PAN AFRICAN UNIVERSITY
INSTITUTE FOR BASIC SCIENCES, TECHNOLOGY AND INNOVATION

MSc RESEARCH THESIS

MSc MATHEMATICS: STATISTICS OPTION

Maximum Likelihood Estimation of the parameters of Exponentiated Generalized Weibull Distribution based on progressive Type II censored data

Author:

Ibrahim SAWADOGO

Supervisors:

1. Prof Leo ODONGO
2. Dr Ibrahim LY

A Research Thesis submitted to Pan African University Institute of Science, Technology and Innovation in partial fulfillment of the requirement for the award of the degree of Master of Science in Mathematics of the Pan African University.

February 2018

Declaration

I, Ibrahim SAWADOGO, declare that this research thesis is my own, and has not been submitted to any other university for examination.

Signature . . Date

Name: Ibrahim SAWADOGO

Pan African University Institute of Science, Technology and Innovation, Nairobi, Kenya.

This research thesis has been submitted for examination with my approval as University Supervisor.

Signature . . Date

Name: Prof Leo ODONGO

Department of Statistics and Actuarial Science, Kenyatta University (KU), Nairobi, Kenya.

This research thesis has been submitted for examination with my approval as University Supervisor.

Signature . . Date

Name: Dr Ibrahim LY

Department of Mathematics, University of Ouagadougou, Ouagadougou, Burkina Faso.

Department of Mathematics, University of Potsdam, Potsdam, Germany.

Acknowledgement

First of all I would like to thank the Almighty God for the gift of life, for granting me the knowledge to carry out this study. I thank the African Union for this opportunity. I would like to thank my supervisors Professor Leo Odongo and Doctor Ibrahim LY for guiding me during the preparation of this thesis.

Dedication

To my Mother.

”Black woman, woman of Africa, O my mother, let me thank you; thank you for all that you have done for me, your son, who, though so far away, is still so close to you!”

Camara Laye

Abstract

The Exponentiated Generalized Weibull distribution is a probability distribution which generalizes the Weibull distribution, introducing two more shapes parameters to best adjust the non-monotonic failure rate. The distribution was derived by Oguntunde et al. in 2015 based on Codeiro et al.'s paper on the exponentiated generalized class of distribution. The parameters of the new probability distribution function are estimated by the maximum likelihood method under progressive type II censored data via Expectation Maximization (EM) algorithm. The performance of estimators are investigated using the Root Mean Square Error RMSE based on simulation for various degrees of censoring and sample sizes. Application to real data is included. It is observed that RMSE decreases with increasing sample size, and also with decreasing censored sample size.

Contents

List of Tables

List of Notations/Abbreviations

cdf	Cumulative distribution function
$f(x)$	The pdf of the distribution under consideration
$F(x)$	The cdf of the distribution under consideration
$L(x)$	The likelihood function
n	Sample size
m	Observed sample size
$n-m$	Censoring size
Y	Observed sample
Z	Censored sample
$X=(Y,Z)$	Complete sample
$S(x)$	Survival function
$h(x)$	Hazard function
$\Gamma(.)$	Gamma function
EGW	Exponentiated Generalized Weibull distribution
GEV	Generalized Extreme Value
EM	Expectation Maximization
GEM	Generalized Expectation Maximization
MLE	Maximum Likelihood Estimation
MSE	Mean Square Error
$RMSE$	Root Mean Square Error

Chapter 1

Introduction

This chapter gives the background of the study by introducing the Weibull distribution, the statement of the problem, justification of the study, the general and specific objectives, scope of the study, significance of the study and definitions of key terms.

1.1 Background of study

The Weibull distribution, named after Waloddi Weibull (1951), is a continuous probability distribution. The Weibull distribution is a special case of the Generalized Extreme Value (GEV) distribution in the same way as the Gumbel distribution or the Fréchet distribution. The probability density function of the Weibull distribution with two parameters say α (shape parameter) and β (scale parameter) is given by:

$$f(x;\alpha,\beta) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^{\alpha}} \tag{1.1}$$

where $x > 0$, $\alpha > 0$, $\beta > 0$.

The term Weibull distribution covers a whole family of distributions, some of them appearing in physics as a consequence of certain hypotheses. This in particular is the case of the exponential distribution and the Rayleigh distribution. In fact by letting $\alpha = 1$ and $\beta = \dfrac{1}{\lambda}$ in (1.1) we get the exponential distribution density function

$$f(x; \lambda) = \lambda e^{-(\lambda x)} \tag{1.2}$$

Likewise if $\alpha = 2$ and $\beta = b\sqrt{2}$ we obtain the Rayleigh distribution density function

$$f(x; b) = \frac{x}{b^2} e^{-(x^2/2b^2)} \tag{1.3}$$

The Weibull distribution is often used in the field of lifetime analysis, thanks to its flexibility. In order to increase that flexibility many authors have proposed several generalisations of the Weibull distribution, like Mudholkar and Srivastava (1993) , Bourguignon et al. (2014), Alshawarbeh (2011), and also Oguntunde et al. (2015) who introduced the Exponentiated Generalized Weibull (EGW) distribution.

1.2 Problem statement

Over the past years Weibull distribution has been used extensively to model survival data due to its flexibility. However, some of the data that arise in survival analysis require distributions with non-monotonic failure rate, see for example Lai et al (2001). A case in point is the human mortality cycle that exhibits a bathtub failure rate which is non-monotonic. Weibull distribution is not useful for modeling phenomenon with non-monotone failure rate, see for example Mahmoudi and Sepahdar (2013). This phenemenon can be modelled adequately by the Exponentiated Generalized Weibull (EGW) distribution developed by Oguntunde et al. (2015).

Procedure of survival analysis past any time-point usually becomes complicated due to the presence of censored observations. Ignoring these censored observations results in loss of potentially valuable information. It is therefore important to carry out a full survival analysis which will incorporate all the information contained in these censored observations.

There is limited work on the maximum likelihood estimation for the parameters of the Exponentiated Generalized Weibull under progressive type II censoring.

Therefore in this study we develop maximum likelihood estimators for the parameters of the Exponentiated Generalized Weibull (EGW) distribution introduced by Oguntunde et al. (2015) under progressive type II censored data.

1.3 Justification of the study

In Oguntunde et al. (2015), the Exponentiated Generalized Weibull distribution has been defined and also some mathematical properties of the new distribution have been investigated such as: the limiting behaviour of the probability density function and the cumulative distribution function, the reliability analysis, expression of the moments, the quantile function and the order statistics. These were done for full sample with no censoring. Up to now nobody has attempted to obtain the Maximum Likelihood Estimators of the Exponentiated Generalized Weibull (EGW) distribution under progressive type II censored data.

1.4 Objective of the study

1.4.1 General objective

The general objective of this study was to develop maximum likelihood estimators for the parameters of the Exponentiated Generalized Weibull distribution under progressive type II censored data.

1.4.2 Specific objectives

In order to achieve the main objective stated above we have the following specific objectives:

1. Derive the maximum likelihood estimators of the parameters of the Exponentiated Generalized Weibull distribution under progressive type II censoring via EM algorithm.
2. Investigate the performance of the MLEs based on simulation for various degrees of censoring and sample sizes using RMSE.
3. Apply the estimation procedure on real data.

1.5 The Scope of the study

The study only considered the maximum likelihood estimation of the Exponentiated Generalized Weibull based on progressive type II censored data. In other words statistical inference on the parameters was not considered.

1.6 Significance of the study

The results of this study will go a long way in making Exponentiated Generalized Weibull (EGW) to be applicable in situation where progressive type II censoring data are encountered, thus contributing to knowledge in the are of survival and reliability analysis.

1.7 Definition of key terms

1.7.1 Lifetime analysis

A lifetime is defined as a positive random variable T, generally the time elapsing between two events.
Examples of events: death, breakdown, entry into unemployment, illness.
The Lifetime analysis is the study of the delay of the occurrence of the event under study.

1.7.2 Censored data

It is data some of which are only known with a lower or upper bound and not a precise value.

1.7.3 Type I censored data

The so-called Type I censoring describes the situation where a test ends at a certain period, and we know that the remaining objects (individuals) have not yet failed.
For example, we start with 100 light bulbs and end the experience after a certain time.
In this case, the censored time is often fixed, and the number of failed objects (individuals) is a random variable.

1.7.4 Type II censored data

The experiment continues until a fixed proportion of objects (individuals) has failed.
For example, we stop the experiment after exactly 50 light bulbs have failed.
In this case, the number of failed objects (individuals) is fixed, and time is a random variable.

1.7.5 Progressive type II censored data

This is a type II censoring in which at each failure a certain number of objects are removed from the remaining data set.

1.7.6 Scale parameter

In Probability and Statistics Theory, a scale parameter is a parameter that governs the flattening of a parametric family of probability distributions. It is mainly a multiplicative factor.

1.7.7 Shape parameter

A shape parameter is a parameter of a probability distribution that is neither a position parameter nor a scale parameter. Such a parameter governs only the shape of the distribution.

1.7.8 Failure rate

The failure rate is an expression of the reliability of equipment and each of their components. It is also called hazard rate or hazard function. It is the frequency at which a component or system fails.

1.7.9 Bathtub shape

That is a shape that has steep sides with a flat bottom, describing a curve that has three parts; decreasing part, constant part and increasing part.

1.7.10 Baseline (parent) distribution

It is an existing distribution that is being generalized or modified. Here our baseline distribution is the Weibull distribution.

Chapter 2

Literature Review

This chapter focuses on reviewing the work done by previous researchers that are relevant to the problem of study. The main goal of this chapter is to offer an overall view on the approaches developed so far in the estimation of the parameters of generalized distributions of the Weibull distribution based on censored data or complete data. This helps to gain an insight of our research while avoiding repetition of the work and mistakes already done by others.

2.1 The Weibull distribution

Various probability density functions have been proposed to perform statistical analysis of lifetime data. The Weibull distribution is one of the most widely used distributions in the analysis of lifetimes data. It was introduced by the French Mathematicians Maurice René Fréchet (1928). Indeed in the 1920s Fréchet developed a distribution to which he gave his name; Fréchet distribution, as an extreme value distribution. This distribution is in fact equal to the reciprocal of the Weibull random variable. The work of Fréchet was used by Rosin (1933). The latter applied it to describe the particle size distribution generated by grinding, milling and crushing operations of materials. This probability distribution has been widely used as a probabilistic model in studies on lifetimes. Mudholkar and Srivastava (1993) introduced the exponentiated Weibull to analyse bathtub failure rate data which we know cannot be handled well by the regular Weibull for its monotonicity. Based on Type-I and Type-II generalized progressive hybrid censoring schemes, Mudholkar and Srivastava derived the maximum likelihood estimators and Bayes estimators for the unknown parameters of exponentiated Weibull lifetime model. The approximate asymptotic variance-covariance matrix and approximate confidence intervals based on the asymptotic normality of the classical estimators were obtained. Independent non-informative types of priors are considered for the unknown parameters to develop the Bayes estimators and corresponding Bayes risks under a squared error loss function. Proposed estimators cannot be expressed in closed forms and can be evaluated numerically by some suitable iterative procedure. Finally, one real data set is analyzed for illustrative purposes. Also Zhang and Xie (2011) worked on bathtub failure data using the truncated Weibull. The characteristics and application of the truncated Weibull distribution were studied in his paper. The distribution is applicable to the situation where the test data are bounded in an interval because of test conditions, cost and other restrictions. An important property of the truncated Weibull distribution is that it can have bathtub-shaped failure rate function. In his paper, the parametric analysis and parameter estimation methods of the distribution were investigated. Both the graphical approach and the maximum likelihood estimation were considered. The applicability of the distribution to modeling lifetime data was illustrated by an example and the results of comparisons to other competitive models in modeling the given data were also presented. Moreover, the possible application of the distribution to modeling component or system failure was discussed. Ghnimi and Gasmi (2014) have given estimation of the parameters of the exponentiated Weibull and the additive Weibull, which are two specific generalization of the Weibull. Their paper gives a study on the performance of two specific modifications of the Weibull distribution which are the exponentiated Weibull distribution and the additive Weibull distribution. These shows how the Weibull distribution has been used since it has been proposed by the Swedish engineer and mathematician Ernst Hjalmar Waloddi Weibull (1887-1979), the Weibull distribution is a probability distribution that is widely used to model lifetime data. Because of its flexibility, some modifications of the Weibull distribution have been made from several researches in order to best adjust the non-monotonic shapes.

2.2 The Exponentiated Generalized Weibull distribution

Cordeiro, et al. (2013) introduced the exponentiated generalized class of distribution which is more general than the two classes of Lehmann (1953) alternative, it is a combination of the Lehmann type I and type II alternatives. Indeed, for any baseline (or parent) distribution it is possible to define the corresponding Exponentiated Generalized family of distribution. Cordeiro, et al. (2013) discussed four special models namely the Exponentiated Generalized Fréchet, the Exponentiated Generalized Normal, the Exponentiated Generalized Gamma and the Exponentiated Generalized Gumbel. They proposed a new method of adding two parameters to a continuous distribution that extends the idea first introduced by Lehmann and studied by Nadarajah and Kotz (2006). This method leads to a new class of exponentiated generalized distributions that can be interpreted as a double construction of Lehmann alternatives. Some special models are dis- cussed. We derive some mathematical properties of this class including the ordinary moments, generating function, mean deviations and order statis- tics. Maximum likelihood estimation is investigated and four applications to real data are presented. In Nadarajah and Kotz (2016) paper, they introduced four more exponentiated type distributions that generalize the standard gamma, standard Weibull, standard Gumbel and the standard Fréchet distributions in the same way the exponentiated exponential distribution generalizes the standard exponential distribution. A treatment of the mathematical properties is provided for each distribution. Oguntunde et al. (2015) have discussed the special case of the Exponentiated Generalized Weibull distribution by using the Weibull distribution as baseline distribution. In his article, a generalization of the Weibull distribution is being studied in some details. The new model is referred to as the Exponentiated Generalized Weibull distribution. The aim is to increase the flexibility of the Weibull distribution. Methods: The concepts introduced in the Exponentiated Generalized family of distributions due to Cordeiro were employed. Findings: Some basic mathematical properties of the resulting model were identified and studied in minute details. Meanwhile, estimation of model parameters was performed using the maximum likelihood method. Application/Improvement: The Exponentiated Generalized Weibull distribution was presented as a competitive model that would be useful in modeling real life situations with inverted bathtub failure rates. The proposed distribution has four parameters (three shape parameters and one scale parameter). The work of Oguntunde et al. (2015) is limited on the mathematics properties of the distribution like the moments, the limiting behaviour of the functions (pdf and cdf), the reliability analysis, and the quantile function. However Oguntunde et al. (2015) did not derive the estimation of the distribution parameters under censored data. In this study the MLEs of the new distribution will be derived using progressive type II censored data via EM algorithm.

2.3 The progressive type II censoring

Several researchers have worked on progressive censoring, see for example Herd (1956), Cohen (1966), Mann (1969), and Thomas and Wilson (1972). In his PhD thesis Herd defined what we call progressive type II censored data. The method of maximum likelihood is employed to estimate the parameters for the exponential, the normal, and the gamma distributions. These estimates are, in certain cases, difficult to obtain. They require iteration; therefore, certain practical limitations exist for their use. A new method of solving the likelihood equations for the normal distribution is introduced, and a delta function is tabulated to facilitate the solution. An extension of the censorship procedure to another general type is considered for estimation by the method of maximum likelihood. The non-parametric estimate of the probability of surviving (quantiles) is obtained, and a general method of 85 estimation based on the quantiles is presented, which will yield reasonable results when the method of maximum likelihood cannot be used, and which will be reasonably efficient in comparison to the maximum likelihood estimates when these are available for comparison. It is shown that the method of estimation from the quantiles yields the maximum-likelihood estimate for the exponential distribution for all rules of censorship and the uniform distribution for a random sample. The quantile method is asymptotically equivalent to the methods of maximum likelihood for the parameters of the normal distribution. The method yields a simple result (best linear unbiased estimate) for the uniform distribution with single or multi-censorship. This is an advantage over the maximum-likelihood method, which does not furnish a simple result. The results are illustrated by a number of examples taken from industrial experiments. It is possible, through the techniques presented, to utilize small samples such as exist in industry, and also, although curtailment exists, to have assurance of a certain number of complete ”life times” from which to make estimates even where no prior knowledge –other than the distributional form – exists on the ”life times” of the items tested. Montanari and Cacciari (1988) illustrated an application of progressive Type-II censoring on aging tests on solid insulating materials. The problem of evaluating the time-to-failure percentiles in progressively-censored tests on solid insulating materials is addressed in their paper. Statistical methods to estimate the parameters of the Weibull distribution (and their confidence limits) are examined on the basis of the results of aging with combined thermal-electrical stresses carried out on XLPE insulated cable models. These tests are performed at the same stresses on samples more than 1 m long and subjected to progressive censoring of aging times, or on short specimens about 20 cm long and subjected to complete, or singly-censored, life tests. This procedure allows the effectiveness of progressively-censored tests in estimating life percentiles to be verified, and the accuracy of the methods to be compared.

Balakrishnan and Aggarwala (2000) have excellent text on progressive censoring. The estimation of parameters from lifetime distributions based on progressive Type-II censoring has been studied by several authors among them Balakrishnan and Kannan (2001), Ng, et al. (2002), Mousa and Jaheen (2002), Balakrishnan, et al. (2003) and Soliman (2005). Balakrishnan and Kannan used the logistic distribution in their paper. The logistic distribution has been widely used as a growth model in many problems. The logistic model has often been selected as an alternative to the normal because of the similarity of the two distributions. In their article, they considered the estimation of the location and scale parameters of the logistic distribution based on progressively Type-II censored samples. The maximum likelihood method yielded equations that do not provide explicit solutions under any censoring scheme. They used an approximation of the cumulative distribution function (cdf) that leads to simplified likelihood equations which yielded explicit estimators. They examined numerically the bias and mean squared error (MSE) of the maximum likelihood estimators (MLEs) and the approximate estimators and showed that the approximation provides estimators that are almost as efficient as the MLEs. The probability coverages of the pivotal quantities (for location and scale parameters) based on asymptotic normality are shown to be unsatisfactory, especially when the effective sample size is small. They suggested the use of unconditional simulated percentage points for the construction of confidence intervals. They also developed estimators based on weighted least squares; these estimators may be used as effective starting values for the numerical algorithm to determine the MLEs, and are themselves quite efficient when the effective sample size is large. A wide range of sample sizes and progressive censoring schemes have been considered in this study. Finally, they present a numerical example to illustrate the methods of inference developed here. In this study MLE's are derived for the EGW distribution proposed by Oguntunde et al. (2015) under progressive type II censored data.

Chapter 3

Methodology

The methods of estimation of the parameters most used are the method of the maximum likelihood estimator (MLE) and the method of the moments (MM). The first is used generally because of its very interesting asymptotic properties, the second for its simplicity. As far as the Weibull distribution is concerned, other methods have been proposed, in particular graphical estimation methods and methods based on order statistics. These include the work of Dubey (1967) and Kappenman (1985). In addition, some authors proposed modifications to the classical methods of maximum likelihood estimator and moments Cohen and Whitten (1982). We consider in this thesis the method of maximum likelihood for its well-known asymptotic properties Lehmann (1983). The asymptotic properties of estimators derived from the maximum likelihood method are well known. In particular, these estimators are consistent, asymptotically unbiased and best asymptotically normal Lehmann (1983).

3.1 Exponentiated Generalized Weibull distribution

The Exponentiated Generalized Weibull distribution has been proposed by Oguntunde et al. (2015) using the Weibull distribution as a baseline distribution in the expontiated generalized class of distribution introduced by Cordeiro et al. (2013). The cumulative distribution function and the survival function of the Exponentiated Generalized Weibull are respectively given by:

$$F(x; a, b, \alpha, \beta) = \left[1 - \left\{e^{-(x/\beta)^{\alpha}}\right\}^{a}\right]^{b} \tag{3.1}$$

and

$$\begin{aligned} S(x; a, b, \alpha, \beta) =& 1 - F(x; a, b, \alpha, \beta) \\ =& 1 - \left[1 - \left\{e^{-(x/\beta)^{\alpha}}\right\}^{a}\right]^{b} \end{aligned} \tag{3.2}$$

where $x > 0$, $a > 0$, $b > 0$, $\alpha > 0$, $\beta > 0$.

The quantile function is given by

$$q(y; a, b, \alpha, \beta) = \beta\left(-\frac{1}{a}\log(1 - y^{1/b})\right)^{1/\alpha} \tag{3.3}$$

where $y > 0$, $a > 0$, $b > 0$, $\alpha > 0$, $\beta > 0$.

The Exponentiated Generalized Weibull generalizes the following distributions:

For $a = 1$, Generalized Weibull;

For $b = 1$, Exponentiated Weibull;

For $a = b = 1$, Weibull distribution;

For $a = b = \alpha$ =1, Exponential distribution.

The probability density function and the hazard function have respectively the following expressions:

$$f(x;a,b,\alpha,\beta) = ab\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\left\{e^{-(x/\beta)^{\alpha}}\right\}^{a}\left[1-\left\{e^{-(x/\beta)^{\alpha}}\right\}^{a}\right]^{b-1} \tag{3.4}$$

and

$$h(x;a,b,\alpha,\beta) = \frac{ab\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\left\{e^{-(x/\beta)^{\alpha}}\right\}^{a}\left[1-\left\{e^{-(x/\beta)^{\alpha}}\right\}^{a}\right]^{b-1}}{1-\left[1-\left\{e^{-(x/\beta)^{\alpha}}\right\}^{a}\right]^{b}} \tag{3.5}$$

3.2 Maximum likelihood estimation

Let us recall the principle of maximum likelihood estimation (MLE).

We have n observations, considered as the realizations of n independent and identically distributed random variables $(X_1, X_2, \ldots, X_n)$.

In the framework of a parametric statistical model $(\mathbb{P}(\theta))_{\theta\in\Theta}$, each X_i $i = 1, \ldots, n$ follows a distribution governed by the vector parameter $\theta \in \mathbb{R}^d$. For example, the X_i's can be identically independently normally distributed, $N(\mu, \sigma^2)$: then $\theta = (\mu, \sigma^2)$, and $\Theta = \mathbb{R}^2$.

The density of X_i is denoted by $f(\theta : x_i)$, $f(x_i;\theta)$ or more often $f(x_i|\theta)$. The likelihood of the sample is the joint density of $X_1, X_2, \ldots, X_n$:

$$L_n(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) \tag{3.6}$$

In the particular case of a discrete distribution, this leads to:

$$L_n(x_1, x_2, \ldots, x_n; \theta) = \mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i) \tag{3.7}$$

Definition: The maximum likelihood estimator of θ is given by

$$\hat{\theta} = arg\max_{\theta} L_n(x_1, x_2, \ldots, x_n; \theta) \tag{3.8}$$

It is therefore, with a fixed sample, the value of the vector of parameters which makes the observations obtained as plausible as possible.

For reasons of analytical convenience, it is often preferred to maximize $\log(L_n)$ rather than L_n. The $\log$ function being strictly increasing, we get the same value, while considerably facilitating the calculations since:

$$\log L_n(x_1, x_2, \ldots, x_n; \theta) = \sum_{i=1}^{n} \log f(x_i; \theta) \tag{3.9}$$

and that it is generally more easier to work on a sum rather than on a product.

3.3 The Exponentiated Generalized Weibull and progressive type II censored data

A type II censored sample is a sample for which only the k smallest observations in a random sample of m elements are observed $(1 \leqslant k \leqslant m)$. Type II censoring is often used in survival analysis, to save time and cost. The type II censored data can be generalized using progressive scheme. In the progressive type II censoring, after the first failing item, R_1 items are removed from the remaining $m-1$ items, then at the second failing item a R_2 items are removed from the remaining $m-2-R_1$, and so forth. The experiment stops after some pre-established number of repetitions of this procedure.

A random variable, X has an exponentiated generalized Weibull distribution with parameter $\theta = (a, b, \alpha, \beta)$, if the probability density function and the survival function of X takes respectively the following form:

$$f(x;a,b,\alpha,\beta) = ab\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\left\{e^{-(x/\beta)^{\alpha}}\right\}^{a}\left[1-\left\{e^{-(x/\beta)^{\alpha}}\right\}^{a}\right]^{b-1} \quad (3.10)$$

and

$$\begin{aligned} S(x;a,b,\alpha,\beta) =& 1 - F(x;a,b,\alpha,\beta) \\ =& 1 - \left[1-\left\{e^{-(x/\beta)^{\alpha}}\right\}^{a}\right]^{b} \end{aligned} \quad (3.11)$$

with $x > 0$, $a > 0$, $b > 0$, $\alpha > 0$, $\beta > 0$.

Let (X_1, R_1), (X_2, R_2), ... , (X_k, R_k), be a progressively type II censored sample, then $X_1 < X_2 < \ldots < X_k$. We make the assuption that $X_1, X_2, \ldots, X_m$, are exponentiated generalized Weibull. A number $k < m$ is determined and the censoring scheme $(R_1, R_2, \ldots, R_k)$ with $R_i > 0$ and $\sum_{i=1}^{k} R_i + k = m$ is satisfied. The likelihood function is expressed as (Cohen 1963):

$$L(x;a,b,\alpha,\beta) = c\prod_{i=1}^{k} f(x_i;a,b,\alpha,\beta)\Big(S(x_i;a,b,\alpha,\beta)\Big)^{R_i} \quad (3.12)$$

where

$$c = m(m - R_1 - 1)(m - R_1 - R_2 - 2)\ldots(m - R_1 - \ldots - R_{k-1} - k + 1),$$

$f(x_i;a,b,\alpha,\beta)$ and $S(x_i;a,b,\alpha,\beta)$ are respectively the exponentiated generalized Weibull probability density and survival functions.

3.4 Expectation maximization algorithm (EM-algorithm)

3.4.1 Principle

The EM-algorithm for Expectation-Maximization algorithm is an iterative algorithm due to Dempster, et al. (1977). It is a parametric estimation method within the general framework of maximum likelihood.

When the only available data do not allow the estimation of the parameters, and (or) the likelihood expression is analytically impossible to maximize, the EM-algorithm can be a solution. Roughly and vaguely, it aims to provide an estimator when this impossibility results from the presence of hidden or missing data or rather, when the knowledge of these data would make it possible to estimate the parameters.

The EM-algorithm derives its name from the fact that at each iteration it operates two distinct steps:

• The word ”Expectation”, often referred to as ”E-step”, proceeds as its name implies, to the estimation of the unknown data, knowing the observed data and the value of the parameters determined at the previous iteration;

• The word ”maximization”, or ”M-step”, then proceeds to the maximization of the likelihood, now made possible by using the estimation of the unknown data completed in the preceding step (E-step), and updates the value of the parameter(s) for the next iteration.

In short, the EM-algorithm proceeds according to an extremely natural mechanism: if there is an obstacle to apply the MLE method, this obstacle is simply blown up and then this method is applied.

The EM algorithm ensures that the likelihood increases with each iteration, which leads to increasingly ”correct” estimators.

3.4.2 The iteration

To formalize somewhat what is stated in the above section:

• We have observations identically independently distribution $X = (X_1, X_2, \ldots, X_n)$ of likelihood denoted $\mathbb{P}(X|\theta)$;

• Maximizing $\log \mathbb{P}(X|\theta)$ is impossible;

• We consider hidden data $Z = (Z_1, \ldots, Z_n)$ whose knowledge would make it possible to maximize the ”likelihood of complete data”, $\log \mathbb{P}(X, Z|\theta)$;

• Since these data Z are not known, we estimate the likelihood of the complete data taking into account all the known information: the estimator is naturally $\mathbb{E}_{Z|X,\theta_m}[\log \mathbb{P}(X, z|\theta)]$ (”E-Step” of the algorithm);

• Finally, this estimated likelihood is maximized to determine the new value of the parameter (” M-step” of the algorithm).

Thus, the passage from the iteration m to the iteration $m+1$ of the algorithm consists in determining:

$$\hat{\theta}_{m+1} = arg \max_{\theta} \mathbb{E}_{Z|X,\hat{\theta}_m}[\log \mathbb{P}(X, z|\theta)] \tag{3.13}$$

Chapter 4

Results and discussions

The results obtained in this study are presented in this chapter; it includes the complete log-likelihood expressions, the EM-algorithm, the conditional expectations, derivations of the complete log-likelihood, simulation study, statistical table, application to real data. Some algorithm for the execution of the analyses in R software are provided in Appendix.

4.1 Parameters estimation

4.1.1 The model

Let us assume that we have n independent variables in a trial, and the ordered m failures are observed under the progressive type-II censoring plan $R = (R_1,\ldots, R_m)$, where $R_j \geq 0$ for $j = 1, \ldots, m$ and $\sum_{j=0}^{m} R_j + m = n$. Let the observed and censored data be respectively $Y = (Y_1, \ldots, Y_m)$ and $Z = (Z_1, \ldots, Z_m)$, where $Z_j = (Z_{j1}, \ldots, Z_{jR_j})$for $j = 1, \ldots, m$. Now let us consider $X = (Y, Z)$ to be the complete data (observed and censored data together). Then the joint probability that the complete sample (the complete data likelihood)is observed is given by

$$L_c(x,a,b,\alpha,\beta)=\prod_{j=1}^{m}\left[f(y_j,a,b,\alpha,\beta)\prod_{k=1}^{R_j}f(z_{jk},a,b,\alpha,\beta)\right], \tag{4.1}$$

(Ng et al 2002).

From which we get the following log-likelihood by using (3.10)

$$\begin{aligned}\log L_c(x,a,b,\alpha,\beta) =& n\log a+n\log b+n\log\alpha-n\log\beta+(\alpha-1)\sum_{j=1}^{m}\log\left(\frac{y_j}{\beta}\right)\\ &-a\sum_{j=1}^{m}\left(\frac{y_j}{\beta}\right)^{\alpha}+(b-1)\sum_{j=1}^{m}\log\left[1-\left\{e^{-(y_j/\beta)^{\alpha}}\right\}^{a}\right]\\ &+(\alpha-1)\sum_{j=1}^{m}\sum_{k=1}^{R_j}\log\left(\frac{z_{jk}}{\beta}\right)-a\sum_{j=1}^{m}\sum_{k=1}^{R_j}\left(\frac{z_{jk}}{\beta}\right)^{\alpha}\\ &+(b-1)\sum_{j=1}^{m}\sum_{k=1}^{R_j}\log\left[1-\left\{e^{-(z_{jk}/\beta)^{\alpha}}\right\}^{a}\right]\end{aligned} \tag{4.2}$$

4.1.2 EM-Algorithm

E-step

In oder to tackle the E-step the conditional expectation of the previous log-likelihood knowing the observed sample $Y=(y_1,y_2,\ldots,y_m)$ is computed. Denote this by $Q(\theta)$ where

$\theta = (a, b, \alpha, \beta)$ is the vector of parameters.

$$
\begin{aligned}
Q(\theta) =& n \log a + n \log b + n \log \alpha - n \log \beta + (\alpha - 1) \sum_{j=1}^{m} \log \left(\frac{y_j}{\beta} \right) \\
& - a \sum_{j=1}^{m} \left(\frac{y_j}{\beta} \right)^{\alpha} + (b-1) \sum_{j=1}^{m} \log \left[1 - \left\{ e^{-(y_j/\beta)^{\alpha}} \right\}^{a} \right] \\
& + (\alpha - 1) \sum_{j=1}^{m} \sum_{k=1}^{R_j} \mathbb{E}_{\theta} \left(\log \left(\frac{z_{jk}}{\beta} \right) | z_{jk} > y_j \right) - a \sum_{j=1}^{m} \sum_{k=1}^{R_j} \mathbb{E}_{\theta} \left(\left(\frac{z_{jk}}{\beta} \right)^{\alpha} | z_{jk} > y_j \right) \\
& + (b-1) \sum_{j=1}^{m} \sum_{k=1}^{R_j} \mathbb{E}_{\theta} \left(\log \left[1 - \left\{ e^{-(z_{jk}/\beta)^{\alpha}} \right\}^{a} \right] | z_{jk} > y_j \right) \qquad (4.3)
\end{aligned}
$$

Thus, to facilitate the E-step, the conditional distribution of Z for given Y and the current value of the parameters, needs to be determined.

In Ng et al (2002) the conditional distribution is given by

$$
f_{Z|Y}(z_{jk}|Y = y, \theta) = \frac{f(z_{jk}, \theta)}{1 - F(y_j, \theta)}, \qquad z_{jk} > y_j \qquad (4.4)
$$

Let us set $A(\theta, y_j) = \mathbb{E}_{\theta} \left(\log \left(\frac{z_{jk}}{\beta} \right) | z_{jk} > y_j \right)$, $\quad B(\theta, y_j) = \mathbb{E}_{\theta} \left(\left(\frac{z_{jk}}{\beta} \right)^{\alpha} | z_{jk} > y_j \right)$

and $C(\theta, y_j) = \mathbb{E}_{\theta} \left(\log \left[1 - \left\{ e^{-(z_{jk}/\beta)^{\alpha}} \right\}^{a} \right] | z_{jk} > y_j \right)$.

Using (4.4) we can obtain the expressions for $A(\theta, y_j)$, $B(\theta, y_j)$ and $C(\theta, y_j)$ as follow.

$$
\begin{aligned}
A(\theta, y_j) &= \mathbb{E}_{\theta} \left(\log \left(\frac{z_{jk}}{\beta} \right) | z_{jk} > y_j \right) \\
&= \frac{1}{1 - F(y_j, \theta)} \int_{y_j}^{\infty} \log \left(\frac{z_{jk}}{\beta} \right) f(z_{jk}, \theta) dz_{jk}
\end{aligned}
$$

Replacing the pdf by its expression (3.10) and applying the following binomial series $(1 + x)^{\mu} = \sum_{i=o}^{\infty} \binom{\mu}{i} x^i, \quad \mu \in \mathbb{R}$ on $\left[1 - \left\{ e^{-(z_{jk}/\beta)^{\alpha}} \right\}^{a} \right]^{b-1}$

we get

$$
\left[1 - \left\{ e^{-(z_{jk}/\beta)^{\alpha}} \right\}^{a} \right]^{b-1} = \sum_{v=0}^{\infty} \binom{b-1}{v} (-1)^v e^{-a(v+1)(z_{jk}/\beta)^{\alpha}}
$$

Hence

$$A(\theta, y_j) = \frac{ab\alpha}{\beta\Big[1 - F(y_j, \theta)\Big]} \int_{y_j}^{\infty} \log\left(\frac{z_{jk}}{\beta}\right)\left(\frac{z_{jk}}{\beta}\right)^{\alpha-1} \sum_{v=0}^{\infty} \binom{b-1}{v} (-1)^v e^{-a(v+1)(z_{jk}/\beta)^{\alpha}} dz_{jk}$$

Due to the Monotone Convergence Theorem, Gallouët and Herbin (2013) we can interchanged the integral and the sum.

$$A(\theta, y_j) = \frac{ab\alpha}{\beta\Big[1 - F(y_j, \theta)\Big]} \sum_{v=0}^{\infty} \binom{b-1}{v} (-1)^v \int_{y_j}^{\infty} \log\left(\frac{z_{jk}}{\beta}\right)\left(\frac{z_{jk}}{\beta}\right)^{\alpha-1} e^{-a(v+1)(z_{jk}/\beta)^{\alpha}} dz_{jk}$$

Now let $x = (z_{jk}/\beta)^{\alpha}$, then

$$\left(\frac{z_{jk}}{\beta}\right)^{\alpha-1} = x^{(\alpha-1)/\alpha} \quad \text{and} \quad \frac{dz_{jk}}{dx} = \frac{\beta}{\alpha} x^{(1-\alpha)/\alpha}$$

Thus

$$A(\theta, y_j) = \frac{ab}{\alpha\Big[1 - F(y_j, \theta)\Big]} \sum_{v=0}^{\infty} \binom{b-1}{v} (-1)^v \int_{(y_j/\beta)^{\alpha}}^{\infty} \log(x) e^{-a(v+1)x} dx$$

Using integration by parts we get

$$A(\theta, y_j) = \frac{ab}{\alpha\Big[1 - F(y_j, \theta)\Big]} \sum_{v=0}^{\infty} \binom{b-1}{v} (-1)^v \left[\frac{\log(y_j/\beta)}{\alpha a(v+1)} e^{-a(v+1)(y_j/\beta)^{\alpha}} + \int_{(y_j/\beta)^{\alpha}}^{\infty} \frac{e^{-a(v+1)x}}{x} dx\right]$$

Let $z = a(v+1)x$ then

$$\begin{aligned} \int_{(y_j/\beta)^{\alpha}}^{\infty} \frac{e^{-a(v+1)x}}{x} dx &= \int_{a(v+1)(y_j/\beta)^{\alpha}}^{\infty} z^{-1} e^{-z} dz \\ &= \Gamma(0, a(v+1)(y_j/\beta)^{\alpha}) \end{aligned}$$

where Γ is upper incomplete gamma function.

Finally we get

$$A(\theta, y_j) = \frac{ab}{\alpha\Big[1 - F(y_j, \theta)\Big]} \sum_{v=0}^{\infty} \binom{b-1}{v} (-1)^v \left[\frac{\log(y_j/\beta)}{\alpha a(v+1)} e^{-a(v+1)(y_j/\beta)^{\alpha}} + \Gamma(0, a(v+1)(y_j/\beta)^{\alpha})\right] \tag{4.5}$$

Next we have

$$B(\theta, y_j) = \mathbb{E}_\theta\left(\left(\frac{z_{jk}}{\beta}\right)^\alpha |z_{jk} > y_j\right)$$
$$= \frac{1}{1 - F(y_j, \theta)} \int_{y_j}^{\infty} \left(\frac{z_{jk}}{\beta}\right)^\alpha f(z_{jk}, \theta) dz_{jk}$$

Replacing the pdf by its expression (3.10) and using the binomial series we have $\left[1 - \left\{e^{-(z_{jk}/\beta)^\alpha}\right\}^a\right]^{b-1} = \sum_{v=0}^\infty \binom{b-1}{v}(-1)^v e^{-a(v+1)(z_{jk}/\beta)^\alpha}$
Which gives

$$B(\theta, y_j) = \frac{ab\alpha}{\beta\Big[1 - F(y_j, \theta)\Big]} \int_{y_j}^{\infty} \left(\frac{z_{jk}}{\beta}\right)^{2\alpha-1} \sum_{v=0}^{\infty} \binom{b-1}{v} (-1)^v e^{-a(v+1)(z_{jk}/\beta)^\alpha} dz_{jk}$$
$$= \frac{ab\alpha}{\beta\Big[1 - F(y_j, \theta)\Big]} \sum_{v=0}^{\infty} \binom{b-1}{v} (-1)^v \int_{y_j}^{\infty} \left(\frac{z_{jk}}{\beta}\right)^{2\alpha-1} e^{-a(v+1)(z_{jk}/\beta)^\alpha} dz_{jk}$$

Now set $x = a(v+1)(z_{jk}/\beta)^\alpha$, then

$$\left(\frac{z_{jk}}{\beta}\right)^{2\alpha-1} = \left(\frac{x}{a(v+1)}\right)^{(2\alpha-1)/\alpha} \quad \text{and} \quad \frac{dz_{jk}}{dx} = \frac{\beta}{\alpha a(v+1)}\left(\frac{x}{a(v+1)}\right)^{(1-\alpha)/\alpha}$$

so that

$$B(\theta, y_j) = \frac{ab}{1 - F(y_j, \theta)} \sum_{v=0}^{\infty} \binom{b-1}{v} \frac{(-1)^v}{(a(v+1))^2} \int_{a(v+1)(y_j/\beta)^\alpha}^{\infty} x e^{-x} dx$$

Therefore

$$B(\theta, y_j) = \frac{ab}{1 - F(y_j, \theta)} \sum_{v=0}^{\infty} \binom{b-1}{v} \frac{(-1)^v}{(a(v+1))^2} \Big(1 + a(v+1)(y_j/\beta)^\alpha\Big) e^{-a(v+1)(y_j/\beta)^\alpha} \tag{4.6}$$

Also we have

$$C(\theta, y_j) = \mathbb{E}_\theta\left(\log\left[1 - \left\{e^{-(z_{jk}/\beta)^\alpha}\right\}^a\right] |z_{jk} > y_j\right)$$
$$= \frac{1}{1 - F(y_j, \theta)} \int_{y_j}^{\infty} \log\left[1 - \left\{e^{-(z_{jk}/\beta)^\alpha}\right\}^a\right] f(z_{jk}, \theta) dz_{jk}$$

Replacing the pdf by its expression (3.10) and using the followings Taylor series

$$\log(1+x) = \sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i} x^i, \qquad x \in (-1, 1] \tag{4.7}$$

and binomial series

$$(1+x)^{\mu} = \sum_{i=o}^{\infty} \binom{\mu}{i} x^i, \qquad \mu \in \mathbb{R} \tag{4.8}$$

so that

$$C(\theta, y_j) =$$

$$\frac{ab\alpha}{\beta\Big[1 - F(y_j,\theta)\Big]} \int_{y_j}^{\infty} \left(\frac{z_{jk}}{\beta}\right)^{\alpha-1} \sum_{i=0}^{\infty} \frac{(-1)^{2i+1}}{i} e^{-ai(z_{jk}/\beta)^{\alpha}} \sum_{v=0}^{\infty} \binom{b-1}{v} (-1)^v e^{-a(v+1)(z_{jk}/\beta)^{\alpha}} dz_{jk}$$

$$= \frac{ab\alpha}{\beta\Big[1 - F(y_j,\theta)\Big]} \int_{y_j}^{\infty} \left(\frac{z_{jk}}{\beta}\right)^{\alpha-1} \sum_{i=0}^{\infty}\sum_{v=0}^{\infty} \frac{(-1)^{2i+1}}{i} \binom{b-1}{v} (-1)^v e^{-a(v+1+i)(z_{jk}/\beta)^{\alpha}} dz_{jk}$$

$$= \frac{ab\alpha}{\beta\Big[1 - F(y_j,\theta)\Big]} \sum_{i=0}^{\infty}\sum_{v=0}^{\infty} \frac{(-1)^{2i+1}}{i} \binom{b-1}{v} (-1)^v \int_{y_j}^{\infty} \left(\frac{z_{jk}}{\beta}\right)^{\alpha-1} e^{-a(v+1+i)(z_{jk}/\beta)^{\alpha}} dz_{jk}$$

Now let $x = a(v+i+1)(z_{jk}/\beta)^{\alpha}$, then

$$\left(\frac{z_{jk}}{\beta}\right)^{\alpha-1} = \left(\frac{x}{a(v+i+1)}\right)^{(\alpha-1)/\alpha} \quad \text{and} \quad \frac{dz_{jk}}{dx} = \frac{\beta}{\alpha a(v+i+1)} \left(\frac{x}{a(v+i+1)}\right)^{(1-\alpha)/\alpha}$$

Therefore

$$C(\theta, y_j) = -\frac{ab}{\Big[1 - F(y_j,\theta)\Big]} \sum_{i=0}^{\infty}\sum_{v=0}^{\infty} \frac{(-1)^v}{ia(v+i+1)} \binom{b-1}{v} \int_{a(v+i+1)(y_j/\beta)^{\alpha}}^{\infty} e^{-x} dx$$

Hence

$$C(\theta, y_j) = -\frac{ab}{\Big[1 - F(y_j,\theta)\Big]} \sum_{i=0}^{\infty}\sum_{v=0}^{\infty} \frac{(-1)^v}{ia(v+i+1)} \binom{b-1}{v} e^{-a(v+i+1)(y_j/\beta)^{\alpha}} \tag{4.9}$$

We therefore obtain an expression for the conditional expectation of the log-likelihood as

$$
\begin{aligned}
Q(\theta) =& n\log a + n\log b + n\log\alpha - n\log\beta + (\alpha-1)\sum_{j=1}^{m}\log\left(\frac{y_j}{\beta}\right) \\
&- a\sum_{j=1}^{m}\left(\frac{y_j}{\beta}\right)^{\alpha} + (b-1)\sum_{j=1}^{m}\log\left[1-\left\{e^{-(y_j/\beta)^{\alpha}}\right\}^{a}\right] \\
&+ (\alpha-1)\sum_{j=1}^{m}R_jA(\theta,y_j) - a\sum_{j=1}^{m}R_jB(\theta,y_j) + (b-1)\sum_{j=1}^{m}R_jC(\theta,y_j)
\end{aligned}
\tag{4.10}
$$

M-step

In the M-step on the p-th iteration of the EM-algorithm, the value of θ which maximizes $Q(\theta,\theta^{(p-1)}) = \mathbb{E}_\theta(\log L_c(x,\theta)|Y,\theta^{(p-1)})$ will be used as the next estimate $\theta^{(p)}$ of θ. Where $\theta^{(p)} = (a^{(p)}, b^{(p)}, \alpha^{(p)}, \beta^{(p)})$ is the vector parameters at the p-th iteration $p \geqslant 1$, and

$\theta^{(0)} = (a^{(0)}, b^{(0)}, \alpha^{(0)}, \beta^{(0)})$ the initial value of the vector parameters.

$$
Q(\theta,\theta^{(p-1)}) = \mathbb{E}_\theta(\log L_c(x,\theta)|Y,\theta^{(p-1)})
$$

Therefore, if at the p-th stage the estimate of θ is $\theta^{(p-1)}$, then $\theta^{(p)}$ can be obtained by maximizing

$$
\begin{aligned}
Q(\theta,\theta^{(p-1)}) =& n\log a + n\log b + n\log\alpha - n\log\beta + (\alpha-1)\sum_{j=1}^{m}\log\left(\frac{y_j}{\beta}\right) - a\sum_{j=1}^{m}\left(\frac{y_j}{\beta}\right)^{\alpha} \\
&+ (b-1)\sum_{j=1}^{m}\log\left[1-\left\{e^{-(y_j/\beta)^{\alpha}}\right\}^{a}\right] + (\alpha-1)\sum_{j=1}^{m}R_jA(\theta^{(p-1)},y_j) \\
&- a\sum_{j=1}^{m}R_jB(\theta^{(p-1)},y_j) + (b-1)\sum_{j=1}^{m}R_jC(\theta^{(p-1)},y_j)
\end{aligned}
\tag{4.11}
$$

Notice that we can write $Q(\theta,\theta^{(p-1)})$ as $T(\theta)$, a function of θ since $\theta^{(p-1)}$ is known (computed in the previous iteration).

Then $\theta^{(p)}$ is solution of the following system of equations

$$\begin{cases} \dfrac{\partial Q(\theta, \theta^{(p-1)})}{\partial a} = 0 \\ \dfrac{\partial Q(\theta, \theta^{(p-1)})}{\partial b} = 0 \\ \dfrac{\partial Q(\theta, \theta^{(p-1)})}{\partial \alpha} = 0 \\ \dfrac{\partial Q(\theta, \theta^{(p-1)})}{\partial \beta} = 0 \end{cases} \tag{4.12}$$

which is equivalent to

$$\begin{cases} \dfrac{n}{a} - \sum_{j=1}^{m} \left(\dfrac{y_j}{\beta}\right)^{\alpha} + (b-1) \sum_{j=1}^{m} \dfrac{\left(\dfrac{y_j}{\beta}\right)^{\alpha} \left\{e^{-(y_j/\beta)^{\alpha}}\right\}^{a}}{1 - \left\{e^{-(y_j/\beta)^{\alpha}}\right\}^{a}} - \sum_{j=1}^{m} R_j B(\theta^{(p-1)}, y_j) = 0 \\ \dfrac{n}{b} + \sum_{j=1}^{m} \log\left[1 - \left\{e^{-(y_j/\beta)^{\alpha}}\right\}^{a}\right] + \sum_{j=1}^{m} R_j C(\theta^{(p-1)}, y_j) = 0 \\ \dfrac{n}{\alpha} + \sum_{j=1}^{m} \log\left(\dfrac{y_j}{\beta}\right) - a \sum_{j=1}^{m} \left(\dfrac{y_j}{\beta}\right)^{\alpha} \log\left(\dfrac{y_j}{\beta}\right) + a(b-1) \sum_{j=1}^{m} \dfrac{\log\left(\dfrac{y_j}{\beta}\right)\left(\dfrac{y_j}{\beta}\right)^{\alpha} \left\{e^{-(y_j/\beta)^{\alpha}}\right\}^{a}}{1 - \left\{e^{-(y_j/\beta)^{\alpha}}\right\}^{a}} \\ + \sum_{j=1}^{m} R_j A(\theta^{(p-1)}, y_j) = 0 \\ -\dfrac{n + (\alpha-1)m}{\beta} + \dfrac{a\alpha}{\beta} \sum_{j=1}^{m} \left(\dfrac{y_j}{\beta}\right)^{\alpha} - \dfrac{a\alpha(b-1)}{\beta} \sum_{j=1}^{m} \dfrac{\left(\dfrac{y_j}{\beta}\right)^{\alpha} \left\{e^{-(y_j/\beta)^{\alpha}}\right\}^{a}}{1 - \left\{e^{-(y_j/\beta)^{\alpha}}\right\}^{a}} = 0 \end{cases} \tag{4.13}$$

From the second equation in the above system we can express $b^{(p)}$ for known $a^{(p)}$, $\alpha^{(p)}$, $\beta^{(p)}$ as:

$$b^{(p)} = -\frac{n}{\sum_{j=1}^{m} \log\left[1 - \left\{e^{-(y_j/\beta^{(p)})^{\alpha^{(p)}}}\right\}^{a^{(p)}}\right] + \sum_{j=1}^{m} R_j C(\theta^{(p-1)}, y_j)} \tag{4.14}$$

The expressions for $a^{(p)}$, $\alpha^{(p)}$ and $\beta^{(p)}$ are not available in closed form.

The solution to the M-step does not exists in closed form. For this case Dempster et al (1977) defined what is called the generalized EM-algorithm (GEM algorithm) for which the M-step requires $\theta^{(p)}$ to be chosen such that

$$Q(\theta^{(p)}, \theta^{(p-1)}) \geqslant Q(\theta^{(p-1)}, \theta^{(p-1)}) \tag{4.15}$$

Since we need only to increase the likelihood, we may replace the M-step with a single iteration of the Newton-Raphson(N-R) algorithm.

4.2 Simulation and remarks

4.2.1 Simulation

For illustration we consider the values of $n = 30$, $m = 20$ and $\theta = (1, 1, 1, 1)$. Progressively Type-II censored sample was generated from the Exponentiated Generalized Weibull distribution using the algorithm in Balakrishnan and Sandhu (1995).
The algorithm is defined as follows:

- Generate m independent Uniform(0,1) observations W_1,W_2, ... ,W_m
- Set $V_i = W_i^{1/(i+R_m+R_{m-1}+\ldots+R_{m-i+1})}$ for $i = 1, 2, \ldots, m$
- Set $U_i = 1 - V_m V_{m-1} \ldots V_{m-i+1}$ for $i = 1, 2, \ldots, m$. Then U_1, U_2,..., U_m, is the required progressive Type-II censored sample from the Uniform (0,1) distribution.
- Finally, we set $X_i = F^{-1}(U_i, \theta)$ for $i = 1, 2, \ldots, m$, where $F^{-1}(., \theta)$ is the inverse cdf of the Exponentiated Generalized Weibull distribution. Then $X_1, X_2, \ldots, X_m$,

is the required progressive Type-II censored sample from the Exponentiated Generalized Weibull distribution.

with censoring scheme $R = (1, 3, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0)$.
The generated sample is given in Table(4.1).

Via the EM algorithm discussed in Section (4.1.2), the computed MLEs of the parameters become:

Table 4.1: Progressive type II data for n = 30, m = 20 and $\theta = (1, 1, 1, 1)$

0.0138	0.0230	0.0447	0.2401	0.3091	0.3264	0.4597	0.5448	0.5841	0.7274,
0.9875	1.1164	1.2090	1.3519	1.4896	1.5041	1.6224	2.9952	3.4537	3.6385

$$\hat{a} = 0.7606,\ \hat{b} = 0.8272,\ \hat{\alpha} = 1.0911 \text{ and } \hat{\beta} = 1.0365$$

In Table(4.2) a Monte Carlo simulation for N=500 was used to compute the RMSE and the mean estimates for different value of n, m and $\theta = (2, 2, 1, 1)$. The following formula was used to compute the RMSE

$$RMSE(\hat{\lambda}) = \sqrt{\sum_{i=1}^{N} \frac{(\hat{\lambda}_i - \lambda)^2}{N}}$$

where $\hat{\lambda}_i$ is the i-th estimates of the parameter λ

Table 4.2: RMSE of the estimators

		Estimates				RMSE			
n	m	$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\beta}$
25	15	1.6346	1.9228	1.2833	0.6301	2.1274	2.0622	1.5970	3.1484
	20	2.0388	1.6436	1.1927	1.0383	1.0591	1.3232	0.9489	1.0364
	25	2.0723	1.9273	0.9889	1.0761	0.4639	0.6423	0.3130	0.6419
40	20	2.0371	1.9960	1.0133	1.0716	1.3787	1.8453	0.8290	0.8120
	30	2.0199	2.0494	0.9433	1.0301	0.4582	0.8808	0.3870	0.2936
	40	2.0011	1.9357	1.0222	1.0030	0.3196	0.5117	0.2313	0.1701
65	30	2.0751	1.9770	1.0153	1.0176	0.3112	0.7451	0.2008	0.1108
	45	2.0148	2.0949	0.9944	1.0012	0.2384	0.3315	0.1546	0.0536
	65	2.0148	1.9991	1.0063	0.9988	0.2610	0.6053	0.1451	0.0720

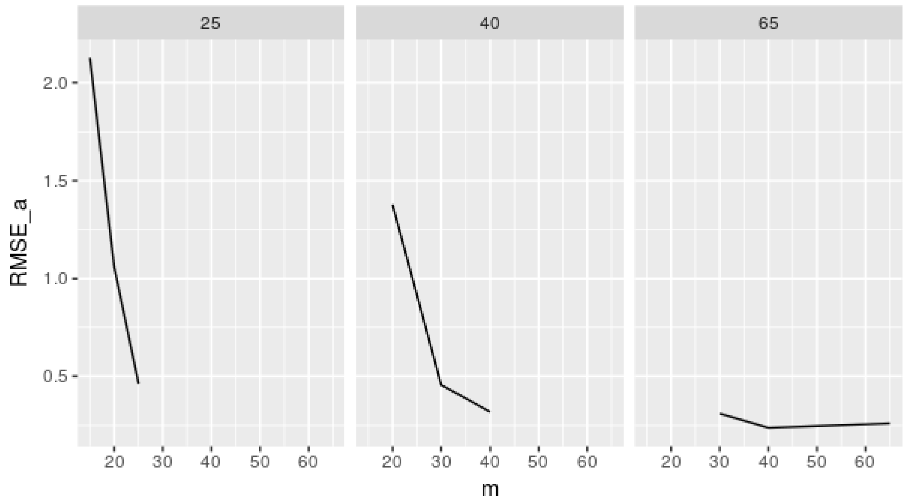

Figure 4.1: Plot of the RMSE of $\hat{a}$

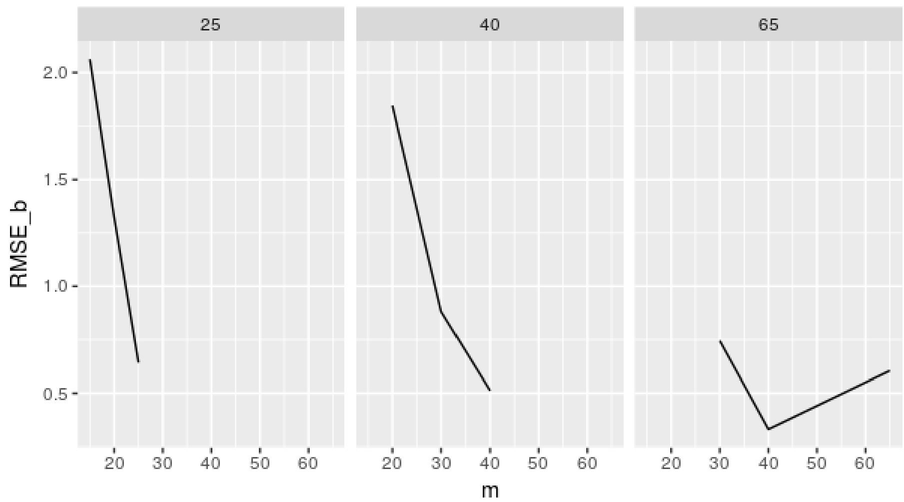

Figure 4.2: Plot of the RMSE of $\hat{b}$

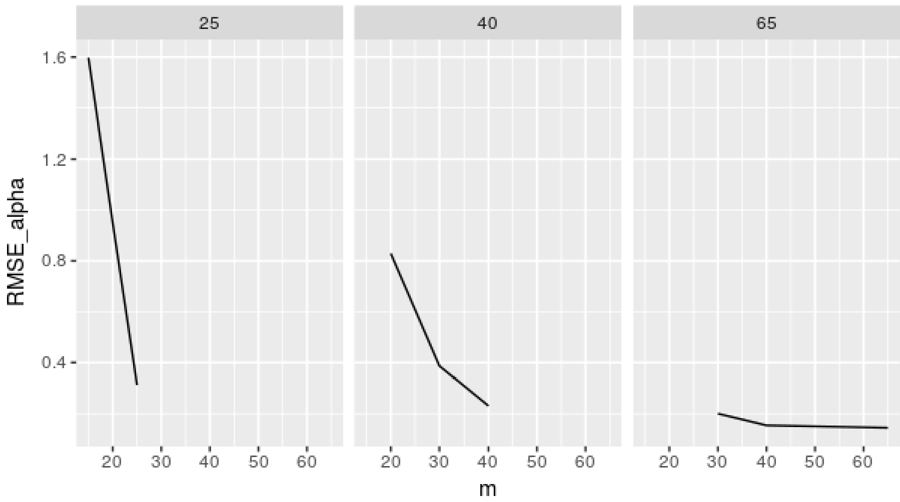

Figure 4.3: Plot of the RMSE of $\hat{\alpha}$

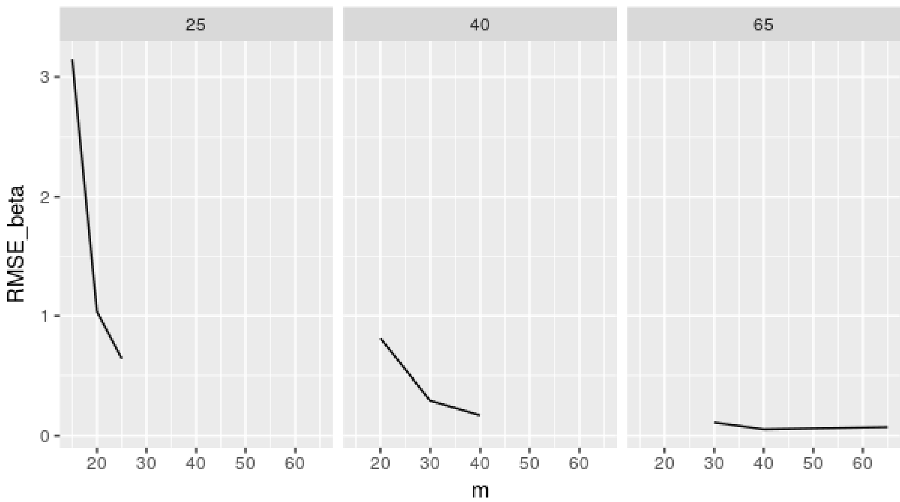

Figure 4.4: Plot of the RMSE of $\hat{\beta}$

4.2.2 Discussion

Three different sample sizes n were used in Table(4.2) and for each sample size we used three different censoring m, which gives nine cases and we compute the estimates and the root mean square error in each case.

In the graphs in Figure(4.1), Figure(4.2), Fgure(4.3), Figure(4.4) we plotted the data in Table(4.2). In each figure we have three graphs each representing respectively the sample of size $n = 25$, $n = 40$, and $n = 65$. Considering:

- the variation of the sample sizes the Root Mean Square Error becomes smaller. That is in other words by increasing the sample size n, the Root Mean Square Error is decreasing. It is noticed by looking at the starting value of each case and the ending value in each case
- the variation of the observed data size for each sample size the RMSE is decreasing. That is in other words for fixed sample size n and by increasing

4.2.3 Remarks

- The largest values of m in each case represent the complete sample case. (see Table(4.2))
- For fixed sample size n and by increasing m, we get smaller RMSE's.(see Table(4.2))
- By increasing the sample size n, we get smaller RMSE's.(see Table(4.2))

4.3 Application to real data

The data set we used is from Nichols and Padgett (2006). It has been analyzed by Mashail and Soliman (2015) to illustrate the estimation of the parameters of the exponentiated Weibull (with two shape parameters) model with adaptive Type-II progressive censored schemes. The data set is composed of 100 observations on breaking stress of carbon fibres (in Gba) as given in Table(4.3):

We take $m = 60$, $R = (20, 0 * 58, 20)$. For clarity $R = (1, 0 * 4, 3)$ is a short form for

Table 4.3: Real data set from Nichols and Padgett

0.39	0.81	0.85	0.98	1.08	1.12	1.17	1.18	1.22	1.25
1.36	1.41	1.47	1.57	1.57	1.59	1.59	1.61	1.61	1.69
1.69	1.71	1.73	1.80	1.84	1.84	1.87	1.89	1.92	2.00
2.03	2.03	2.05	2.12	2.17	2.17	2.17	2.35	2.38	2.41
2.43	2.48	2.48	2.50	2.53	2.55	2.55	2.56	2.59	2.67
2.73	2.74	2.76	2.77	2.79	2.81	2.81	2.82	2.83	2.85
2.87	2.88	2.93	2.95	2.96	2.97	2.97	3.09	3.11	3.11
3.15	3.15	3.19	3.19	3.22	3.22	3.27	3.28	3.31	3.31
3.33	3.39	3.39	3.51	3.56	3.60	3.65	3.68	3.68	3.68
3.70	3.75	4.20	4.38	4.42	4.70	4.90	4.91	5.08	5.56

$R = (1, 0, 0, 0, 0, 3)$

So the observed data is given in Table(4.4): Based on the above progressive Type-II censored data we compute estimates of the parameters of the Exponentiated Generalized Weibull distribution using the algorithm describe above. We get

$$(\hat{a}, \hat{b}, \hat{\alpha}, \hat{\beta}) = c(1.5819, 0.8477, 1.8117, 3.6395)$$

Table 4.4: Observed data set

0.39	0.85	0.98	1.12	1.17	1.18	1.22	1.36	1.41	1.57
1.57	1.59	1.61	1.61	1.69	1.69	1.71	1.73	1.80	1.84
1.84	1.87	1.92	2.03	2.03	2.12	2.17	2.17	2.17	2.35
2.38	2.41	2.48	2.48	2.50	2.53	2.55	2.55	2.56	2.59
2.67	2.74	2.77	2.79	2.81	2.82	2.83	2.87	2.88	2.93
2.95	2.96	2.97	2.97	3.09	3.11	3.11	3.15	3.15	3.19

Chapter 5

Summary, conclusion and recommendation

5.1 Summary

In this study we presented an estimation procedure for estimating the parameters of the Exponentiated Generalized Wiebull based on progressive type II censored data. Since the progressive type II censored data can be considered as a missing data problem the Expectation Maximization algorithm (EM-algorithm) is used to compute the maximum likelihood estimates (MLE's). In the first specific objective we derived the MLE's via the EM-algorithm, where we computed in the E-step (Expectation step) conditional expectations (conditioned by the observed data Y) and then used the Newton Raphson Algorithm instead of the M-step (Maximization step) since the closed is not obtained from in the M-step. The simulation for various degrees of censoring and sample sizes conducted in the second specific objective led us to investigate the performance of the estimators. Finally using the data from Nichols and Padgett paper we applied the estimation to a real data case for the third and last specific objective of the study

5.2 Conclusion

The parameters of the Exponentiated Generalized Weibull distribution were estimated using maximum likelihood estimation method via Expectation Maximization (EM) algorithm. The Root Mean Square Error were computed for different values of the sample size n and failures (observed sample size) m (various degrees of censoring). It was observed that the RMSE's were smaller for fixed sample size n and increasing the size m of the observed data, and also for the increasing sample size n. Application to real data was also given to estimate the parameters of the EGW distribution.

5.3 Recommendations

The study only focused on developing maximum likelihood estimators for the parameters of the Exponentiated Generalized Weibull based on progressive type II censored data. One may be interested in the asymptotic variance-covariance matrix of the MLE's and use it to construct the asymptotic 95% confidence interval for the parameters. One can also be interested on Bayesian estimation of parameters of the Exponentiated Generalized Weibull.

The results of this study will go a long way in making Exponentiated Generalized Weibull (EGW) to be applicable in situation where progressive type II censoring data are encountered, thus contributing to knowledge in the are of survival and reliability analysis.

References

Alshawarbeh, E. (2011). *Beta-Cauchy Distribution and Its Applications*. PhD thesis, Michigan, USA: Central Michigan University.

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

Balakrishnan, N. and Kannan, N. (2001). Point and interval estimation for parameters of the logistic distribution based on progressively type-ii censored samples. *Handbook of statistics*, 20:431–456.

Balakrishnan, N., Kennan, N., Lin, C.-T., and Ng, H. T. (2003). Point and interval estimation for gaussian distribution, based on progressively type-ii censored samples. *IEEE Transactions on Reliability*, 52(1):90–95.

Balakrishnan, N. and Sandhu, R. (1995). A simple simulational algorithm for generating progressive type-ii censored samples. *The American Statistician*, 49(2):229–230.

Bourguignon, M., Silva, R. B., and Cordeiro, G. M. (2014). The weibull-g family of probability distributions. *Journal of Data Science*, 12(1):53–68.

Cohen, A. (1966). Chemotherapy of tuberculosis. *The Journal of the American Osteopathic Association*, 65(9):956–963.

Cohen, C. A. and Whitten, B. (1982). Modified maximum likelihood and modified moment estimators for the three-parameter weibull distribution. *Communications in Statistics-Theory and Methods*, 11(23):2631–2656.

Cordeiro, G. M., Ortega, E. M., and da Cunha, D. C. (2013). The exponentiated generalized class of distributions. *Journal of Data Science*, 11(1):1–27.

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 (methodological)*, pages 1–38.

Dubey, S. D. (1967). Some percentile estimators for weibull parameters. *Technometrics*, 9(1):119–129.

Fréchet, M. (1928). Sur la loi de probabilité de l'écart maximum. *Annales de la societe Polonaise de Mathematique*, 6:93–116.

Gallouët, T. and Herbin, R. (2013). Mesure, intégration, probabilités.

Ghnimi, S. and Gasmi, S. (2014). Parameter estimations for some modifications of the weibull distribution. *Open Journal of Statistics*, 4(08):597.

Herd, G. R. (1956). *Estimation of the parameters of a population from a multi-censored sample*. PhD thesis, Iowa State University.

Johnson, N. L. (1966). Life testing and early failure. *Technometrics*, 8(3):539–545.

Kappenman, R. F. (1985). Estimation for the three-parameter weibull, lognormal, and gamma distributions. *Computational Statistics and Data Analysis*, 3:11–23.

Lai, C., Xie, M., and Murthy, D. (2001). *Bathtub-shaped failure rate life distributions*, volume 20. Elsevier Science BV, Handbook in Reliability, N. Balakrishnan and C. R. Rao.

Lehmann, E. (1983). *Theory of Point Estimation*. John Wiley, New York.

Lehmann, E. L. (1953). The power of rank tests. *The Annals of Mathematical Statistics*, 24(1):23–43.

Mahmoudi, E. and Sepahdar, A. (2013). Exponentiated weibull–poisson distribution: Model, properties and applications. *Mathematics and Computers in Simulation*, 92:76–97.

Mann, N. R. (1969). Exact three-order-statistic confidence bounds on reliable life for a weibull model with progressive censoring. *Journal of the American Statistical Association*, 64(325):306–315.

Mashail, M. and Soliman, A. A. (2015). Estimation for the exponentiated weibull model with adaptive type-ii progressive censored schemes.

Montanari, G. and Cacciari, M. (1988). Progressively-censored aging tests on xlpe-insulated cable models. *IEEE Transactions on Electrical Insulation*, 23(3):365–372.

Mousa, M. A. and Jaheen, Z. (2002). Statistical inference for the burr model based on progressively censored data. *Computers and Mathematics with Applications*, 43(10):1441–1449.

Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42(2):299–302.

Ng, H., Chan, P., and Balakrishnan, N. (2002). Estimation of parameters from progressively censored data using em algorithm. *Computational Statistics and Data Analysis*, 39(4):371–386.

Nichols, M. D. and Padgett, W. (2006). A bootstrap control chart for weibull percentiles. *Quality and reliability engineering international*, 22(2):141–151.

Oguntunde, P., Odetunmibi, O., and Adejumo, A. (2015). On the exponentiated generalized weibull distribution: A generalization of the weibull distribution. *Indian Journal of Science and Technology*, 8(35):1–7.

Rosin, P. (1933). The laws governing the fineness of powdered coal. *J. Inst. Fuel.*, 7:29–36.

Soliman, A. A. (2005). Estimation of parameters of life from progressively censored data using burr-xii model. *IEEE Transactions on Reliability*, 54(1):34–42.

Thomas, D. R. and Wilson, W. M. (1972). Linear order statistic estimation for the two-parameter weibull and extreme-value distributions from type ii progressively censored samples. *Technometrics*, 14(3):679–691.

Weibuil, W. (1951). A statisticai distribution function of wide application. *Journal of Applied Mechanics*.

Zhang, T. and Xie, M. (2011). On the upper truncated weibull distribution and its reliability implications. *Reliability Engineering and System Safety*, 96(1):194–200.

Appendices

R codes for the simulation

```
# The pdf of the Exponentiated Generalized Weibull EGW(a,b,alpha,beta)
f <- function(a,b,alpha,beta,x){
y = a*b*(alpha/beta)*(x/beta)^(alpha-1)*exp(-a*(x/beta)^alpha)*
(1-exp(-a*(x/beta)^alpha))^(b-1)
return(y)
}

# The cdf of the Exponentiated Generalized Weibull EGW(a,b,alpha,beta)
F <- function(a,b,alpha,beta,x){
y = (1-exp(-a*(x/beta)^alpha))^b
return(y)
}

# Generating progressive type II censored
rsample <- function(n,m,par){
p = 0.2 # p is the proba that an individual unit is removed form the te
        # at the i-th failure, i = 1, 2, ... , m - 1
a = par[1]
b = par[2]
alpha = par[3]
```

```
beta = par[4]
#quantile function
regw <- function(x, aa=a, bb=b, aalpha=alpha, bbeta=beta){
q = bbeta*(-(1/aa)*log(1-x^(1/bb)))^(1/aalpha)
return(q)
}
random = regw(runif(m), a, b, alpha, beta)
y = random[order(random)]
r = c()
r[1] = rbinom(1, n-m, p)
int = seq(2, m-1)
s = r[1]
for (i in int){
r[i] = rbinom(1, n-m-s, p)
s = s+r[i]}
if (n-m-s >0)
r[m] = n-m-s
else
r[m] = 0
D = data.frame(y, r)
return(D)
}

# Generating progressive type II censored
# Using algorithm from N. Balakrishnan & R. A. Sandhu (1995)
```

```
balasan <- function(n,m,par){
a = par[1]
b = par[2]
alpha = par[3]
beta = par[4]
regw <- function(x,aa=a,bb=b,aalpha=alpha,bbeta=beta){#quantile functio
q = bbeta*(-(1/aa)*log(1-x^(1/bb)))^(1/aalpha)
return(q)
}
w = runif(m)
r = c()
r[1] = rbinom(1,n-m,0.2)
int = seq(2,m-1)
s = r[1]
for (i in int){
r[i] = rbinom(1,n-m-s,0.2)
s = s+r[i]}
if (n-m-s >0)
r[m] = n-m-s
else
r[m] = 0
p = c()
x = c()
v = c()
u = c()
```

```
y = c()
p = r[m]
for (i in 1:m){
v[i] = w[i]^(1/(i+p))
p = p+r[m-i]}
x = v[m]
for (i in 1:m){
u[i] = 1-x # progressive type II sample from U(0,1)
x = x*v[m-i]
y[i] = regw(u[i],a,b,alpha,beta)}
D = data.frame(y,r)
return(D)
}

#The conditional expectations A, B and C
#A
Aa <- function(par,m){
a = par[1]
b = par[2]
alpha = par[3]
beta = par[4]
y = c()
c = c()
for(j in 1:m){
c[j] = 1/(1-F(a,b,alpha,beta,frame$y[j]))
```

```
g <- function(x){
log(x/beta)*f(a,b,alpha,beta,x)
}
y[j] = c[j]*integrate(g,lower=frame$y[j],upper=Inf)$value
}
return(y)
}

#B
Bb <- function(par,m){
a = par[1]
b = par[2]
alpha = par[3]
beta = par[4]
y = c()
c = c()
for(j in 1:m){
c[j] = 1/(1-F(a,b,alpha,beta,frame$y[j]))
g <- function(x){
((x/beta)^alpha)*f(a,b,alpha,beta,x)
}
y[j] = c[j]*integrate(g,lower=frame$y[j],upper=Inf)$value
}
return(y)
}
```

```
#C
Cc <- function(par,m){
a = par[1]
b = par[2]
alpha = par[3]
beta = par[4]
y = c()
c = c()
for(j in 1:m){
c[j] = 1/(1-F(a,b,alpha,beta,frame$y[j]))
g<-function(x){
log(1-(exp(-(x/beta)^alpha))^a)*f(a,b,alpha,beta,x)
}
y[j] = c[j]*integrate(g,lower=frame$y[j],upper=Inf)$value
}
return(y)
}

# The function Q, that is the complete log likelihood
Q <- function(par){
a = par[1]
b = par[2]
alpha = par[3]
beta = par[4]
```

```
n*log((a*b*alpha)/beta)+(alpha-1)*sum(log(frame$y/beta))-a*sum((frame$y
}

library(numDeriv)
library(Matrix)
library(rootSolve)
# The em algorithm
em <- function(frame, n, m, a, b, alpha, beta, eps = 1/100000){
err <- 1
iter <- 0
param = c(a, b, alpha, beta)
results <- NULL
results <- rbind(results, param)
while(err > eps ) {
A = Aa(frame,param,m)
B = Bb(frame,param,m)
C = Cc(frame,param,m)
gvec <- param - solve(hessian(Q, param)) %*% gradient(Q, param)[1,]
old.param <- param
param <- gvec[,1]
err <- max(abs((old.param - param)/param))
iter <- iter + 1
results <- rbind(results, gvec[,1])
}
return(results)
```

```
}

# Simulation for n = 30, m = 20 and (a, b, alpha, beta) = (1, 1, 1, 1)
n = 30
m = 20
a = 1
b = 1
alpha = 1
beta = 1
par = c(a, b, alpha, beta)
frame = balasan(n, m, par)
em(frame, n, m, a, b, alpha, beta)
N = 1
mle_a <- c(rep(0,N))
mle_b <- c(rep(0,N))
mle_alpha <- c(rep(0,N))
mle_beta <- c(rep(0,N))
n = 30
m = 20
param = c(1, 1, 1, 1)
par = param
a = par[1]
b = par[2]
alpha = par[3]
beta = par[4]
```

```
for (i in 1:N){
frame=balasan(n,m,par) #can also use rsample(n,m,par)
A = Aa(par,m)
B = Bb(par,m)
C = Cc(par,m)
gvec = par # initiale value
gvec <- gvec - solve(hessian(Q, par)) %*% gradient(Q, par)[1,]
mle.vec <- gvec[,1]
mle_a[i] <- mle.vec[1]
mle_b[i] <- mle.vec[2]
mle_alpha[i] <- mle.vec[3]
mle_beta[i] <- mle.vec[4]
}
print(cbind(mle_a, mle_b, mle_alpha, mle_beta))
### mean estimates
mean_a <- sum(mle_a)/N
mean_b <- sum(mle_b)/N
mean_alpha <- sum(mle_alpha)/N
mean_beta <- sum(mle_beta)/N
### Calculate Average Bias
ABias_a <- sum(mle_a-a)/N
ABias_b <- sum(mle_b-b)/N
ABias_alpha <- sum(mle_alpha-alpha)/N
ABias_beta <- sum(mle_beta-beta)/N
### Calculate RMSE
```

```
RMSE_a <- sqrt(sum((a-mle_a)^2)/N)
RMSE_b <- sqrt(sum((b-mle_b)^2)/N)
RMSE_alpha <- sqrt(sum((alpha-mle_alpha)^2)/N)
RMSE_beta <- sqrt(sum((beta-mle_beta)^2)/N)
print(cbind(mean_a, mean_b, mean_alpha, mean_beta))
print(cbind(ABias_a,ABias_b,ABias_alpha,ABias_beta))
print(cbind(RMSE_a,RMSE_b,RMSE_alpha,RMSE_beta))

# Monte Carlo Simulations N = 500, n = 25 and various censored size
n = 25
m = 15 # 20, 25
N = 500
mle_a <- c(rep(0,N))
mle_b <- c(rep(0,N))
mle_alpha <- c(rep(0,N))
mle_beta <- c(rep(0,N))
param = c(2, 2, 1, 1)
par=param
a=par[1]
b=par[2]
alpha=par[3]
beta=par[4]
for (i in 1:N){
  frame=balasan(n,m,par) #can also use rsample(n,m,par)
  d <- em(frame, n, m, a, b, alpha, beta)
```

```r
  mle.vec <- d[nrow(d), ]
  mle_a[i] <- mle.vec[1]
  mle_b[i] <- mle.vec[2]
  mle_alpha[i] <- mle.vec[3]
  mle_beta[i] <- mle.vec[4]
}
print(cbind(mle_a, mle_b, mle_alpha, mle_beta))
### mean estimates
mean_a<-sum(mle_a)/N
mean_b<-sum(mle_b)/N
mean_alpha<-sum(mle_alpha)/N
mean_beta<-sum(mle_beta)/N
### Calculate Average Bias
ABias_a<-sum(mle_a-a)/N
ABias_b<-sum(mle_b-b)/N
ABias_alpha<-sum(mle_alpha-alpha)/N
ABias_beta<-sum(mle_beta-beta)/N
### Calculate RMSE
RMSE_a<-sqrt(sum((a-mle_a)^2)/N)
RMSE_b<-sqrt(sum((b-mle_b)^2)/N)
RMSE_alpha<-sqrt(sum((alpha-mle_alpha)^2)/N)
RMSE_beta<-sqrt(sum((beta-mle_beta)^2)/N)
print(cbind(mean_a, mean_b, mean_alpha, mean_beta))
print(cbind(ABias_a,ABias_b,ABias_alpha,ABias_beta))
print(cbind(RMSE_a,RMSE_b,RMSE_alpha,RMSE_beta))
```

```
# Monte Carlo Simulations N = 500, n = 40 and various censored size
n = 40
m = 20 # 30, 40
N = 500
mle_a <- c(rep(0,N))
mle_b <- c(rep(0,N))
mle_alpha <- c(rep(0,N))
mle_beta <- c(rep(0,N))
param = c(2, 2, 1, 1)
par=param
a=par[1]
b=par[2]
alpha=par[3]
beta=par[4]
for (i in 1:N){
frame=balasan(n,m,par) #can also use rsample(n,m,par)
d <- em(frame, n, m, a, b, alpha, beta)
mle.vec <- d[nrow(d), ]
mle_a[i] <- mle.vec[1]
mle_b[i] <- mle.vec[2]
mle_alpha[i] <- mle.vec[3]
mle_beta[i] <- mle.vec[4]
}
print(cbind(mle_a, mle_b, mle_alpha, mle_beta))
```

```
### mean estimates
mean_a<-sum(mle_a)/N
mean_b<-sum(mle_b)/N
mean_alpha<-sum(mle_alpha)/N
mean_beta<-sum(mle_beta)/N
### Calculate Average Bias
ABias_a<-sum(mle_a-a)/N
ABias_b<-sum(mle_b-b)/N
ABias_alpha<-sum(mle_alpha-alpha)/N
ABias_beta<-sum(mle_beta-beta)/N
### Calculate RMSE
RMSE_a<-sqrt(sum((a-mle_a)^2)/N)
RMSE_b<-sqrt(sum((b-mle_b)^2)/N)
RMSE_alpha<-sqrt(sum((alpha-mle_alpha)^2)/N)
RMSE_beta<-sqrt(sum((beta-mle_beta)^2)/N)
print(cbind(mean_a, mean_b, mean_alpha, mean_beta))
print(cbind(ABias_a,ABias_b,ABias_alpha,ABias_beta))
print(cbind(RMSE_a,RMSE_b,RMSE_alpha,RMSE_beta))

# Monte Carlo Simulations N = 500, n = 65 and various censored size
n = 65
m = 30 # 45, 65
N = 500
mle_a <- c(rep(0,N))
mle_b <- c(rep(0,N))
```

```
mle_alpha <- c(rep(0,N))
mle_beta <- c(rep(0,N))
param = c(2, 2, 1, 1)
par=param
a=par[1]
b=par[2]
alpha=par[3]
beta=par[4]
for (i in 1:N){
frame=balasan(n,m,par) #can also use rsample(n,m,par)
d <- em(frame, n, m, a, b, alpha, beta)
mle.vec <- d[nrow(d), ]
mle_a[i] <- mle.vec[1]
mle_b[i] <- mle.vec[2]
mle_alpha[i] <- mle.vec[3]
mle_beta[i] <- mle.vec[4]
}
print(cbind(mle_a, mle_b, mle_alpha, mle_beta))
### mean estimates
mean_a<-sum(mle_a)/N
mean_b<-sum(mle_b)/N
mean_alpha<-sum(mle_alpha)/N
mean_beta<-sum(mle_beta)/N
### Calculate Average Bias
ABias_a<-sum(mle_a-a)/N
```

```
ABias_b<-sum(mle_b-b)/N
ABias_alpha<-sum(mle_alpha-alpha)/N
ABias_beta<-sum(mle_beta-beta)/N
### Calculate RMSE
RMSE_a<-sqrt(sum((a-mle_a)^2)/N)
RMSE_b<-sqrt(sum((b-mle_b)^2)/N)
RMSE_alpha<-sqrt(sum((alpha-mle_alpha)^2)/N)
RMSE_beta<-sqrt(sum((beta-mle_beta)^2)/N)
print(cbind(mean_a, mean_b, mean_alpha, mean_beta))
print(cbind(ABias_a,ABias_b,ABias_alpha,ABias_beta))
print(cbind(RMSE_a,RMSE_b,RMSE_alpha,RMSE_beta))

N = 500
mle_a <- c(rep(0,N))
mle_b <- c(rep(0,N))
mle_alpha <- c(rep(0,N))
mle_beta <- c(rep(0,N))
n = 25
m = 15
param = c(2, 2, 1, 1)
par=param
a=par[1]
b=par[2]
alpha=par[3]
beta=par[4]
```

```
for (i in 1:N){
frame=balasan(n,m,par) #can also use rsample(n,m,par)
A=Aa(par,m)
B=Bb(par,m)
C=Cc(par,m)
gvec=par # initiale value
gvec <- gvec - solve(hessian(Q, par)) %*% gradient(Q, par)[1,]
mle.vec <- gvec[,1]
mle_a[i] <- mle.vec[1]
mle_b[i] <- mle.vec[2]
mle_alpha[i] <- mle.vec[3]
mle_beta[i] <- mle.vec[4]
}
print(cbind(mle_a, mle_b, mle_alpha, mle_beta))
### mean estimates
mean_a<-sum(mle_a)/N
mean_b<-sum(mle_b)/N
mean_alpha<-sum(mle_alpha)/N
mean_beta<-sum(mle_beta)/N
### Calculate Average Bias
ABias_a<-sum(mle_a-a)/N
ABias_b<-sum(mle_b-b)/N
ABias_alpha<-sum(mle_alpha-alpha)/N
ABias_beta<-sum(mle_beta-beta)/N
### Calculate RMSE
```

```
RMSE_a<-sqrt(sum((a-mle_a)^2)/N)
RMSE_b<-sqrt(sum((b-mle_b)^2)/N)
RMSE_alpha<-sqrt(sum((alpha-mle_alpha)^2)/N)
RMSE_beta<-sqrt(sum((beta-mle_beta)^2)/N)
print(cbind(mean_a, mean_b, mean_alpha, mean_beta))
print(cbind(ABias_a,ABias_b,ABias_alpha,ABias_beta))
print(cbind(RMSE_a,RMSE_b,RMSE_alpha,RMSE_beta))
```