo-complete data, and are used as initial values. This method is used in the iterative algorithm throughout this chapter. The bias and mean square error of the parameter estimates for the EM and NR methods are reported in Table 4.1.

Table 4.1: Bias (B) and mean square error (MSE) for the EM algorithm and the NR method.

$n = 100$						
(κ, θ)	Trunc.	Method	B($\hat{\kappa}$)	B($\hat{\theta}$)	MSE($\hat{\kappa}$)	MSE($\hat{\theta}$)
(4, 8)	10%	EM	0.394	-0.374	1.178	4.903
		NR	0.394	-0.373	1.180	4.910
	15%	EM	0.336	-0.313	1.112	4.560
		NR	0.337	-0.313	1.113	4.565
(5, 5)	10%	EM	0.375	-0.195	1.309	1.261
		NR	0.376	-0.196	1.311	1.262
	15%	EM	0.434	-0.279	1.353	1.160
		NR	0.435	-0.280	1.355	1.161
$n = 200$						
(κ, θ)	Trunc.	Method	B($\hat{\kappa}$)	B($\hat{\theta}$)	MSE($\hat{\kappa}$)	MSE($\hat{\theta}$)
(4, 8)	10%	EM	0.256	-0.397	0.539	2.478
		NR	0.256	-0.396	0.540	2.480
	15%	EM	0.254	-0.400	0.509	2.284
		NR	0.254	-0.399	0.510	2.287
(5, 5)	10%	EM	0.288	-0.229	0.662	0.653
		NR	0.289	-0.231	0.662	0.653
	15%	EM	0.286	-0.250	0.577	0.588
		NR	0.287	-0.251	0.578	0.588
$n = 300$						
(κ, θ)	Trunc.	Method	B($\hat{\kappa}$)	B($\hat{\theta}$)	MSE($\hat{\kappa}$)	MSE($\hat{\theta}$)
(4, 8)	10%	EM	0.217	-0.427	0.348	1.708
		NR	0.217	-0.426	0.348	1.709
	15%	EM	0.195	-0.368	0.336	1.580
		NR	0.195	-0.368	0.336	1.582
(5, 5)	10%	EM	0.204	-0.203	0.386	0.428
		NR	0.205	-0.204	0.387	0.428
	15%	EM	0.195	-0.194	0.362	0.396
		NR	0.196	-0.195	0.362	0.396

Table 4.2: Coverage probabilities for the two asymptotic confidence intervals for θ for different nominal confidence levels.

n=100				
θ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
8	10%	90%	0.833*	0.833*
		95%	0.872*	0.872*
	15%	90%	0.837*	0.835*
		95%	0.872*	0.872*
5	10%	90%	0.830*	0.830*
		95%	0.883*	0.883*
	15%	90%	0.836*	0.836*
		95%	0.882*	0.882*

n=200				
θ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
8	10%	90%	0.834*	0.834*
		95%	0.881*	0.880*
	15%	90%	0.828*	0.828*
		95%	0.871*	0.871*
5	10%	90%	0.836*	0.836*
		95%	0.890*	0.890*
	15%	90%	0.839*	0.838*
		95%	0.905*	0.905*

n=300				
θ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
8	10%	90%	0.815*	0.815*
		95%	0.864*	0.864*
	15%	90%	0.841*	0.840*
		95%	0.896*	0.895*
5	10%	90%	0.840*	0.839*
		95%	0.901*	0.901*
	15%	90%	0.843*	0.843*
		95%	0.898*	0.898*

*Note: The * values are significantly different from the nominal level with 95% confidence.*

Table 4.3: Coverage probabilities for the two asymptotic confidence intervals for κ for different nominal confidence levels.

n=100				
κ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
4	10%	90%	0.908	0.909
		95%	0.951	0.951
	15%	90%	0.898	0.898
		95%	0.954	0.954
5	10%	90%	0.893	0.893
		95%	0.951	0.951
	15%	90%	0.890	0.890
		95%	0.951	0.951

n=200				
κ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
4	10%	90%	0.894	0.893
		95%	0.951	0.951
	15%	90%	0.886	0.885
		95%	0.95	0.95
5	10%	90%	0.884	0.884
		95%	0.932*	0.932*
	15%	90%	0.899	0.899
		95%	0.942	0.941

n=300				
κ	Truncation	Nominal CL	Coverage Probability	
			EM	NR
4	10%	90%	0.879*	0.879*
		95%	0.944	0.944
	15%	90%	0.880*	0.880*
		95%	0.948	0.948
5	10%	90%	0.889	0.889
		95%	0.939	0.939
	15%	90%	0.887	0.887
		95%	0.942	0.942

*Note: The * values are significantly different from the nominal level with 95% confidence.*

From Table 4.1, it is clearly seen that the two methods are in quite close agreement, in terms of bias and MSE of the parameter estimates. Table 4.2 gives the coverage probabilities for θ, corresponding to the two asymptotic confidence intervals based on the EM algorithm and the NR method. As the time required for the construction of parametric bootstrap confidence intervals is very high, it was not estimated in the simulation study. But, for illustrative purposes, the parametric bootstrap confidence intervals are constructed in the next section.

It can be noticed from Table 4.2 that the coverage probabilities corresponding to the two methods are quite close, but they are consistently below the nominal level. Analogous to Table 4.2, Table 4.3 gives the coverage probabilities for κ.

It can be noticed from Table 4.3 that the coverage probabilities corresponding to the two methods are again quite close, and are very close to the nominal level also.

4.5 Illustrative example

4.5.1 Point and interval estimation

The sample size used for the illustration is 100, the truncation percentage is 15, and the true parameter value is $(\kappa, \theta) = (5, 5)$. The simulated data that is used for this illustration is presented in Table C.1 in Appendix C. The initial value, calculated by the method of moments explained in the preceding section, is found to be (3.293, 9.895). The tolerance level was set at 0.0001. The EM algorithm took 114 steps to

Table 4.4: Confidence intervals obtained by different methods.

Parameter	Nominal CL	EM	NR	Bootstrap
$\theta = 5$	90%	(3.212, 6.581)	(3.211, 6.580)	(3.453, 6.748)
	95%	(2.889, 6.904)	(2.888, 6.903)	(3.138, 7.064)
$\kappa = 5$	90%	(3.460, 6.461)	(3.461, 6.462)	(2.918, 6.286)
	95%	(3.173, 6.749)	(3.173, 6.750)	(2.596, 6.609)

give the final estimates of (κ, θ) to be (4.961, 4.897), while the NR method took 6 steps to converge to (4.962, 4.896). However, the time taken by the two methods in terms of CPU usage is almost the same. The asymptotic confidence intervals based on the EM algorithm, the NR method and the parametric bootstrap technique are presented in Table 4.4.

From Table 4.4, we observe that confidence intervals based on the EM algorithm and the NR method are quite close. It is also observed that the parametric bootstrap confidence intervals are narrower compared to other asymptotic confidence intervals for θ, while they are slightly wider than the other two for κ.

By employing the missing information principle, the complete information matrix for this data is obtained to be $\begin{pmatrix} 20.6016 & 19.2816 \\ & 20.0112 \end{pmatrix}$, while the observed information is obtained to be $\begin{pmatrix} 17.7724 & 15.2849 \\ & 14.0989 \end{pmatrix}$. Thus, the ratio of the determinants of the observed information matrix to the complete information matrix is 0.4185, from which we could provide 58.15% as the proportion of information lost due to censoring. Alternatively, we can compute the corresponding variance-covariance matrices as $\begin{pmatrix} 0.4943 & -0.4763 \\ & 0.5089 \end{pmatrix}$ and $\begin{pmatrix} 0.8322 & -0.9022 \\ & 1.0490 \end{pmatrix}$ from which we can compute the trace-efficiency of the estimates based on censored data as compared to complete data to be 53.33% and the determinant-efficiency to be 41.85%.

4.5.2 Prediction

Refer to the 95-th unit in Table C.1, for example. For this unit, Y_{ins} is 1977, i.e., the unit is left truncated; also, it is right censored, with censoring year being 2008. The probability that this unit will be working till 2016 is estimated, by using (4.3.12) and the estimated parameters κ and θ, to be 0.415. The standard error of this probability estimate, obtained by using (4.3.13) and the estimated variance-covariance matrix of the MLEs as $\begin{pmatrix} 0.8322 & -0.9022 \\ & 1.0490 \end{pmatrix}$, turns out to be 0.066. In fact, an approximate 95% confidence interval for this probability is (0.286, 0.544). Similarly, for the 15-th unit for which the installation year is 1984 (i.e., not left truncated and also right censored), the probability that the unit will be working till 2016 is estimated to be 0.476, with a standard error of 0.060. An approximate 95% confidence interval for this probability is then (0.358, 0.594). Note that the second unit (installed in 1984) has a higher probability of working till 2016 than the first unit (installed in 1977), as one would expect.

Chapter 5

Inference for generalized gamma distribution

5.1 Introduction

The generalized gamma distribution is a flexible family which includes lognormal, exponential, gamma and Weibull distributions, among others, as special cases. In this chapter, we develop the EM algorithm for estimating the parameters of a three-parameter generalized gamma distribution, based on left truncated and right censored lifetime data. The asymptotic variance-covariance matrix of the MLEs are obtained by using the missing information principle (Louis, 1982), using which asymptotic confidence intervals for the parameters are constructed. The parametric bootstrap confidence intervals for the parameters are also constructed. Then, all these inferential methods are examined through a Monte Carlo simulation study and the corresponding results are presented. A prediction problem regarding the future lifetime of a unit is addressed as well. Finally, a numerical example is presented to illustrate all the results developed here.

5.2 Likelihood function

Let Y be a generalized gamma random variable representing the lifetime variable with scale parameter α and two shape parameters η and κ, whose density is given by (1.2.7). The corresponding survival function of X is (see (1.2.8))

$$S_Y(y) = \frac{\Gamma(\kappa, \left(\frac{y}{\alpha}\right)^{\eta})}{\Gamma(\kappa)}, \quad x > 0, \alpha > 0, \eta > 0, \kappa > 0,$$

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

Let C denote the censoring time variable, and δ_i denote the censoring indicator, i.e., δ_i is 0 if the i-th unit is censored, and 1 if it is not. Similarly, let ν_i denote the truncation indicator, i.e., ν_i is 0 if the i-th unit is truncated, and 1 if it is not. Let τ_i^L denote the left-truncation time, and S_1 and S_2 denote the index sets corresponding to the units which are not left truncated and left truncated, respectively. Then, from (1.6.1), we have the likelihood function for the left truncated and right censored data as

$$L(\alpha, \eta, \kappa) = \prod_{i \in S_1} \left\{f_Y(y_i)\right\}^{\delta_i} \left\{S_Y(y_i)\right\}^{1-\delta_i} \times \prod_{i \in S_2} \left\{\frac{f_Y(y_i)}{S_Y(\tau_i^L)}\right\}^{\delta_i} \left\{\frac{S_Y(y_i)}{S_Y(\tau_i^L)}\right\}^{1-\delta_i}.$$

Upon substituting the pdf and survival function of the generalized gamma, we obtain the loglikelihood function as

$$\begin{aligned}\log L(\alpha,\eta,\kappa) = \ & \sum_{i=1}^{n}\Bigg[\delta_i\Bigg\{\log\eta-\log\alpha+(\kappa\eta-1)\log\left(\frac{y_i}{\alpha}\right)-\left(\frac{y_i}{\alpha}\right)^{\eta}\Bigg\} \\ & +(1-\delta_i)\log\Gamma\left(\kappa,\left(\frac{y_i}{\alpha}\right)^{\eta}\right)\Bigg]-\sum_{i=1}^{n}\Bigg[\nu_i\log\Gamma(\kappa) \\ & +(1-\nu_i)\log\Gamma\left(\kappa,\left(\frac{\tau_i^L}{\alpha}\right)^{\eta}\right)\Bigg].\end{aligned}$$

5.3 Inferential methods

5.3.1 The EM algorithm

We first obtain the complete data likelihood, in order to build the steps of the EM algorithm. When $\boldsymbol{t}$ denotes the lifetime data and $\boldsymbol{\theta}=(\alpha,\eta,\kappa)$ denotes the parameter vector, with all other notation remaining the same as before, had there been no censoring, the complete data likelihood function would be

$$L_c(\boldsymbol{t};\boldsymbol{\theta})=\prod_{i\in S_1}\left\{\frac{\eta(\frac{t_i}{\alpha})^{\kappa\eta-1}}{\alpha\Gamma(\kappa)}\exp\{-(t_i/\alpha)^{\eta}\}\right\}\times\prod_{i\in S_2}\left\{\frac{\eta(\frac{t_i}{\alpha})^{\kappa\eta-1}}{\alpha\Gamma(\kappa,(\frac{\tau_i^L}{\alpha})^{\eta})}\exp\{-(t_i/\alpha)^{\eta}\}\right\}.$$

The loglikelihood function, using the truncation indicator ν_i, then becomes

$$\begin{aligned}\log L_c(\boldsymbol{t};\boldsymbol{\theta}) = \ & n(\log\eta-\log\alpha)+\sum_{i=1}^{n}\left[(\kappa\eta-1)\log(t_i/\alpha)-(t_i/\alpha)^{\eta}\right] \\ & -\sum_{i=1}^{n}\left[\nu_i\log\Gamma(\kappa)+(1-\nu_i)\log\Gamma(\kappa,(\tau_i^L/\alpha)^{\eta})\right].\end{aligned}$$

The E-step: Let $\boldsymbol{y}=(y_1,y_2,...,y_n)'$ denote the observed data vector, where $y_i=\min(t_i,c_i)$, as the data are right censored, and $\boldsymbol{\delta}=(\delta_1,\delta_2,...,\delta_n)'$ denote the vector of censoring indicators. Let the current value of the parameter vector, at the r-th step, be $\boldsymbol{\theta}^{(r)}$. Then we need to obtain

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(r)}) = E_{\boldsymbol{\theta}^{(r)}}[\log L_c(\boldsymbol{t}; \boldsymbol{\theta})|\boldsymbol{y}, \boldsymbol{\delta}]. \tag{5.3.1}$$

It can be readily seen that the expectations of interest are $E_{\boldsymbol{\theta}^{(r)}}[\log(T_i/\alpha)|T_i > y_i]$ and $E_{\boldsymbol{\theta}^{(r)}}[(T_i/\alpha)^{\eta}|T_i > y_i]$. These conditional expectations are based on the conditional density of T_i, given $T_i > y_i$, which is given by

$$f_{T_i|Y_i=y_i}(t_i) = \frac{\eta}{\alpha\Gamma(\kappa, (y_i/\alpha)^{\eta})}\left(\frac{t_i}{\alpha}\right)^{\kappa\eta-1} \exp\left\{-\left(\frac{t_i}{\alpha}\right)^{\eta}\right\}, \quad t_i > y_i, \alpha > 0, \eta > 0, \kappa > 0. \tag{5.3.2}$$

Using this conditional density, it can be readily obtained that

$$E_{\boldsymbol{\theta}^{(r)}}[(T_i/\alpha)^{\eta}|T_i > y_i] = \left\{\frac{\alpha^{(r)}}{\alpha}\right\}^{\eta} \frac{\Gamma(\kappa^{(r)} + \eta/\eta^{(r)}, (y_i/\alpha^{(r)})^{\eta^{(r)}})}{\Gamma(\kappa^{(r)}, (y_i/\alpha^{(r)})^{\eta^{(r)}})}.$$

For the other required expectation, we see that

$$E_{\boldsymbol{\theta}^{(r)}}[\log(T_i/\alpha)|T_i > y_i] = E_{\boldsymbol{\theta}^{(r)}}[\log T_i|T_i > y_i] - \log\alpha = E_i^{(r)} - \log\alpha.$$

Now, note that

$$\begin{aligned} E_i^{(r)} &= E_{\boldsymbol{\theta}^{(r)}}[\log T_i|T_i > y_i] \\ &= \log\alpha^{(r)} + \frac{1}{\eta^{(r)}\Gamma(\kappa^{(r)}, (\frac{y_i}{\alpha^{(r)}})^{\eta^{(r)}})} \int_{(\frac{y_i}{\alpha^{(r)}})^{\eta^{(r)}}}^{\infty} u_i^{\kappa^{(r)}-1} \log u_i e^{-u_i} du_i. \end{aligned} \tag{5.3.3}$$

Then, using the result of Geddes, Glasser, Moore and Scott (1990), we have

$$\int_{(y_i/\alpha^{(r)})^{\eta^{(r)}}}^{\infty} u_i^{\kappa^{(r)}-1} \log u_i e^{-u_i} du_i = \left[\frac{d}{da}\Gamma(a, (y_i/\alpha^{(r)})^{\eta^{(r)}})\right]_{a=\kappa^{(r)}}. \tag{5.3.4}$$

Upon substituting the expression of the integral in (5.3.4) into (5.3.3), we get

$$\begin{aligned} E_{\boldsymbol{\theta}^{(r)}}[\log T_i | T_i > y_i] &= \log \alpha^{(r)} + \frac{1}{\eta^{(r)} \Gamma(\kappa^{(r)}, (y_i/\alpha^{(r)})^{(\eta^{(r)})})} \Bigg[\Psi(\kappa^{(r)}) \Gamma(\kappa^{(r)}) \Bigg\{ 1 - e^{-(y_i/\alpha^{(r)})^{\eta^{(r)}}} \\ &\times \sum_{p=0}^{\infty} \frac{(y_i/\alpha^{(r)})^{\eta^{(r)}(\kappa^{(r)}+p)}}{\Gamma(\kappa^{(r)}+p+1)} \Bigg\} - \Gamma(\kappa^{(r)}) e^{-(y_i/\alpha^{(r)})^{\eta^{(r)}}} \Bigg\{ \log\{(y_i/\alpha^{(r)})^{\eta^{(r)}}\} \\ &\times \sum_{p=0}^{\infty} \frac{(y_i/\alpha^{(r)})^{\eta^{(r)}(\kappa^{(r)}+p)}}{\Gamma(\kappa^{(r)}+p+1)} - \sum_{p=0}^{\infty} \frac{(y_i/\alpha^{(r)})^{\eta^{(r)}(\kappa^{(r)}+p)} \Psi(\kappa^{(r)}+p+1)}{\Gamma(\kappa^{(r)}+p+1)} \Bigg\} \Bigg]. \end{aligned}$$

Upon substituting these two expectations into (5.3.1), we finally obtain

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(r)}) = \quad & n(\log \eta - \log \alpha) + (\kappa\eta - 1) \Big\{ \sum_{i:\delta_i=1} (\log t_i - \log \alpha) + \sum_{i:\delta_i=0} (E_i^{(r)} - \log \alpha) \Big\} \\ & - \Bigg\{ \sum_{i:\delta_i=1} \left(\frac{t_i}{\alpha}\right)^{\eta} + \sum_{i:\delta_i=0} \left(\frac{\alpha^{(r)}}{\alpha}\right)^{\eta} \frac{\Gamma(\kappa^{(r)} + \eta/\eta^{(r)}, (y_i/\alpha^{(r)})^{\eta^{(r)}})}{\Gamma(\kappa^{(r)}, (y_i/\alpha^{(r)})^{\eta^{(r)}})} \Bigg\} \\ & - \sum_{i=1}^{n} \big[\nu_i \log \Gamma(\kappa) + (1-\nu_i) \log \Gamma(\kappa, (\tau_i^L/\alpha)^{\eta}) \big]. \end{aligned} \quad (5.3.5)$$

In the next step (M-step), the objective is to maximize the expression in (5.3.5) with respect to $\boldsymbol{\theta}$.

The M-step: In this step, the objective is to maximize $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(r)})$ with respect to $\boldsymbol{\theta}$ over the parameter space Θ to determine the $(r+1)$-th step iterate as

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

Clearly, we can see from (5.3.5) that this maximization is not straightforward.

Here, we use a slight variation of the EM-gradient algorithm of Lange (1995). In the M-step, we use a two stage process as follows. In the $r-$th step, first for a given value of the shape parameter $\kappa^{(r)}$, we perform a one-step Newton-Raphson on the $Q-$function to obtain the updated values $\alpha^{(r+1)}$ and $\eta^{(r+1)}$. Then, with these updated values, we perform another one-step Newton-Raphson on the $Q-$function to obtain the updated value $\kappa^{(r+1)}$. Then with these updated values of all the parameters, we go back to the E-step and the algorithm is continued till convergence is achieved to the desired level of accuracy. The convergence criterion has been taken as the difference between the values of the $Q-$function in two successive steps.

In a Monte Carlo simulation study to be described in Section 5.4, it is observed that the numerical MLEs converge to the true parameter values, taking on an average 20 iterative steps.

5.3.2 Asymptotic variances and covariance of the MLEs

As described in Section 1.5.2, we use the missing information principle here to obtain the asymptotic variance-covariance matrix of the MLEs.

The elements of the complete information matrix, given by (1.5.1), are given by

$$
\begin{aligned}
-E\left[\tfrac{\partial^2 \log L_c}{\partial \alpha^2}\right] &= -\frac{n\kappa\eta}{\alpha^2} + \frac{\eta(\eta+1)}{\alpha^2}\left[\sum_{i=1}^{n}\{\nu_i\kappa + (1-\nu_i)\frac{\Gamma(\kappa+1,(\tau_i^L/\alpha)^\eta)}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)}\}\right] \\
&\quad + \sum_{i=1}^{n}(1-\nu_i)\left[e^{-(\tau_i^L/\alpha)^\eta}\frac{\frac{\eta}{\alpha^2}(\tau_i^L/\alpha)^{\eta\kappa}\{-\eta\kappa - 1 + \eta(\tau_i^L/\alpha)^\eta\}}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)}\right. \\
&\quad \left. -\left\{\frac{\frac{\eta}{\alpha}(\tau_i^L/\alpha)^{\eta\kappa}e^{-(\tau_i^L/\alpha)^\eta}}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)}\right\}^2\right],
\end{aligned}
$$

$$
\begin{aligned}
-E\left[\tfrac{\partial^2 \log L_c}{\partial \eta^2}\right] &= \frac{n}{\eta^2} + \sum_{i=1}^{n}\left[\nu_i\frac{\kappa}{\eta^2}\{\Psi'(\kappa+1) + \Psi^2(\kappa+1)\} + (1-\nu_i)\frac{V_2(\kappa+1,\tau_i^L)}{\eta^2\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)}\right] \\
&\quad - \sum_{i=1}^{n}(1-\nu_i)\left[\frac{\{\log(\frac{\tau_i^L}{\alpha})\}^2 e^{-(\tau_i^L/\alpha)^\eta}(\frac{\tau_i^L}{\alpha})^{\eta\kappa}\{\kappa - (\frac{\tau_i^L}{\alpha})^\eta\}}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)}\right. \\
&\quad \left. + \left\{\frac{\log(\frac{\tau_i^L}{\alpha})(\frac{\tau_i^L}{\alpha})^{\eta\kappa}e^{-(\tau_i^L/\alpha)^\eta}}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)}\right\}^2\right],
\end{aligned}
$$

$$-E\left[\frac{\partial^2 \log L_c}{\partial \kappa^2}\right] = \sum_{i=1}^{n}\left[\nu_i \Psi'(\kappa) + (1-\nu_i)\left\{\frac{V_2(\kappa,\tau_i^L)}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)} - \left\{\frac{V_1(\kappa,\tau_i^L)}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)}\right\}^2\right\}\right],$$

$$\begin{aligned}-E\left[\frac{\partial^2 \log L_c}{\partial \alpha \partial \eta}\right] &= \frac{n\kappa}{\alpha} - \frac{1}{\alpha}\sum_{i=1}^{n}\Bigg[\nu_i(\kappa + \kappa\Psi(\kappa+1)) + (1-\nu_i)\Bigg(\frac{\Gamma(\kappa+1,(\tau_i^L/\alpha)^\eta)}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)} \\ &+ \frac{V_1(\kappa+1,\tau_i^L)}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)}\Bigg)\Bigg] - \sum_{i=1}^{n}\frac{(1-\nu_i)}{\alpha}\Bigg[\frac{(\frac{\tau_i^L}{\alpha})^{\eta\kappa}e^{-(\tau_i^L/\alpha)^\eta}}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)}\{\eta(\frac{\tau_i^L}{\alpha})^\eta \log(\frac{\tau_i^L}{\alpha}) \\ &- \eta\kappa\log(\frac{\tau_i^L}{\alpha}) - 1\} - \frac{\frac{\eta}{\alpha}\log(\frac{\tau_i^L}{\alpha})\{(\tau_i^L/\alpha)^{\eta\kappa}e^{-(\tau_i^L/\alpha)^\eta}\}^2}{\{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)\}^2}\Bigg],\end{aligned}$$

$$\begin{aligned}-E\left[\frac{\partial^2 \log L_c}{\partial \alpha \partial \kappa}\right] &= \frac{n\eta}{\kappa} + \sum_{i=1}^{n}(1-\nu_i)\Bigg[\frac{\frac{\eta^2}{\alpha}(\frac{\tau_i^L}{\alpha})^{\eta\kappa}\log(\frac{\tau_i^L}{\kappa})e^{-(\tau_i^L/\alpha)^\eta}}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)} \\ &- \frac{\frac{\eta}{\alpha}(\frac{\tau_i^L}{\alpha})^{\eta\kappa}e^{-(\tau_i^L/\alpha)^\eta}V_1(\kappa,\tau_i^L)}{\{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)\}^2}\Bigg],\end{aligned}$$

$$\begin{aligned}-E\left[\frac{\partial^2 \log L_c}{\partial \eta \partial \kappa}\right] &= -\sum_{i=1}^{n}\left[\nu_i\frac{\Psi(\kappa)}{\eta} + (1-\nu_i)\frac{V_1(\kappa,\tau_i^L)}{\eta\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)}\right] \\ &- \sum_{i=1}^{n}(1-\nu_i)\Bigg[\frac{\eta\{\log(\frac{\tau_i^L}{\alpha})\}^2(\frac{\tau_i^L}{\alpha})^{\eta\kappa}e^{-(\tau_i^L/\alpha)^\eta}}{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)} \\ &- \frac{\log(\frac{\tau_i^L}{\alpha})(\frac{\tau_i^L}{\alpha})^{\eta\kappa}e^{-(\tau_i^L/\alpha)^\eta}V_1(\kappa,\tau_i^L)}{\{\Gamma(\kappa,(\tau_i^L/\alpha)^\eta)\}^2}\Bigg],\end{aligned}$$

where $\Psi(\cdot)$ and $\Psi'(\cdot)$ are the digamma and trigamma functions, respectively, and the functions $V_1(\cdot,\cdot)$ and $V_2(\cdot,\cdot)$ are given by

$$V_1(a,b) = \frac{\partial}{\partial u}\Gamma(u,(b/\alpha)^\eta)|_{u=a},$$

$$V_2(a,b) = \frac{\partial^2}{\partial u^2}\Gamma(u,(b/\alpha)^\eta)|_{u=a},$$

respectively.

The conditional density required for the derivation of the Fisher information matrix in the $i-$th observation which is censored, referred to as missing information matrix in short, can be derived to be [see Ng, Chan and Balakrishnan (2002)]

$$f_{C_i|Y_i}(c_i|C_i > y_i, \alpha, \eta, \kappa) = \frac{\eta}{\alpha\Gamma(\kappa, (y_i/\alpha)^\eta)}\left(\frac{c_i}{\alpha}\right)^{\kappa\eta-1}\exp\left\{-\left(\frac{c_i}{\eta}\right)^\eta\right\},$$
$$c_i > y_i, \alpha > 0, \eta > 0, \kappa > 0.$$

Using this conditional density, we can easily derive the following expressions:

$$-\tfrac{\partial^2}{\partial\alpha^2}\log f_{C_i|Y_i} = \frac{1}{\alpha^2}\Bigg[-\frac{\eta(\eta\kappa+1)(y_i/\alpha)^{\eta\kappa}e^{-(y_i/\alpha)^\eta}}{\Gamma(\kappa,(y_i/\alpha)^\eta)} + \frac{\eta^2(y_i/\alpha)^{\eta\kappa+\eta}e^{-(y_i/\alpha)^\eta}}{\Gamma(\kappa,(y_i/\alpha)^\eta)} - \left\{\frac{\eta(y_i/\alpha)^{\eta\kappa}e^{-(y_i/\alpha)^\eta}}{\Gamma(\kappa,(y_i/\alpha)^\eta)}\right\}^2 - \kappa\eta + \eta(\eta+1)\left(\frac{c_i}{\alpha}\right)^\eta\Bigg],$$

$$-\tfrac{\partial^2}{\partial\eta^2}\log f_{C_i|Y_i} = \frac{1}{\eta^2} - \frac{\kappa\{\log(y_i/\alpha)\}^2(y_i/\alpha)^{\eta\kappa}e^{-(y_i/\alpha)^\eta}}{\Gamma(\kappa,(y_i/\alpha)^\eta)} + \frac{\{\log(y_i/\alpha)\}^2(y_i/\alpha)^{\eta\kappa+\eta}e^{-(y_i/\alpha)^\eta}}{\Gamma(\kappa,(y_i/\alpha)^\eta)} - \left\{\frac{\log(y_i/\alpha)(y_i/\alpha)^{\eta\kappa}e^{-(y_i/\alpha)^\eta}}{\Gamma(\kappa,(y_i/\alpha)^\eta)}\right\}^2 + \left(\frac{c_i}{\alpha}\right)^2\{\log(c_i/\alpha)\}^2,$$

$$-\tfrac{\partial^2}{\partial\kappa^2}\log f_{C_i|Y_i} = \frac{V_2(\kappa, y_i)}{\Gamma(\kappa,(y_i/\alpha)^\eta)} - \left\{\frac{V_1(\kappa, y_i)}{\Gamma(\kappa,(y_i/\alpha)^\eta)}\right\}^2,$$

$$-\tfrac{\partial^2}{\partial\alpha\partial\eta}\log f_{C_i|Y_i} = \frac{1}{\alpha}\Bigg[\frac{(y_i/\alpha)^{\eta\kappa}e^{-(y_i/\alpha)^\eta}}{\Gamma(\kappa,(y_i/\alpha)^\eta)} + \frac{\log(y_i/\alpha)\eta\kappa(y_i/\alpha)^{\eta\kappa}e^{-(y_i/\alpha)^\eta}}{\Gamma(\kappa,(y_i/\alpha)^\eta)} - \frac{\log(y_i/\alpha)(y_i/\alpha)^{\eta}e^{-(y_i/\alpha)^\eta}\eta(y_i/\alpha)^\eta}{\Gamma(\kappa,(y_i/\alpha)^\eta)} + \kappa + \frac{\eta\log(y_i/\alpha)\{(y_i/\alpha)^{\eta\kappa}e^{-(y_i/\alpha)^\eta}\}^2}{\{\Gamma(\kappa,(y_i/\alpha)^\eta)\}^2} - \eta\left(\frac{c_i}{\alpha}\right)^\eta\log(c_i/\alpha) - \left(\frac{c_i}{\alpha}\right)^\eta\Bigg],$$

$$-\tfrac{\partial^2}{\partial\alpha\partial\kappa}\log f_{C_i|Y_i} = \frac{1}{\alpha}\Bigg[\frac{\eta^2\log(y_i/\alpha)(y_i/\alpha)^{\eta\kappa}e^{-(y_i/\alpha)^\eta}}{\Gamma(\kappa,(y_i/\alpha)^\eta)} - \frac{\eta(y_i/\alpha)^{\eta\kappa}e^{-(y_i/\alpha)^\eta}V_1(\kappa, y_i)}{\{\Gamma(\kappa,(y_i/\alpha)^\eta)\}^2} + \eta\Bigg],$$

$$-\frac{\partial^2}{\partial\eta\partial\kappa}\log f_{C_i|Y_i} = -\frac{\eta\{\log(y_i/\alpha)\}^2(y_i/\alpha)^{\eta\kappa}e^{-(y_i/\alpha)^\eta}}{\Gamma(\kappa,(y_i/\alpha)^\eta)} + \frac{\log(y_i/\alpha)(y_i/\alpha)^{\eta\kappa}e^{-(y_i/\alpha)^\eta}V_1(\kappa,y_i)}{\{\Gamma(\kappa,(y_i/\alpha)^\eta)\}^2} - \log(c_i/\alpha).$$

Clearly, the expectations that are needed in the implementation of the missing information principle are

$$E[(C_i/\alpha)^\eta|C_i > y_i], E[\log(C_i/\alpha)|C_i > y_i], E[(C_i/\alpha)^\eta\log(C_i/\alpha)|C_i > y_i],$$
$$E[(C_i/\alpha)^\eta\{\log(C_i/\alpha)\}^2|C_i > y_i].$$

Using the conditional density of C_i, given $C_i > y_i$, and the result of Geddes, Glasser, Moore and Scott (1990) in the same way earlier, we obtain

$$E[(C_i/\alpha)^\eta|C_i > y_i] = \frac{\Gamma(\kappa+1,(y_i/\alpha)^\eta)}{\Gamma(\kappa,(y_i/\alpha)^\eta)},$$
$$E[\log(C_i/\alpha)|C_i > y_i] = \frac{V_1(\kappa,y_i)}{\eta\Gamma(\kappa,(y_i/\alpha)^\eta)},$$
$$E[(C_i/\alpha)^\eta\log(C_i/\alpha)|C_i > y_i] = \frac{V_1(\kappa+1,y_i)}{\eta\Gamma(\kappa,(y_i/\alpha)^\eta)},$$
$$E[(C_i/\alpha)^\eta\{\log(C_i/\alpha)\}^2|C_i > y_i] = \frac{V_2(\kappa+1,y_i)}{\eta^2\Gamma(\kappa,(y_i/\alpha)^\eta)}.$$

Using these expressions of expectations and the negative of the second derivatives of the logarithm of the conditional density of C_i given $C_i > y_i$, the Fisher information matrix in the $i-$th observation which is censored can be obtained. Then, by using (1.5.3), the expected missing information matrix can be obtained readily. Finally, (1.5.4) would yield the observed information matrix, and the asymptotic variance-covariance matrix of the MLEs can then be obtained by inverting $I_Y(\hat{\boldsymbol{\theta}})$.

5.3.3 Confidence intervals

After obtaining the MLEs and their asymptotic variance-covariance matrix, we can construct asymptotic confidence intervals for the parameters α, η and κ by using the asymptotic normality of the MLEs. However, we observe in our simulation study that the asymptotic variances and covariances of the parameters obtained by the missing information principle highly overestimate the quantiles, and so the corresponding asymptotic confidence intervals become too wide. As a result, the coverage probabilities of the confidence intervals become 100%, at least for sample sizes such as 200. Thus, the asymptotic confidence intervals based on the asymptotic variance-covariance matrix from the missing information principle do not turn out to be a good choice.

As an alternative, we can construct parametric bootstrap confidence intervals for the parameters as described in Section 1.5.3. We constructed parametric boostrap confidence intervals here, and observed that their coverage probabilities are indeed close to the nominal level.

Also, we can construct asymptotic confidence intervals for the parameters using the empirical variances of the parameter estimates. The empirical variances are obtained from a Monte Carlo simulation with 1000 runs, and then these variances are used for constructing the asymptotic confidence intervals for the parameters in another 1000 Monte Carlo runs. We observed in such a simulation study that these empirical asymptotic confidence intervals also have coverage probabilities close to the nominal level, but not for the parameter κ.

5.3.4 An application to prediction

Using the MLEs of α, η and κ, the probability for a censored unit to work until a future year, given that it has worked until Y_{cen} (the right censoring point), can be easily obtained. Suppose a unit is installed in the year Y_{ins}, before 1980, i.e., the unit is left truncated. Clearly, the left truncation time for the unit is $\tau^L = 1980 - Y_{ins}$. Then, the probability that this unit will be working until a future year Y_{fut}, given that it is right censored at Y_{cen}, is given by

$$\pi = \frac{S^*(Y_{fut} - Y_{ins})}{S^*(Y_{cen} - Y_{ins})},$$

where $S^*(\cdot)$ is the survival function of the left truncated random variable. The above probability reduces to

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

$S(\cdot)$ being the survival function of the untruncated lifetime variable and $g(\cdot)$ being a function of $\boldsymbol{\theta}$. It can be easily noticed that this is also the probability of the same event for a unit which is not left truncated. An estimate $\hat{\pi}$ of π, by using the MLE $\hat{\boldsymbol{\theta}}$, can then be obtained as

$$\hat{\pi} = \frac{\Gamma(\hat{\kappa}, (a/\hat{\theta})^{\hat{\eta}})}{\Gamma(\hat{\kappa}, (b/\hat{\theta})^{\hat{\eta}})} = g(\hat{\boldsymbol{\theta}}), \tag{5.3.6}$$

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

One can also obtain an estimate of the standard error of the above estimate $\hat{\pi}$ by using the delta-method and the asymptotic variance-covariance matrix of the MLE $\hat{\boldsymbol{\theta}}$. 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

$$\begin{aligned}\widehat{\mathrm{Var}(\hat{\pi})} = & \Big(\big(\frac{\partial g}{\partial \alpha}\big)^2 \mathrm{Var}(\hat{\alpha}) + \big(\frac{\partial g}{\partial \eta}\big)^2 \mathrm{Var}(\hat{\eta}) + \big(\frac{\partial g}{\partial \kappa}\big)^2 \mathrm{Var}(\hat{\kappa}) \\ & + 2\big(\frac{\partial g}{\partial \alpha}\big)\big(\frac{\partial g}{\partial \eta}\big) \mathrm{Cov}(\hat{\alpha}, \hat{\eta}) + 2\big(\frac{\partial g}{\partial \alpha}\big)\big(\frac{\partial g}{\partial \kappa}\big) \mathrm{Cov}(\hat{\alpha}, \hat{\kappa}) \\ & + 2\big(\frac{\partial g}{\partial \eta}\big)\big(\frac{\partial g}{\partial \kappa}\big) \mathrm{Cov}(\hat{\eta}, \hat{\kappa}) \Big) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \end{aligned} \tag{5.3.7}$$

Examples of this technique are demonstrated in the illustrative example presented in Section 5.5.

5.4 Simulation

5.4.1 Setup

The inferential methods developed in the preceding sections are examined through a Monte Carlo simulation study, using the R software. We now give a brief description of the simulation process. The years of truncation and censoring, as mentioned before, are taken to be 1980 and 2008, respectively. First, a certain truncation percentage is fixed. The installation years are divided into two parts: (1960 - 1979) and (1980 - 1995), around the point of truncation. Unequal probabilities are assigned to the different years as follows: for the period 1980 - 1995, a probability of 0.1 is attached to each of the first six years, and a probability of 0.04 is attached to each of the remaining years of this period. For the period 1960 - 1979, a probability of 0.15 is attached to each of the first five years, and the remaining probability is distributed equally over the rest of the years of this period. The installation years are then sampled from these two sets, through unequal probability with replacement sampling. Then, from a generalized gamma distribution with scale parameter α and shape parameters η and κ, the lifetimes of the machines, in years, are sampled. These lifetimes are then added to the installation years of the machines to obtain the corresponding failure years. The year of failure of a machine determines whether it is a censored observation or not. 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 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 corresponding lifetime are simulated for that unit. This setup produced, along with the fixed proportion of truncated observations, sufficiently many censored observations.

For the simulation study, the sample size used is 200. The truncation percentage is fixed as 20, and the generalized gamma parameters used are $(\alpha, \eta, \kappa) = (15, 3, 5)$. This setup produced 50% censored observations. Now, Balakrishnan and Cohen (1991) observed for complete samples that in order to have the asymptotic variance estimates meaningful, the numerical value of the shape parameter of a gamma distribution must be at least 2.5. As mentioned in the last chapter, we observed the same in the case of left truncated and right censored gamma data. As the κ parameter of the generalized gamma is in fact the gamma shape parameter, we observed the same in the case of generalized gamma distribution as well. We chose the κ parameter to be 5 in our simulation study in order to ensure that the asymptotic variances are reliable.

Also, we have observed that for higher censoring percentages, the asymptotic variances of the MLEs get overestimated and in these situations the corresponding coverage probabilities get close to 100%.

5.4.2 Results and discussion

For choosing the initial values of the parameters, we adopted a two-stage grid-search method. First, an interval around the true value of κ is divided into discrete numbers, and a number is chosen from that set at random, which is then used as the initial value for κ. Then, a two-dimensional grid is formed, of which the different points are the values of (α, η), the intervals being taken around the true values of these parameters. Given the value of κ, that point in this two-dimensional grid at which the observed likelihood is maximized is chosen as the initial value for (α, η).

Table 5.1 presents the bias and mean square errors of the parameter estimates for the above simulation settings based on 1000 Monte Carlo runs. The tolerance required for convergence of the algorithm was taken as 0.001. AT and AI in Table

5.1 denote the average tolerance and average number of iterations, respectively, over 1000 Monte Carlo runs.

Table 5.1: Bias (B) and mean square error (MSE) of the MLEs based on EM algorithm when $n = 200$.

(α, η, κ)	$\hat{\alpha}$		$\hat{\eta}$		$\hat{\kappa}$		AT	AI
	Bias	MSE	Bias	MSE	Bias	MSE		
(15, 3, 5)	0.186	1.941	0.120	0.126	-0.015	0.414	0.0005	21.173

Table 5.2 presents the coverage probabilities for the asymptotic confidence intervals corresponding to the empirical method and the parametric bootstrap technique. Here, NCL denotes nominal confidence level.

Table 5.2: Coverage probabilities of asymptotic confidence intervals by the empirical method and the parametric bootstrap technique when $n = 200$.

Parameter	NCL = 90%		NCL = 95%	
	Empirical	Bootstrap	Empirical	Bootstrap
$\alpha = 15$	0.915	0.905	0.970	0.962
$\eta = 3$	0.897	0.933	0.940	0.961
$\kappa = 5$	0.965	0.923	0.995	0.973

It can be seen from Table 5.1 that the bias and MSEs of the estimates are quite reasonable. From Table 5.2, it can be observed that the parametric bootstrap confidence intervals have coverage probabilties close to the nominal level. The coverage probabilities of the empirical confidence intervals, though close to the nominal level for α and η, it is not so for κ as they are too high.

Thus, under this setup, with regard to confidence intervals of the parameters, we suggest the use of parametric bootstrap, even though it would be computationally quite demanding and may be difficult to use in case of small sample sizes as the estimation method may fail to converge in some instances and for large samples, the program takes immensely longer time to run.

5.5 Illustrative example

5.5.1 Point and interval estimation

We present a numerical example in this section to demonstrate the inferential methods developed in the preceding sections. For this purpose, we use a simulated sample of size 200, with truncation percentage 20 and true parameter value (α, η, κ) = (15, 3, 5). The censoring percentage for this data turns out to be 49.5. These data are presented in Table D.1 of Appendix D. The initial value, chosen as described in the preceding section, turns out to be (14.40, 3.0, 5.4). The tolerance level is set as 0.001. In 34 steps, the EM algorithm converges to the final estimate of (15.07, 3.18, 5.11). Here, we present empirical confidence intervals and the parametric bootstrap confidence intervals for the parameters, and these are reported in Table 5.3.

It can be clearly seen that the confidence intervals corresponding to the empirical method are slightly wider than the parametric boostrap confidence intervals, from which the higher cooverage probabilities for the empirical confidence intervals observed in Table 5.2 can be explained. In this case, the parametric bootstrap confidence

Table 5.3: Asymptotic confidence intervals for parameters based on the empirical method and the parametric bootstrap technique.

Parameter	Nominal CL	Empirical	Bootstrap
$\alpha = 15$	90%	(12.797, 17.337)	(12.829, 16.945)
	95%	(12.362, 17.772)	(12.435, 17.339)
$\eta = 3$	90%	(2.632, 3.730)	(2.515, 3.601)
	95%	(2.526, 3.835)	(2.411, 3.705)
$\kappa = 5$	90%	(4.057, 6.172)	(4.095, 6.176)
	95%	(3.854, 6.374)	(3.896, 6.376)

intervals is the better choice as mentioned earlier, even though it is computationally much more demanding.

5.5.2 Prediction

By using the missing information principle, the asymptotic variance-covariance matrix for the MLEs in this case turns out to be $\begin{pmatrix} 93.1235 & 17.6524 & -51.7954 \\ & 3.3630 & -9.7859 \\ & & 28.9063 \end{pmatrix}$. Clearly, compared to the empirical variances (can be calculated from Table 5.1) of the estimates, these asymptotic variances are considerably high. However, for illustrative purpose, we construct here the confidence interval for the predictive probability as described in Section 5.3.4.

Consider unit no. 4 in Table D.1, which was installed in the year 1990. Suppose we are interested in estimating the probability that this unit will be working till a future year 2016, say. The probability estimate, using the formula (5.3.6) and the parameter estimates $\hat{\alpha}, \hat{\eta}, \hat{\kappa}$ given above, turns out to be 0.361. By using the asymptotic variance-covariance matrix of the MLEs and by using (5.3.7), we obtain the standard error of this estimate to be 0.039. Thus, an approximate 95% confidence interval for this predictive probability becomes (0.285, 0.437).

Chapter 6

Model discrimination

6.1 Introduction

In chapters 2 to 5, we have discussed the fitting of lognormal, Weibull, gamma and generalized gamma distributions to left truncated and right censored data through the EM algorithm. All these distributions are commonly used in the reliability literature for modelling lifetime data. There are other distributions as well which are used in this context. So, for a given lifetime data, it is desirable to choose the most appropriate model from an array of potential candidate models. This problem has been studied in the statistical literature from many perspectives, and is termed as the model discrimination problem in general.

In this chapter, we look into a model discrimination problem specifically in the context of analyzing left truncated and right censored data. Thus, for a given data, if we are to determine which of these distributions provides a better fit, we may employ different techniques depending on their suitability. In the following sections, we discuss this problem and some approaches in detail.

6.2 Different methods

Several methods may be used for model discrimination purposes. These methods can be categorized based on the criteria on which they are based. For example, the most commonly used criteria are the likelihood-based criteria and the information-based criteria. Depending on the nature of the problem, one of these may be used in a particular situation. We now describe these criteria briefly.

6.2.1 Information-based criteria

For a given data, the information-based criteria determines which model provides the maximum information for the data at hand. Clearly, the model that provides the most information for a given data will be adjudged the most appropriate model. The commonly used information-based criteria are the Akaike's information criterion and the Bayesian information criterion.

Akaike's information criterion

Akaike's information criterion or AIC (Akaike, 1974) provides a way for model selection by evaluating an information measure based on the given data. It is given by the expression

$$AIC = 2k - 2\log L,$$

where k is the number of parameters in the model and L is the likelihood value determined at the MLEs of the parameters for the assumed model. For a given data, when there are a number of models as potential candidates, the model with the lowest AIC is the best model to be fitted to that data, according to this criterion. Note that along with the maximized likelihood function, AIC contains a penalty term $2k$, which depends on the number of parameters in the model. Inclusion of this penalty term ensures that this criterion does not favor overfitting, i.e., for a given data, along with maximizing the likelihood function, one should also not fit a model with more parameters than what is essential.

The main advantages of AIC come from two different aspects. First of all, it is simple to apply, and easy to interpret. For a given data and a given set of models, one can calculate the AIC values, and choose the model with minimum AIC to be the best model for that data. Another important advantage of AIC is that it can be applied to models which are not nested; this comes mainly from the fact that the use of AIC is not linked to testing of hypotheses.

Bayesian information criterion

Another commonly used information-based method is the Bayesian information criterion or BIC (Schwarz, 1978) which is given by the expression

$$BIC = k \log n - 2 \log L,$$

where k and L, respectively, are the number of model parameters and the likelihood function value determined at the MLEs of the parameters as before, and n is the sample size. Observe that the penalty term in this case depends on both the number of parameters in the model and the sample size.

Here, we use AIC and BIC for the model discrimination purpose. Through Monte Carlo simulations, we study the behavior of AIC and BIC for the models under consideration.

6.2.2 Likelihood-based criteria

The likelihood-based criteria can be used when the models under consideration are nested. The reason for this is that the likelihood-based criteria test hypotheses regarding the fit. For nested models, specific values of one or more parameters indicate which particular distribution under a parsimonious family is being considered. Then, the hypotheses are formed in such a way that the null and the alternative hypotheses represent the specific distribution under consideration (i.e., the special case of the parsimonious family being considered) and the general case (i.e., all other members of the parsimonious model).

There are several tests that are used in the literature for model discrimination using likelihood-based criteria. Here, we mention the most common among them briefly.

Likelihood-ratio test

The likelihood-ratio test is a statistical test that compares the fit of two different models to a given data. As mentioned before, these models must be nested. Let $\boldsymbol{Y}$ denote the data vector and $\boldsymbol{\theta}$ denote the vector of parameters. Suppose we are interested in testing

$$H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \quad \text{against} \quad H_1 : \boldsymbol{\theta} \in \Theta_1,$$

where $\boldsymbol{\Theta}_0$ is the partition of the parametric space $\boldsymbol{\Theta}$ that corresponds to the distribution under the null hypothesis, and $\boldsymbol{\Theta}_1$ is the partition of $\boldsymbol{\Theta}$ that corresponds to the distribution under the alternative hypothesis. Note that as the models under consideration are nested, there must be a parsimonious family of distributions for which, when the parameter vector $\boldsymbol{\theta}$ belongs to $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}_1$, we get two different distributions (corresponding to the null and the alternative hypotheses, respectively). The likelihood-ratio test statistic is then defined as

$$\lambda(\boldsymbol{y}) = \frac{\text{Sup}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} L(\boldsymbol{\theta}|\boldsymbol{y})}{\text{Sup}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_1} L(\boldsymbol{\theta}|\boldsymbol{y})}.$$

Asymptotically, under some regularity conditions (see Sen and Singer (1993) for details),

$$-2 \log \lambda(\boldsymbol{y}) \longrightarrow \chi^2_1,$$

where χ_1^2 is the χ^2-distribution with 1 degree of freedom.

6.3 Simulation study

Our first objective is to discriminate between the three popular lifetime models – lognormal, Weibull and gamma. We maintain the same setup for simulating left truncated and right censored samples from these distributions, as detailed in preceding chapters. The procedure adopted proceeds as follows. First of all, a left truncated and right censored sample from the lognormal distribution is simulated and the parameters of the lognormal distribution are estimated based on that sample. Based on the same sample, the parameters of the Weibull and gamma distributions are also estimated. Then, using these estimated parameters, the AIC and BIC for the respective models are calculated. As the sample is from lognormal distribution, we would expect the AIC and BIC corresponding to the lognormal model to be minimum among the three. This process is then repeated several times, each time generating a new sample from the lognormal distribution, and the numbers of times the correct model (i.e., lognormal) as well as the wrong model (i.e., Weibull and gamma) is chosen are noted. It is expected that the lognormal model will be chosen as the appropriate model by AIC and BIC with higher probability than the other two models.

Similar procedure was repeated based on samples from the Weibull and gamma distributions as well. Everytime a sample is chosen from one of these distributions, all three distributions are fitted to the data and the AIC and BIC values corresponding to the models are noted. This process gives rise to a 3×3 table, of which the diagonal elements represent the probabilities of choosing the correct model by AIC and BIC. The off-diagonal elements of this table represent the probabilities of choosing another model (from which the sample did not come).

For lognormal distribution, the model parameters used are $(\mu, \sigma) = (3.5, 0.5)$, and the truncation percentage is 60; for Weibull distribution, the model parameters used are $(\alpha, \eta) = (35, 3)$, and the truncation percentage is 40; finally, for gamma distribution, the model parameters used are $(\kappa, \theta) = (5, 5)$, and the truncation percentage is 15. The sample sizes used for all the distributions are 100 and 300, and all the results are based on 500 Monte Carlo simulations. The results so obtained are presented in Table 6.1.

From Table 6.1, we note that AIC and BIC perform without any difference for these models, which is very much expected since all considered models have the same number of parameters and the same sample sizes. We observe that the diagonal elements, which estimate the probabilities of choosing the correct model, are always higher than the off-diaginal elements. This displays that AIC and BIC can successfully discriminate between these models under this setup. Also, we note that the probability of choosing the correct model increases with sample size. This means that, with increasing sample size, AIC and BIC become more accurate in determining the correct underlying distribution of the lifetimes even in the presence of truncation and censoring.

Table 6.1: Proportion of times the model is chosen by AIC and BIC.

n=100				
Original data	Criteria	Fit		
		Lognormal	Weibull	Gamma
Lognormal	AIC	0.750	0.062	0.188
	BIC	0.750	0.062	0.188
Weibull	AIC	0.060	0.702	0.238
	BIC	0.060	0.702	0.238
Gamma	AIC	0.340	0.248	0.412
	BIC	0.340	0.248	0.412
n=300				
Original data	Criteria	Fit		
		Lognormal	Weibull	Gamma
Lognormal	AIC	0.796	0.008	0.196
	BIC	0.796	0.008	0.196
Weibull	AIC	0.000	0.864	0.136
	BIC	0.000	0.864	0.136
Gamma	AIC	0.180	0.118	0.702
	BIC	0.180	0.118	0.702

Next, we investigate the model discrimination problem from another perspective. We simulate a left truncated and right censored data from the generalized gamma distribution, and fit all four distributions to the data, i.e., lognormal, Weibull, gamma and generalized gamma. We calculate the AIC and BIC values for all the models, and based on these values, we determine which of the considered models is to be selected. We repeat this procedure 500 times and observe once again the proportion of times the different models are selected.

The setting used for the generalized gamma distribution is as follows: $(\alpha, \eta, \kappa) =$ (15, 3, 5), truncation percentage 20, sample sizes 500 and 800. The results obtained from this simulation study are presented in Table 6.2.

Table 6.2: Proportion of times the model is chosen by AIC and BIC when the data are from generalized gamma distribution.

Original data: Generalized gamma	Criteria	Fit			
		Lognormal	Weibull	Gamma	Generalized gamma
$n = 500$	AIC	0.004	0.100	0.310	0.586
	BIC	0.004	0.238	0.672	0.086
$n = 800$	AIC	0.004	0.032	0.210	0.754
	BIC	0.004	0.102	0.526	0.368

From Table 6.2 we observe that AIC selects the generalized gamma model with higher probability than all other models; and this probability increases, as we would expect, when the sample size increases. However, according to BIC, in both cases, the model that is chosen most of the times is the gamma distribution. Even though the probability of selecting the generalized gamma distribution by BIC increases with sample size, it still remains less than the probability corresponding to the gamma distribution. This may be due to the fact that the penalty term of BIC corresponding to the generalized gamma distribution (having three parameters) turns out to be more severe than necessary. Having said that, the simulation results also display that with an increase in sample size, BIC is able to discriminate between the models more clearly. In general, based on the above results, we recommend the use of AIC for model discrimination under this setup, when the left truncated and right censored data arise from the generalized gamma distribution.

In the next section, we give some numerical examples to illustrate the use of AIC and BIC criteria.

6.4 Illustrative example

For the first illustrative example, we use the same datasets which were used in the numerical examples in the preceding chapters dealing with lognormal, Weibull and gamma distributions. For lognormal distribution, the model parameters are $(\mu, \sigma) =$ (3.5, 0.5), the truncation percentage is 60 and the sample size 100, and the data are as presented in Table A.1. Then, lognormal, Weibull and gamma distributions are fitted to these data, and the AIC and BIC values corresponding to all the models are calculated. For Weibull distribution, the model parameters are $(\alpha, \eta) =$ (35, 3), the truncation percentage is 40 and the sample size is 100, and the data are as presented in Table B.1. Similarly, for gamma distribution, the model parameters are $(\kappa, \theta) =$ (5, 5), the truncation percentage 15 and the sample size is 100, and the data are as presented in Table C.1. The same process is repeated for these two data sets as well. The AIC and BIC values so calculated are presented in Table 6.3.

The model with the minimum AIC or BIC values is the most appropriate model, and it can be seen that the AIC and BIC values always agree with the correct model for these datasets.

To illustrate the model discrimination problem when the sample is obtained from

Table 6.3: The AIC and BIC values for different models.

Original data	Criteria	Fit		
		Lognormal	Weibull	Gamma
Lognormal	AIC	423.2089	425.3809	423.6208
	BIC	428.4192	430.5913	428.8312
Weibull	AIC	425.2395	418.8227	421.5560
	BIC	430.4498	424.0330	426.7664
Gamma	AIC	453.2284	454.1567	451.9672
	BIC	458.4387	459.3670	457.1776

the generalized gamma distribution, the model parameters used are $(\alpha, \eta, \kappa) = (15, 3, 5)$, truncation percentage is 20 and the sample sizes are 500 and 800. Table 6.4 presents the corresponding results.

Table 6.4: The AIC and BIC values when the data are from generalized gamma distribution.

Original data: Generalized gamma	Criteria	Fit			
		Lognormal	Weibull	Gamma	Generalized gamma
$n = 500$	AIC	1596.389	1596.808	1590.568	1587.682
	BIC	1604.818	1605.237	1598.998	1600.326
$n = 800$	AIC	2510.796	2521.563	2505.580	2504.048
	BIC	2520.165	2530.932	2514.949	2518.102

We observe from Table 6.4 that when the sample size is small (500), though AIC correctly discriminates between the models, the BIC is unable to do so. However, when the sample size is increased to 800, then the BIC still fails to do so. This is in conformance with the conclusions drawn in Table 6.2 based on an extensive Monte Carlo study.

Though the generalized gamma distribution is close to all three considered models, Figures 6.1 to 6.3 do suggest that the generalized gamma distribution is closest to the gamma distribution. In fact, for the above example, when the sample size is 500, the Kolmogorov-Smirnov statistic measuring the distances between the generalized gamma and each of lognormal, Weibull and gamma distributions turn out to be 0.033, 0.029 and 0.019, respectively. These show that the differences between the special cases and the generalized gamma distribution are indeed small, with the difference for the gamma being the smallest. Incidentally, this may also explain as to why gamma distribution is selected as the best model by AIC a number of times and by BIC most of the time, as presented in Tables 6.2 and 6.4.

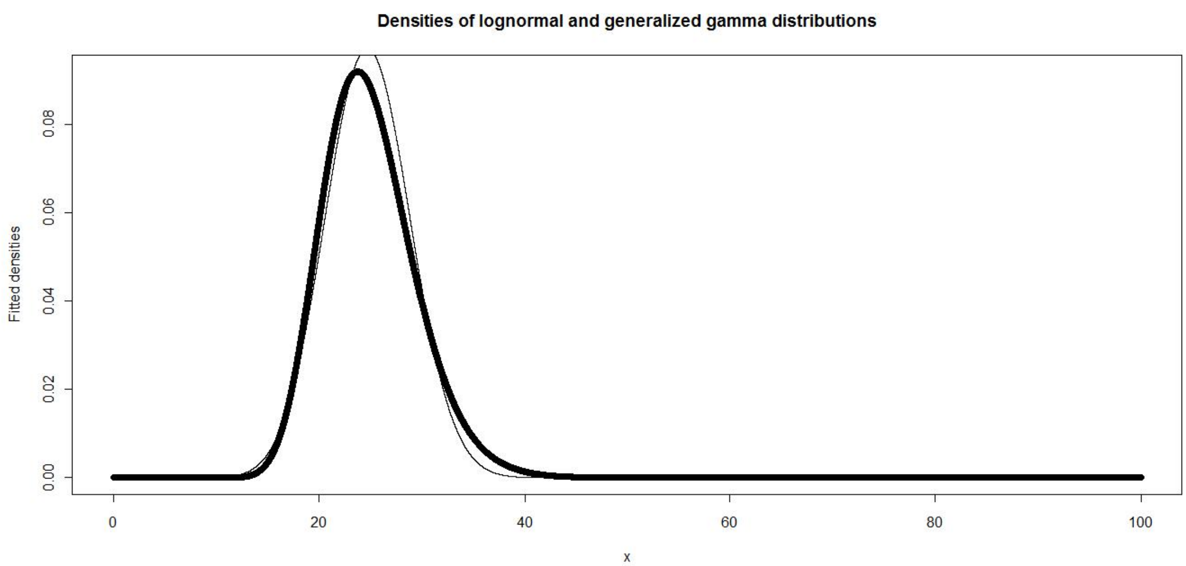

Figure 6.1: Fitted densities of lognormal distribution (represented by the bold line) and generalized gamma distribution (represented by the thin line)

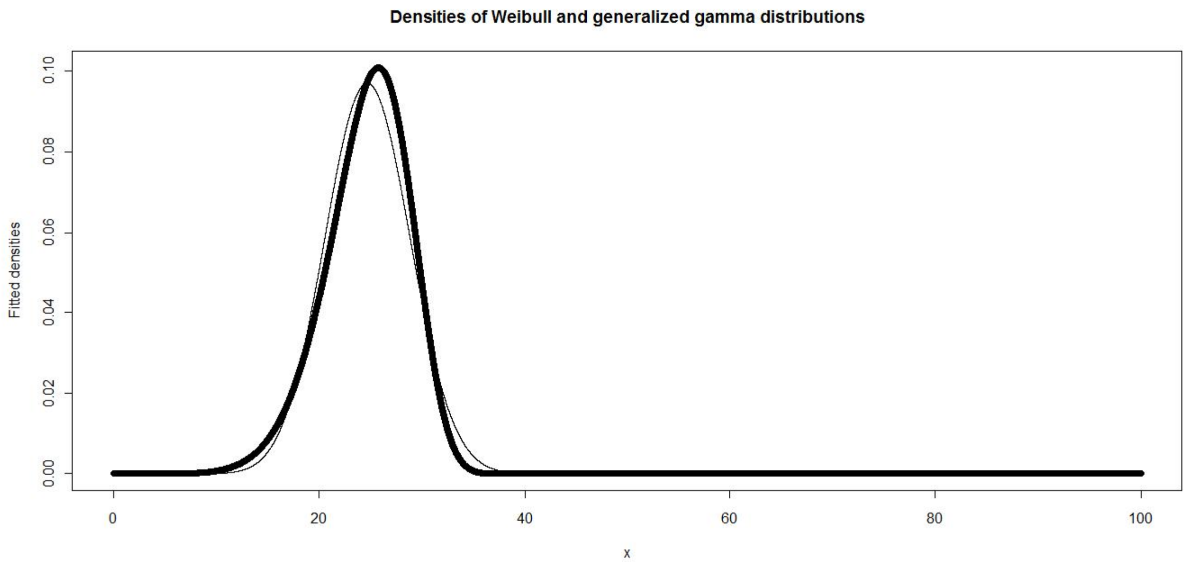

Figure 6.2: Fitted densities of Weibull distribution (represented by the bold line) and generalized gamma distribution (represented by the thin line)

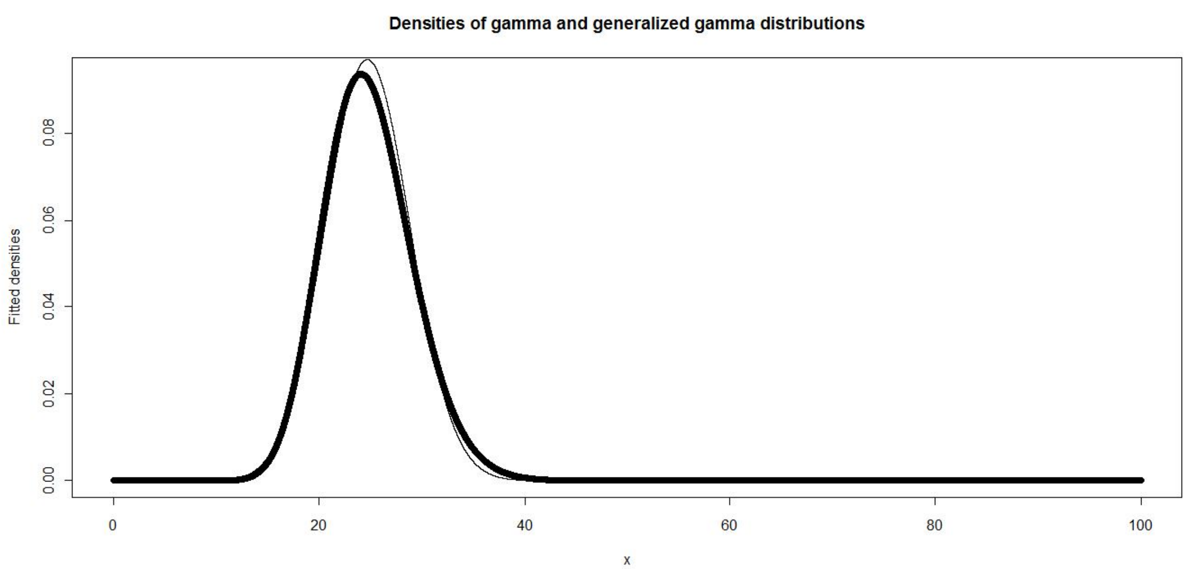

Figure 6.3: Fitted densities of gamma distribution (represented by the bold line) and generalized gamma distribution (represented by the thin line)

Chapter 7

Concluding remarks

7.1 Summary of research

In lifetime data analysis, left truncation and right censoring are quite common phenomena. Some important lifetime distributions used in practice are the lognormal, Weibull and gamma distributions, all of which are special cases of the generalized gamma distribution.

For the case of the lognormal distribution, the EM algorithm is developed here for estimating the model parameters based on left truncated and right censored data, after performing the log transformation. The maximization step is carried out by two different methods; one of them involves an approximation of the normal hazard function by Taylor series expansion, and the other is the known EM gradient algorithm which involves a one-step Newton-Raphson method. For comparison purpose, the MLEs of the parameters are also obtained by the Newton-Raphson method. It is observed through an extensive Monte Carlo simulation study that all the methods perform quite closely with respect to parameter estimation. The asymptotic variance-covariance matrix of the MLEs corresponding to the EM algorithm is derived by using the missing information principle; then, using these and the asymptotic normality of MLEs, the asymptotic confidence intervals for the parameters are constructed. Asymptotic variances of the MLEs are also obtained from the observed information matrix corresponding to the Newton-Raphson method, using which the corresponding asymptotic confidence intervals are obtained. The parametric bootstrap confidence intervals for the parameters are also constructed. Then, all these confidence intervals are compared in terms of coverage probabilities through an extensive Monte carlo simulation study. It is observed that the methods give close results. With the estimated model parameters, the predictive probability that a unit will work until a future year, given that it is right censored, is estimated. By using the delta method, the asymptotic confidence interval for this probability is also derived. All the methods of inference discussed are illustrated with a numerical example.

Next, necessary steps of the EM algorithm for estimating the model parameters of the Weibull distribution are developed, based on log-transformed left truncated and right censored data, thus taking the advantage of the extreme value distribution model. The maximization step is carried out by the EM gradient algorithm. For comparison purpose, the Newton-Raphson method is used once again, and in the Monte Carlo simulation studies, it is observed that the two methods of estimation perform quite closely. The asymptotic variance-covariance matrix of the MLEs are derived by using the missing information principle as before, and comparison in terms of coverage probabilities of the asymptotic confidence intervals of the parameters corresponding to the EM algorithm with the confidence intervals corresponding to the Newton-Raphson method in the simulation study shows that the two methods give close results, with slightly higher coverage probabilities for σ, the scale parameter of the extreme value distribution, based on the NR method. The predictive probability regarding the future lifetime of a censored unit is estimated and its asymptotic confidence interval is derived using the delta method as before. An illustrative example is given for all the methods of inference discussed.

The gamma distribution is another important lifetime distribution. The necessary steps of the EM algorithm are developed for fitting a gamma distribution to a left truncated and right censored data. For comparative purpose, the Newton-Raphson method is used once again. The asymptotic variance-covariance matrix of the MLEs are obtained by using the missing information principle under the EM framework, and also by using the observed information matrix corresponding to the Newton-Raphson method. Then, through an extensive Monte Carlo simulation study, all the methods of inference are examined. It is observed that they all give quite close results. The predictive probability of a censored unit's future lifetime is estimated, and its asymptotic confidence interval is obtained by using the delta method. All the methods of inference discussed are illustrated through a numerical example.

The generalized gamma distribution is a parsimonious model that includes the exponential, lognormal, Weibull and gamma distribution as special cases. The necessary steps of the EM algorithm for estimating the parameters of a generalized gamma distribution are developed based on left truncated and right censored data. The asymptotic variance-covariance matrix of the MLEs is derived by using the missing information principle, as before. Then, through a Monte Carlo simulation study, the proposed methods of inference are examined. The predictive probability regarding the future lifetime of a censored unit is estimated and an asymptotic confidence interval for this probability is provided by using the delta method. A numerical example is also presented.

For a given left truncated and right censored data, it would be of natural interest to know which of the above distributions provides a better fit. This model discrimination problem is dealt with through information criteria. Akaike's information criterion (AIC) and Bayesian information criterion (BIC) are used to select the model when the data come from one of the three distributions: lognormal, Weibull and gamma. It is observed in a Monte Carlo simulation study that these criteria choose the correct model among the three models with highest probability always, and that this probability increases with an increase in sample size. However, when the data come from generalized gamma distribution, AIC and BIC cannot distinguish between the models well, especially when the sample size is smaller. For larger samples, AIC picks generalized gamma as the true model to be fitted with highest probability among the four, but not so for BIC especially when the sample size is small. Based on our empirical study, we would recommend the use of AIC method for the purpose of model discrimination.

7.2 Future work

Here, some of the future directions are mentioned that could be considered as natural extensions of the the work in this thesis.

7.2.1 Models with covariates

An immediate extension of this work would be to consider models with inclusion of some pertinent covariates. For example, the failure of a unit may depend on several external factors; the failure of a power transformer may depend on factors such as heat, wind velocity, humidity etc. These factors can be included in the models as covariates with the use of some appropriate link functions, and the EM algorithm for estimating the model paramaters then can be developed for such models. These covariate-included models will certainly enhance the applicability of this method of inference.

7.2.2 Semiparametric modelling

Instead of looking at the probability density function, it will be possible to model a left truncated and right censored data through the hazard function. An appropriate baseline hazard function can be considered, and theh the related inference can be carried out. Modelling the hazard function this way leads to semiparametric inference which may be considered to be a more general form of analysis.

7.2.3 Other forms of censoring

Other forms of censoring, such as random censoring (through assuming a censoring distribution), progressive censoring (to account for loss of units), informative censoring, and so on, can be considered under this setup and then necessary inference may be developed.

7.2.4 Goodness-of-fit tests

After fitting a distribution to a left truncated and right censored data, one may want to know how good the fit of the model to data is. Formal statistical tests may be devised to examine the goodness-of-fit in such a situation.

Appendix A

Corresponding to Chapter 2

Table A.1 gives a simulated data from lognormal distribution with sample size 100, truncation percentage 60, and $(\mu,\ \sigma) = (3.5,\ 0.5)$. The illustrative example of lognormal distribution (in Chapter 2) are based on these data.

Table A.1: A simulated dataset, for sample size 100, truncation percentage 60, and the true parameter value of (μ, σ) as (3.5, 0.5). 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	1992	1	*	*	0	2.773	2.773
2	1984	1	*	*	0	3.178	3.178
3	1985	1	*	*	0	3.135	3.135
4	1988	1	*	*	0	2.996	2.996
5	1981	1	*	*	0	3.296	3.296
6	1984	1	*	*	0	3.178	3.178
7	1993	1	*	*	0	2.708	2.708
8	1984	1	*	*	0	3.178	3.178
9	1983	1	*	*	0	3.219	3.219

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
10	1995	1	*	*	0	2.565	2.565
11	1994	1	*	*	0	2.639	2.639
12	1993	1	*	*	0	2.708	2.708
13	1985	1	*	*	0	3.135	3.135
14	1980	1	*	*	0	3.332	3.332
15	1994	1	*	2008	1	2.639	*
16	1993	1	*	*	0	2.708	2.708
17	1988	1	*	*	0	2.996	2.996
18	1984	1	*	*	0	3.178	3.178
19	1989	1	*	2002	1	2.565	*
20	1984	1	*	*	0	3.178	3.178
21	1995	1	*	*	0	2.565	2.565
22	1984	1	*	2005	1	3.045	*
23	1991	1	*	*	0	2.833	2.833
24	1986	1	*	*	0	3.091	3.091
25	1981	1	*	2007	1	3.258	*
26	1988	1	*	*	0	2.996	2.996
27	1982	1	*	*	0	3.258	3.258
28	1980	1	*	*	0	3.332	3.332
29	1982	1	*	2003	1	3.045	*
30	1985	1	*	*	0	3.135	3.135
31	1980	1	*	1999	1	2.944	*
32	1986	1	*	*	0	3.091	3.091
33	1983	1	*	*	0	3.219	3.219
34	1983	1	*	*	0	3.219	3.219
35	1985	1	*	*	0	3.135	3.135
36	1984	1	*	*	0	3.178	3.178
37	1981	1	*	*	0	3.296	3.296
38	1984	1	*	1994	1	2.303	*
39	1983	1	*	*	0	3.219	3.219
40	1993	1	*	*	0	2.708	2.708
41	1962	0	2.890	1995	1	3.497	*
42	1960	0	2.996	*	0	3.871	3.871
43	1974	0	1.792	1992	1	2.890	*
44	1964	0	2.773	2007	1	3.761	*
45	1962	0	2.890	*	0	3.829	3.829
46	1963	0	2.833	1985	1	3.091	*

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
47	1960	0	2.996	*	0	3.871	3.871
48	1979	0	0.000	*	0	3.367	3.367
49	1962	0	2.890	2003	1	3.714	*
50	1962	0	2.890	1985	1	3.135	*
51	1967	0	2.565	1995	1	3.332	*
52	1963	0	2.833	1992	1	3.367	*
53	1963	0	2.833	1996	1	3.497	*
54	1967	0	2.565	*	0	3.714	3.714
55	1964	0	2.773	2004	1	3.689	*
56	1969	0	2.398	*	0	3.664	3.664
57	1960	0	2.996	*	0	3.871	3.871
58	1960	0	2.996	1982	1	3.091	*
59	1969	0	2.398	1985	1	2.773	*
60	1970	0	2.303	1984	1	2.639	*
61	1963	0	2.833	1983	1	2.996	*
62	1961	0	2.944	2006	1	3.807	*
63	1965	0	2.708	1990	1	3.219	*
64	1965	0	2.708	1987	1	3.091	*
65	1960	0	2.996	2007	1	3.850	*
66	1961	0	2.944	*	0	3.850	3.850
67	1960	0	2.996	*	0	3.871	3.871
68	1979	0	0.000	*	0	3.367	3.367
69	1970	0	2.303	*	0	3.638	3.638
70	1961	0	2.944	*	0	3.850	3.850
71	1963	0	2.833	1991	1	3.332	*
72	1963	0	2.833	*	0	3.807	3.807
73	1967	0	2.565	2001	1	3.526	*
74	1962	0	2.890	1990	1	3.332	*
75	1964	0	2.773	1999	1	3.555	*
76	1962	0	2.890	*	0	3.829	3.829
77	1961	0	2.944	2004	1	3.761	*
78	1962	0	2.890	*	0	3.829	3.829
79	1963	0	2.833	1986	1	3.135	*
80	1964	0	2.773	2005	1	3.714	*
81	1961	0	2.944	1990	1	3.367	*
82	1963	0	2.833	*	0	3.807	3.807
83	1961	0	2.944	1988	1	3.296	*

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
84	1960	0	2.996	1995	1	3.555	*
85	1972	0	2.079	1999	1	3.296	*
86	1960	0	2.996	1982	1	3.091	*
87	1962	0	2.890	2007	1	3.807	*
88	1977	0	1.099	2001	1	3.178	*
89	1963	0	2.833	1984	1	3.045	*
90	1960	0	2.996	2000	1	3.689	*
91	1974	0	1.792	*	0	3.526	3.526
92	1979	0	0.000	2003	1	3.178	*
93	1964	0	2.773	1995	1	3.434	*
94	1973	0	1.946	*	0	3.555	3.555
95	1960	0	2.996	1992	1	3.466	*
96	1964	0	2.773	2001	1	3.611	*
97	1960	0	2.996	1982	1	3.091	*
98	1961	0	2.944	1997	1	3.584	*
99	1963	0	2.833	1982	1	2.944	*
100	1961	0	2.944	*	0	3.850	3.850

Appendix B

Corresponding to Chapter 3

In Section 3.3, the M-step of the EM algorithm is carried out by a one-step Newton-Raphson method, using Eqs. (3.3.7) - (3.3.11). 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. (3.3.7) - (3.3.11), with $a = e^{\xi_i^{(k)}}$, $i = 1, ..., n$). Clearly, the first of the above three integrals is the upper incomplete gamma function $\Gamma(2, a)$. The other integrals can be evaluated by using the following lemma.

Lemma: We have

$$\textbf{(i)}\ \int_a^\infty w(\log w)e^{-w}dw = e^{-a}(a\log a + 1) + I_0(a),$$

$$\textbf{(ii)}\ \int_a^\infty w(\log w)^2e^{-w}dw = a(\log a)^2e^{-a} + I_1(a) + 2I_0(a),$$

where

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

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

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 point a. The mgf of Z is obtained to be

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

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

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

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

Using these, 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 B.1 below presents an example of left truncated and right censored data. All the results of the illustrative example in chapter 3 have been obtained based on these data.

Table B.1: A simulated dataset from Weibull distribution, for sample size 100, truncation percentage 40, and the true parameter value of (μ, σ) as (3.55, 0.33) wherein $*$ means not applicable. 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

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
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

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
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	*
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	*

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
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	*

Appendix C

Corresponding to Chapter 4

The first and second derivatives of the upper incomplete gamma function $\Gamma(\kappa, y/\theta)$ with respect to θ can be easily obtained by using Leibnitz's rule. For obtaining the derivatives with respect to κ, we use the following infinite series expansion of incomplete gamma function:

$$\Gamma(\kappa, y/\theta) = \Gamma(\kappa)\left[1 - e^{-y/\theta}\sum_{p=0}^{\infty}\frac{(y/\theta)^{\kappa+p}}{\Gamma(\kappa+p+1)}\right].$$

The first and second derivatives with respect to κ are obtained to be

$$\begin{aligned}\tfrac{\partial}{\partial\kappa}\Gamma(\kappa, y/\theta) = \;& \Big\{\Psi(\kappa)\Gamma(\kappa) + \Gamma(\kappa)\log(y/\theta)\Big\}\left[1 - e^{-y/\theta}\sum_{p=0}^{\infty}\frac{(y/\theta)^{\kappa+p}}{\Gamma(\kappa+p+1)}\right] - \Gamma(\kappa)\log(y/\theta)\\ &+\Gamma(\kappa)e^{-y/\theta}\sum_{p=0}^{\infty}\frac{(y/\theta)^{\kappa+p}\Psi(\kappa+p+1)}{\Gamma(\kappa+p+1)},\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2}{\partial \kappa^2}\Gamma(\kappa, y/\theta) = \; & \Big\{\Psi'(\kappa)\Gamma(\kappa) + \Psi^2(\kappa)\Gamma(\kappa)\Big\}\Bigg[1 - e^{-y/\theta}\sum_{p=0}^{\infty}\frac{(y/\theta)^{\kappa+p}}{\Gamma(\kappa+p+1)}\Bigg] \\
& -\{2\Psi(\kappa)\Gamma(\kappa)\log(y/\theta) + \Gamma(\kappa)\{\log(y/\theta)\}^2\}e^{-y/\theta}\sum_{p=0}^{\infty}\frac{(y/\theta)^{\kappa+p}}{\Gamma(\kappa+p+1)} \\
& +2\{\Gamma(\kappa)\Psi(\kappa) + \Gamma(\kappa)\log(y/\theta)\}e^{-y/\theta}\sum_{p=0}^{\infty}\frac{(y/\theta)^{\kappa+p}\Psi(\kappa+p+1)}{\Gamma(\kappa+p+1)} \\
& +\Gamma(\kappa)e^{-y/\theta}\Bigg[\sum_{p=0}^{\infty}\frac{(y/\theta)^{\kappa+p}\Psi'(\kappa+p+1)}{\Gamma(\kappa+p+1)} - \sum_{p=0}^{\infty}\frac{(y/\theta)^{\kappa+p}\Psi^2(\kappa+p+1)}{\Gamma(\kappa+p+1)}\Bigg].
\end{aligned}
$$

Table C.1 presents the left truncated and right censored data from the gamma distribution that are analyzed as an illustrative example in chapter 4.

Table C.1: A simulated dataset from gamma distribution, with sample size 100, truncation percentage 15, and the true parameter value of (κ, θ) as (5, 5).

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
1	1986	1	*	*	0	22	22
2	1987	1	*	2008	1	21	*
3	1980	1	*	1999	1	19	*
4	1995	1	*	*	0	13	*
5	1987	1	*	1997	1	10	*
6	1993	1	*	2006	1	13	*
7	1992	1	*	*	0	16	16
8	1993	1	*	*	0	15	15
9	1980	1	*	2006	1	26	*
10	1985	1	*	2007	1	22	*
11	1982	1	*	1995	1	13	*
12	1990	1	*	1998	1	8	*
13	1982	1	*	*	0	26	26
14	1983	1	*	2005	1	22	*
15	1984	1	*	*	0	24	24
16	1980	1	*	1998	1	18	*
17	1992	1	*	1998	1	6	*
18	1982	1	*	1996	1	14	*

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
19	1987	1	*	*	0	21	21
20	1982	1	*	*	0	26	26
21	1984	1	*	2007	1	23	*
22	1988	1	*	*	0	20	20
23	1980	1	*	2006	1	26	*
24	1993	1	*	*	0	15	15
25	1989	1	*	*	0	19	19
26	1987	1	*	*	0	21	21
27	1985	1	*	2002	1	17	*
28	1984	1	*	1989	1	5	*
29	1980	1	*	*	0	28	28
30	1980	1	*	*	0	28	28
31	1985	1	*	1999	1	14	*
32	1985	1	*	*	0	23	23
33	1982	1	*	*	0	26	26
34	1982	1	*	1996	1	14	*
35	1983	1	*	*	0	25	25
36	1980	1	*	2003	1	23	*
37	1980	1	*	1999	1	19	*
38	1982	1	*	*	0	26	26
39	1987	1	*	*	0	21	21
40	1988	1	*	*	0	20	20
41	1984	1	*	*	0	24	24
42	1986	1	*	*	0	22	22
43	1982	1	*	2008	1	26	*
44	1983	1	*	1999	1	16	*
45	1981	1	*	1994	1	13	*
46	1990	1	*	*	0	18	18
47	1983	1	*	2001	1	18	*
48	1983	1	*	1997	1	14	*
49	1993	1	*	*	0	15	15
50	1984	1	*	1999	1	15	*
51	1981	1	*	1999	1	18	*
51	1982	1	*	2000	1	18	*
52	1984	1	*	*	0	24	24
53	1989	1	*	2005	1	16	*

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
54	1986	1	*	2002	1	16	*
55	1981	1	*	*	0	27	27
56	1982	1	*	1998	1	16	*
57	1981	1	*	1990	1	9	*
58	1984	1	*	*	0	24	24
59	1982	1	*	2008	1	26	*
60	1983	1	*	*	0	25	25
61	1990	1	*	2003	1	13	*
62	1986	1	*	*	0	22	22
63	1985	1	*	*	0	23	23
64	1991	1	*	*	0	17	17
65	1984	1	*	*	0	24	24
66	1982	1	*	1991	1	9	*
67	1984	1	*	2002	1	18	*
68	1984	1	*	2008	1	24	*
69	1982	1	*	*	0	26	26
70	1983	1	*	*	0	25	25
71	1983	1	*	1998	1	15	*
72	1985	1	*	*	0	23	23
73	1984	1	*	1999	1	15	*
74	1994	1	*	*	0	14	14
75	1985	1	*	*	0	23	23
76	1989	1	*	*	0	19	19
77	1982	1	*	1995	1	13	*
78	1990	1	*	2005	1	15	*
79	1980	1	*	*	0	28	28
80	1993	1	*	*	0	15	15
81	1984	1	*	2008	1	24	*
82	1982	1	*	1995	1	13	*
83	1985	1	*	*	0	23	23
84	1994	1	*	*	0	14	14
85	1981	1	*	2006	1	25	*
86	1964	0	16	1983	1	19	*
87	1963	0	17	*	0	45	45
88	1962	0	18	1991	1	29	*
89	1963	0	17	1981	1	18	*

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
90	1964	0	16	2000	1	36	*
91	1962	0	18	1989	1	27	*
92	1961	0	19	2002	1	41	*
93	1974	0	6	1992	1	18	*
94	1961	0	19	1985	1	24	*
95	1977	0	3	*	0	31	31
96	1967	0	13	1989	1	22	*
97	1965	0	15	1990	1	25	*
98	1964	0	16	1987	1	23	*
99	1963	0	17	1986	1	23	*
100	1963	0	17	1980	1	17	*

Appendix D

Corresponding to Chapter 5

Table D.1 presents the left truncated and right censored data that are analyzed as an illustrative example in chapter 5.

Table D.1: A simulated dataset from generalized gamma distribution with sample size 200, truncation percentage 20, and the true parameter value of (α, η, κ) as (15, 3, 5), wherein $*$ means not applicable.

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
1	1992	1	*	*	0	16	16
2	1991	1	*	*	0	17	17
3	1980	1	*	2004	1	24	*
4	1990	1	*	*	0	18	18
5	1985	1	*	*	0	23	23
6	1988	1	*	*	0	20	20
7	1982	1	*	2008	1	26	*
8	1993	1	*	*	0	15	15
9	1994	1	*	*	0	14	14
10	1987	1	*	*	0	21	21

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
11	1982	1	*	*	0	26	26
12	1986	1	*	*	0	22	22
13	1983	1	*	2006	1	23	*
14	1989	1	*	*	0	19	19
15	1984	1	*	*	0	24	24
16	1984	1	*	2006	1	22	*
17	1982	1	*	*	0	26	26
18	1982	1	*	2007	1	25	*
19	1992	1	*	*	0	16	16
20	1990	1	*	*	0	18	18
21	1983	1	*	*	0	25	25
22	1981	1	*	2006	1	25	*
23	1981	1	*	2002	1	21	21
24	1990	1	*	*	0	18	18
25	1981	1	*	2007	1	26	*
26	1986	1	*	*	0	22	22
27	1992	1	*	*	0	16	16
28	1988	1	*	*	0	20	20
29	1988	1	*	*	0	20	20
30	1985	1	*	*	0	23	23
31	1984	1	*	2006	1	22	22
32	1995	1	*	*	0	13	13
33	1991	1	*	*	0	17	17
34	1981	1	*	2006	1	25	*
35	1980	1	*	2007	1	27	*
36	1981	1	*	2002	1	21	*
37	1990	1	*	*	0	18	18
38	1992	1	*	*	0	16	16
39	1995	1	*	*	0	13	13
40	1992	1	*	2003	1	11	*
41	1983	1	*	2007	1	24	*
42	1995	1	*	*	0	13	13
43	1984	1	*	*	0	24	24
44	1981	1	*	*	0	27	27
45	1985	1	*	*	0	23	23
46	1991	1	*	*	0	17	17
47	1988	1	*	*	0	20	20
48	1981	1	*	2004	1	23	*

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
49	1984	1	*	*	0	24	24
50	1982	1	*	2003	1	21	*
51	1994	1	*	*	0	14	14
52	1985	1	*	*	0	23	23
53	1980	1	*	2005	1	25	*
54	1990	1	*	2007	1	17	*
55	1983	1	*	*	0	25	25
56	1983	1	*	*	0	25	25
57	1984	1	*	2007	1	23	*
58	1980	1	*	2005	1	25	*
59	1980	1	*	2004	1	24	*
60	1994	1	*	*	0	14	14
61	1984	1	*	2008	1	24	*
62	1993	1	*	*	0	15	15
63	1986	1	*	*	0	22	22
64	1982	1	*	2006	1	24	*
65	1982	1	*	2007	1	25	*
66	1982	1	*	*	0	26	26
67	1985	1	*	*	0	23	23
68	1984	1	*	*	0	24	24
69	1980	1	*	2003	1	23	*
70	1991	1	*	*	0	17	17
71	1993	1	*	*	0	15	15
72	1982	1	*	*	0	26	26
73	1993	1	*	*	0	15	15
74	1992	1	*	*	0	16	16
75	1984	1	*	2001	1	17	*
76	1989	1	*	*	0	19	19
77	1989	1	*	*	0	19	19
78	1986	1	*	2003	1	17	*
79	1988	1	*	*	0	20	20
80	1987	1	*	*	0	21	21
81	1981	1	*	2003	1	22	*
82	1992	1	*	*	0	16	16
83	1981	1	*	2005	1	24	*
84	1989	1	*	*	0	19	19
85	1984	1	*	*	0	24	24
86	1986	1	*	*	0	22	22

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
87	1994	1	*	*	0	14	14
88	1991	1	*	*	0	17	17
89	1987	1	*	2007	1	20	*
90	1980	1	*	2001	1	21	*
91	1988	1	*	*	0	20	20
92	1994	1	*	*	0	14	14
93	1987	1	*	*	0	21	21
94	1984	1	*	*	0	24	24
95	1989	1	*	2008	1	19	19
96	1981	1	*	*	0	27	27
97	1984	1	*	2006	1	22	*
98	1983	1	*	2007	1	24	*
99	1982	1	*	*	0	26	26
100	1983	1	*	2007	1	24	*
101	1984	1	*	*	0	24	24
102	1983	1	*	*	0	25	25
103	1991	1	*	*	0	17	17
104	1985	1	*	*	0	23	23
105	1983	1	*	2002	1	19	*
106	1989	1	*	*	0	19	19
107	1988	1	*	*	0	20	20
108	1993	1	*	*	0	15	15
109	1994	1	*	*	0	14	14
110	1982	1	*	2007	1	25	*
111	1981	1	*	2003	1	22	*
112	1993	1	*	*	0	15	15
113	1982	1	*	2004	1	22	*
114	1991	1	*	*	0	17	17
115	1981	1	*	2007	1	26	*
116	1983	1	*	2008	1	25	*
117	1987	1	*	*	0	21	21
118	1982	1	*	2004	1	22	*
119	1980	1	*	*	0	28	28
120	1995	1	*	*	0	13	13
121	1984	1	*	*	0	24	24
122	1985	1	*	*	0	23	23
123	1987	1	*	*	0	21	21
124	1980	1	*	2005	1	25	*

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
125	1992	1	*	*	0	16	16
126	1988	1	*	*	0	20	20
127	1982	1	*	*	0	26	26
128	1992	1	*	*	0	16	16
129	1982	1	*	2005	1	23	*
130	1981	1	*	2001	1	20	*
131	1984	1	*	2008	1	24	*
132	1981	1	*	1998	1	17	*
133	1980	1	*	2003	1	23	*
134	1988	1	*	2005	1	17	*
135	1991	1	*	*	0	17	17
136	1989	1	*	2006	1	17	*
137	1989	1	*	*	0	19	19
138	1985	1	*	2007	1	22	*
139	1986	1	*	2007	1	21	*
140	1980	1	*	2005	1	25	*
141	1988	1	*	*	0	20	20
142	1985	1	*	*	0	23	23
143	1984	1	*	2008	1	24	*
144	1994	1	*	*	0	14	14
145	1993	1	*	*	0	15	15
146	1984	1	*	2008	1	24	*
147	1982	1	*	2006	1	24	*
148	1982	1	*	2008	1	26	*
149	1983	1	*	*	0	25	25
150	1990	1	*	*	0	18	18
151	1995	1	*	*	0	13	13
152	1993	1	*	*	0	15	15
153	1982	1	*	2004	1	22	*
154	1984	1	*	2006	1	22	*
155	1987	1	*	*	0	21	21
156	1982	1	*	2008	1	26	*
157	1988	1	*	*	0	20	20
158	1989	1	*	*	0	19	19
159	1988	1	*	*	0	20	20
160	1980	1	*	2007	1	27	*
161	1961	0	19	1987	1	26	*
162	1960	0	20	1985	1	25	*

Serial No.	Installation year	Truncation indicator	Truncation time	Failure year	Censoring indicator	Lifetime	Censoring time
163	1968	0	12	1989	1	21	*
164	1977	0	3	1998	1	21	*
165	1965	0	15	1990	1	25	*
166	1964	0	16	1993	1	29	*
167	1961	0	19	1983	1	22	*
168	1977	0	3	2002	1	25	*
169	1961	0	19	1985	1	24	*
170	1961	0	19	1983	1	22	*
171	1971	0	9	1991	1	20	*
172	1961	0	19	1992	1	31	*
173	1966	0	14	1991	1	25	*
174	1976	0	4	2006	1	30	*
175	1979	0	1	2000	1	21	*
176	1960	0	20	1989	1	29	*
177	1960	0	20	1983	1	23	*
178	1961	0	19	1982	1	21	*
179	1962	0	18	1987	1	25	*
180	1962	0	18	1991	1	29	*
181	1973	0	7	1994	1	21	*
182	1961	0	19	1986	1	25	*
183	1979	0	1	2003	1	24	*
184	1960	0	20	1989	1	29	*
185	1977	0	3	2005	1	28	*
186	1963	0	17	1988	1	25	*
187	1964	0	16	1991	1	27	*
188	1965	0	15	1993	1	28	*
189	1960	0	20	1992	1	32	*
190	1977	0	3	1997	1	20	*
191	1962	0	18	1988	1	26	*
192	1963	0	17	1986	1	23	*
193	1974	0	6	1997	1	23	*
194	1979	0	1	2006	1	27	*
195	1962	0	18	1991	1	29	*
196	1970	0	10	1989	1	19	*
197	1974	0	6	2005	1	31	*
198	1960	0	20	1987	1	27	*
199	1977	0	3	2004	1	27	*
200	1964	0	16	1992	1	28	*

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatics Control*, 19:716 – 723.

Balakrishnan, N. (2003). Progressive censoring methodology: An appraisal (with discussions). *Test*, 16:211 – 296.

Balakrishnan, N. and Aggarwala, R. (2000). *Progressive Censoring: Theory, Methods, and Applications*. Birkhäuser, Boston.

Balakrishnan, N. and Caroni, C. (2013). *An Introduction to Lifetime Data Analysis*. Birkhäuser, Boston (to appear).

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

Balakrishnan, N. and Kundu, D. (2013). Hybrid censoring: Models, inferential results and applications. *Computational Statistics & Data Analysis*, 57:166 – 209.

Balakrishnan, N., Kundu, D., Ng, H. K. T., and 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. and Peng, Y. (2006). Generalized gamma frailty model. *Statistics in Medicine*, 25:2797 – 2816.

Bowman, K. O. and Shenton, L. R. (1988). *Properties of Estimators for the Gamma Distribution*. Marcel Dekker, New York.

Childs, A., Chandrasekar, B., Balakrishnan, N., and Kundu, D. (2003). Exact likelihood inference based on Type-I and Type-II hybrid censored samples from the exponential distribution. *Annals of the Institute of Statistical Mathematics*, 55:319 – 330.

Cohen, A. C. (1991). *Truncated and Censored Samples*. Marcel Dekker, New York.

Cohen, A. C. and Whitten, B. J. (1988). *Parameter Estimation in Reliability and Life Span Models*. Marcel Dekker, New York.

Crow, E. L. and Shimizu, K. (1988). *Lognormal Distributions: Theory and Applications*. Marcel Dekker, New York.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1 – 38.

Geddes, K. O., Glasser, M. L., Moore, R. A., and Scott, T. C. (1990). Evaluation of classes of definite integrals involving elementary functions via differentiation of special functions. *Applicable Algebra in Engineering, Communication and Computing*, 1:149 – 165.

Hong, Y., Meeker, W. Q., and 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., and Balakrishnan, N. (1994). *Continuous Univariate Distributions - Vol. 1, Second Edition*. 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. and Krishnan, T. (2008). *The EM Algorithm and Extensions, Second Edition*. John Wiley & Sons, Hoboken, New Jersey.

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

Schwarz, E. D. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461 – 464.

Sen, P. K. and Singer, J. M. (1994). *Large Sample Methods in Statistics. An Introduction with Applications*. Chapman & Hall, New York - London.