MIXTURES OF BETA WEIBULL G FAMILY OF DISTRIBUTIONS AND APPLICATIONS

BY

TEFO BARAKI

Reg. No:12000553

BSc Applied Science (Mathematics and Statistics), (BIUST, BW)

Department of Mathematics and Statistical Sciences

Faculty of Sciences

Botswana International University of Science and Technology

E-mail: tefo.baraki@studentmail.biust.ac.bw

A Thesis Submitted to the Faculty of Sciences in Partial Fulfilment of the Requirement for the Award of the Degree of Masters in Statistics of BIUST

Supervisors:

Dr Nkumbuludzi Ndwapi
Department of Mathematics
and Statistical Sciences
Faculty of Sciences, BIUST
Email: ndwapin@biust.ac.bw

Dr Thabiso Maupong
Department of Computer
Science and Information Systems
Faculty of Sciences, BIUST
Email: maupongt@biust.ac.bw

Signature:..
Date: 23 July 2021

Signature..
Date 23 July 2021

November, 2020

Declaration and Copyright

I <u>Tefo Baraki</u>, with student ID 12000553, hereby declare that this thesis has been composed by myself and that it has never been submitted for a degree at this or any other university or institution of higher learning. I confirm that the work contained in this thesis is my own research work under the supervision of Dr N. Ndwapi and Dr T. Maupong. Any work done by others has been acknowledged and referenced accordingly.

Signature:.. Date: 23 July 2021

Certification

The undersigned certify that they have read and hereby recommend for acceptance by the Faculty of Science a thesis titled: MIXTURES OF BETA WEIBULL G FAMILY OF DISTRIBUTIONS AND APPLICATIONS, in fulfilment of the requirements for the Degree of Masters in Statistics of BIUST.

Supervisors:

Dr Nkumbuludzi Ndwapi
Department of Mathematics
and Statistical Sciences
Faculty of Sciences, BIUST
Email: ndwapin@biust.ac.bw

Signature:...
Date: 23 July 2021

Dr Thabiso Maupong
Department of Computer
Science and Information Systems
Faculty of Sciences, BIUST
Email: maupongt@biust.ac.bw

Signature...
Date 23 July 2021

Acknowledgements

I would like to thank my supervisor Dr Nkumbuludzi Ndwapi for his guidance and support during my studies. His insightful discussions have steered me in the right direction, whenever he thought I needed it. He consistently allowed this thesis to be my own work, whilst giving unwavering support. Secondly, I would like to thank Dr Thabiso Maupong for being a co-supervisor to my work. He has availed himself whenever I needed guidance, and he has particularly made a great contribution to my research writing.

I would like to thank my friends and family for their encouragement throughout my academic journey.

Finally, I will like to express my sincere gratitude to the Botswana Insurance Holdings Limited company, for funding my studies through the Professor Thomas Tlou Scholarship.

Contents

List of Figures

List of Tables

Abstract

Mixture models have gained popularity in statistical analyses because of their flexibility in capturing local variations in heterogeneous populations. Model based approaches to classification use mixture models to fit data via maximum likelihood based approaches and provide labels to unlabelled observations. Over the years model based approaches have grown into an important sub-field of classification because they provide the uncertainty of classifying the unlabelled observations as probabilities. Despite many advances in model based approaches to classification, not much work is evidenced in the literature where reliability data is concerned. The Weibull mixtures are often used in modelling reliability data but they are limited to data with monotone failure rates. To this end we introduce a Beta Weibull G (BWG) mixture that provides an appealing framework for handling reliability data with non monotone failure rate functions. Parametric estimation is executed by the Expectation Maximization algorithm, which is an extension of maximum likelihood estimation. The Bayesian Information Criterion is used for model selection. Model based clustering and mixture discriminant analysis techniques are used to assign labels to unlabelled observation. These labels are cross validated by the Adjusted Rand Index. Additionally, parsimony is introduced to the BWG mixtures, by adding constraints on some of the parameter estimates. The constrained models give rise to simple models with great explanatory predictive power. To demonstrate the utility of the proposed approaches, different data sets are simulated to mimic reliability data with non monotone failure rates. The findings of this thesis demonstrate that mixtures of the BWG family of distributions fit heterogeneous population with non monotone hazard rates better than mixtures of the Weibull distributions as evidenced by higher values of BIC for BWG mixtures. The BWG mixtures performed better than Weibull mixtures in both model based clustering and mixture discriminant analysis as demonstrated by high values of the ARI.

Chapter 1

Introduction

Reliability studies investigate time-to-event of objects in a population. This could be the time it takes for car tyres to be worn out, light bulbs to die or yarn fabrics to break. Conventionally, the Weibull distribution is used to fit reliability data, see Nelson (1985), Rinne (2008) and Meeker and Escobar (2014) . Fitting reliability data with the Weibull distribution is done under the assumption that the data has a monotone failure rate function. The monotone failure rate functions only capture failure rates that either decrease or increase over time. These are events when defective items are removed early from a population or when items become defective at a later stage in a population. However, there are cases where the failure rate function is non monotone. Examples of such cases include bathtube failure rate function discussed by Almalki and Nadarajah (2015) and unimodal failure rates studied in Merovci and Elbatal (2015). In these cases the Weibull distribution is not a reasonable fit due to the non monotone failure rate functions.

Over the years, there have been some generalisations of the Weibull distribution to accommodate various forms of the failure rate function. Examples include new generalised class of modified Weibull distributions investigated by Oluyede et al., (2015) , gamma Weibull G family of distributions suggested by Oluyede (2018) and the Beta Weibull G family of distributions (BWG) introduced by Fagbamigbe et al., (2018) . Even though the BWG family of distributions is flexible in modelling various failure rates, it does not provide a sufficient fit for a heterogeneous population.

Heterogeneous data usually occur when the observed data aggregate into two or more groups which result in unknown distributional shapes. Authors such as McLachlan and Peel (2000) and Titterington et al., (1985) argue that heterogeneous data are not sufficiently fitted using one statistical distribution due to their unknown distributional shapes. A common approach when data do not follow one statistical distribution is to use mixture distributions. These are compositions of more than one statistical distributions. Mixture distributions are convenient in fitting heterogeneous data due to their flexibility in capturing local variations in observed data, as opposed to using one classical distribution.

In the context of mixture modelling, data are viewed to be arising from more than one population. The distributions of the underlying sub-populations are called the component distributions and their weighted sum is a probability distribution called the mixture distribution. Mixture models have been used by several authors in fitting heterogeneous data. A classical example is the work of Pearson (1894), who applied a mixture of two normal distributions to a data set from Weldon (1892) , where the normal distribution was a poor fit for the skewed data.

Apart from fitting heterogeneous data, mixture models are a good tool to use in classification problems, see Ko et al., (2007), McNicholas (2010) and McNicholas (2016). Classification is concerned with assigning labels to observations where either all observations do not have labels or some observations do not have labels. Classification can be explained in terms of clustering and discriminant analysis.

1.1 Clustering and Discriminant Analysis

Clustering partitions data into groups with similar characteristics, on the assumption that none of the observations has a prior class label. Some of clustering techniques include approaches such as hierarchical clustering, K-means clustering and model-based clustering. K-means and hierarchical clustering are simple techniques but they are not always ideal for mixture models. This is because K- means clustering is sensitive to noise and outliers while hierarchical clustering is such that once a particular merge or split decision is made then the method cannot backtrack and make any adjustments. Moreover, the method does not have explicit clusters.

A superior approach to clustering is the model based clustering technique. Model based clustering is explained by McNicholas (2010) as an unsupervised learning technique that uses mixture models to aggregate data into clusters. Observations following the same distribution are assigned to the same cluster.

Discriminant analysis on the other hand is a type of classification in which some observations have class labels whilst other observations do not have labels and only the observations which are prior labelled into groups are used to infer labels of unlabelled observations. Fisher′s discriminant function, Linear discriminant analysis (LDA) and Quadratic discriminant analysis (QDA) are the most commonly used techniques due to their ease in implementation. Unfortunately these methods have strong assumptions. For example, Fisher′s discriminant method assumes equal covariance structure of the underlying population while LDA assumes normality and homogeneous class covariance and QDA assumes normality with heterogeneous class covariance. When these assumptions are violated, an alternative approach is the mixture discriminant analysis (MDA).

Mixture discriminant analysis, explained in detail by Hastie and Tibshirani (1996), is a supervised learning technique that uses mixture models to develop a discriminant rule. In MDA, labelled observations are used to derive the discriminating rule, then the unlabelled observations are assigned labels based on this discriminating rule.

1.2 Contributions

An extensive literature review on mixture models of reliability data, yields mixtures of the classical Weibull distributions, see Bucar et al., (2004) , Qin et al., (2012), Razali and Al-Wakeel (2013) and Shin et al., (2016). However, it can be argued that mixtures of the classical Weibull distribution do not model heterogeneous population with non monotone failure rates effectively. To this end, the use of the Beta Weibull G family of distributions in mixture modelling is proposed.

A hypothesis that the BWG mixture out performs the mixture of Weibull distributions is made, because the BWG family of distributions is more flexible than the Weibull distribution. In this thesis two special cases of the mixtures of BWG family of distributions are developed. These are mixtures of Beta Weibull log logistic distribution (BWLLoG) and mixtures of Beta Weibull Exponential distribution (BWE). These extensions have applications in the textile industry as demonstrated by Fagbamigbe et al., (2018) , and these mixtures could also be applied in other areas where the reliability data follow non monotone failure rate functions. The utility of the proposed mixtures will be demonstrated by real data sets and simulated data sets. Parameter estimation for the mixture models will be done using the Expectation Maximisation (EM) algorithm and the Bayesian Information Criterion (BIC) will be used for model selection.

The contributions of this thesis are as follows.

• Mixtures of the Beta Weibull G Family of Distributions.

The proposed mixtures fill the gap where fitting heterogeneous reliability data sets by the classical Weibull mixture is not a reasonable fit. This typically occurs when the underlying sub populations have non monotone hazard functions. Also some constraints are imposed on parameters of the BWG mixture models, resulting in simpler models with high explanatory and predictive power.

- Model Based Classification with Mixtures of the Beta Weibull G Family of Distributions.

  We propose model based approaches to classification that are tailored for reliability data: where the underlying sub populations have non monotone hazard functions. These model based approaches to classification are also extended to using BWG mixture models with fewer parameter estimates.

1.3 Thesis Structure

1.3.1 Chapter 2

This chapter provides background information on;

- Mixture distributions with emphasis on parameter estimation using the Expectation Maximisation (EM) algorithm as an extension of maximum likelihood estimation.
- The Newton Raphson algorithm and EM algorithm and their application in parameter estimation.
- Criterion used for model selection.
- Model based clustering (MBC).
- Mixture Discriminant Analysis (MDA).
- Performance diagnostics in clustering and discriminant analysis.
- The Beta Weibull G family of distributions.

1.3.2 Chapter 3

This chapter consists of the contributions of this thesis.

- The construction of mixtures of the Beta Weibull G family of distributions.
- An overview of how the EM algorithm will be used for parameter estimation in mixtures of Beta Weibull G Family of distribution is given.

- An overview of how model based clustering and mixture discriminant analysis with mixtures of Beta Weibull G Family of Distributions.
- Special cases of mixtures of the Beta Weibull G family of distribution being the mixture of Beta Weibull log logistic (BWLLoG) distribution and the mixture of Beta Weibull Exponential (BWE) distribution.

1.3.3 Chapter 4

In this chapter the utility of the proposed techniques in Chapter 2 is demonstrated by;

- Providing simulated data sets and some real life data sets to be used in the analyses.
- Fitting both the full and parsimonious mixture models of the BWG family of distributions.
- Performing model based clustering and mixture discriminant analysis using the BWG mixtures.

1.3.4 Chapter 5

This chapter highlights the conclusions reached from this work and makes suggestions for future work.

Chapter 2

Background

This chapter provides background information on mixture distributions with emphasis on parameter estimation using the Expectation Maximisation (EM) algorithm as an extension of maximum likelihood estimation. A detailed account of the Newton Raphson algorithm and EM algorithm is also provided. In addition this chapter addresses convergence criterion of the EM algorithm and model selection. A discussion on model based clustering and mixture discriminant analysis then follows. Within the chapter, diagnostics in clustering and discriminant analysis are also discussed. A discussion of the Beta Weibull G family of Distributions concludes the chapter.

2.1 Mixture Distribution

Mixture distributions model heterogeneous population by a weighted sum of probability density functions with non negative mixing proportions that sum to unity. The statistical distributions of the sub-populations are called the components of the mixture and their weighted sum is a probability distribution called the mixture distribution.

A formal definition of a parametric finite mixture given by Titterington et al., (1985), is as follows.

Let X be a set of observations from a distribution with a density function

$$f(x \mid \mathbf{\Theta}) = \sum_{g=1}^{G} \pi_g f_g(x \mid \boldsymbol{\theta}_g) \tag{2.1}$$

then $x \in$ X arises from a *parametric finite mixture* distribution, where G is the number of components of the mixture, $\pi_g > 0$ are mixing proportions such that $\sum_{g=1}^{G} \pi_g = 1$, $\mathbf{\Theta} = [\pi_1, ..., \pi_G, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_G]$ is the vector of parameters and $f_g(x \mid \boldsymbol{\theta}_g)$ is the component probability density function of the g^{th} component. Usually the component distributions are taken from the same family of distribution.The vector of parameters $\mathbf{\Theta} = [\pi_1, ..., \pi_G, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_G]$ is unknown and it should be estimated in order to fully characterise the mixture model.

2.2 Parametric Estimation of Mixture Distributions

There are several methods that can be used to approximate the vector of parameters $\mathbf{\Theta} = [\pi_1, ..., \pi_G, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_G]$. These include methods of moments, maximum likelihood estimation (MLE), Bayes′ estimation and minimum χ^2. Authors such as Day (1969), Tan and Chang (1972), Holgersson and Jorner (1978) have compared the afore mentioned techniques and concluded that the best fit is achieved by the maximum likelihood estimation.

A formal definition of the *likelihood function of a mixture distribution* given by McLachlan and Peel (2000), is

$$\ell(\mathbf{\Theta} \mid x) := \prod_{i=1}^{n} \sum_{g=1}^{G} \pi_g f(x_i \mid \boldsymbol{\theta}_g), \tag{2.2}$$

where n is the number of observations, x_i is the i^{th} observation and G, π_g, θ_g are the same as in equation 2.1.

Then the *log likelihood function of a mixture* is defined by

$$\ell\ell(\mathbf{\Theta} \mid x) = log \prod_{i=1}^{n} \sum_{g=1}^{G} \pi_g f(x_i \mid \boldsymbol{\theta}_g). \tag{2.3}$$

To simplify the computation of the log likelihood function of the mixture for each of the groups the variable z_{ig} associated with the i^{th} observation in the g^{th} group is introduced. The variable, z_{ig} equals to 1 if the observation x_i belongs to a sub-population g and 0 otherwise and the vector $\mathbf{z} = [z_{i1}, z_{i2}, ..., z_{nG}]$.

The G component mixture distribution is such that the observation x_i belongs to exactly one of the sub-populations $g = 1, 2, ..., G$. The observed data are viewed to be incomplete because the latent variables associated with component labels are unobserved. McLachlan and Peel (2000) incorporated the latent variable z_{ig} in the log likelihood function of mixture models, resulting in the complete log likelihood function of a mixture distributions. The *complete log likelihood function of a mixture distribution* is defined by

$$\ell\ell_c(\boldsymbol{\Theta} \mid x, \mathbf{z}) := \sum_{i=1}^{n} \sum_{g=1}^{G} z_{ig}(log\pi_g + logf(x_i \mid \boldsymbol{\theta}_g)). \tag{2.4}$$

The complete log likelihood plays a key role in parameter estimation via the EM algorithm as outlined in section 2.2.1.

2.2.1 The Expectation Maximisation Algorithm

The EM algorithm proposed by Dempster et al., (1977) , is a general iterative method for maximum likelihood parameter estimation when the data are incomplete. The EM algorithm works well in estimating parameters in mixture models. The algorithm alternates between two steps, the expectation step (E-step) and the maximisation step (M-step). On each E-step, the expected value of the complete-data log-likelihood is calculated conditional on current parameter estimates.

Let $Q = \ell\ell_c(\boldsymbol{\Theta} \mid x, \mathbf{z})$ be the complete log likelihood of the mixture model defined by

$$Q = \sum_{i=1}^{n} \sum_{g=1}^{G} z_{ig} log(\pi_g f(x_i \mid \boldsymbol{\theta}_g)).$$

The expected value of the complete log likelihood is calculated as follows,

$$E(Q) = \sum_{i=1}^{n} \sum_{g=1}^{G} log(\pi_g f(x_i \mid \boldsymbol{\theta}_g)) E[z_{ig} \mid x_i]. \tag{2.5}$$

By Bayes theorem, the expected value (e_{ig}) of the latent variable z_{ig} is calculated as

$$\begin{aligned} E[z_{ig} \mid x_i] &= P(z_{ig} = 1 \mid x_i) \\ &= \frac{\pi_g f(x_i \mid \boldsymbol{\theta}_g)}{\sum_{h=1}^{G} \pi_h f(x_i \mid \boldsymbol{\theta}_h)} \\ &= e_{ig}. \end{aligned} \tag{2.6}$$

From (2.5) and (2.6) it follows that,

$$
\begin{aligned}
E(Q) &= \sum_{i=1}^{n}\sum_{g=1}^{G} log(\pi_g f(x_i \mid \boldsymbol{\theta}_g))e_{ig} \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G}(e_{ig} log\pi_g) + \sum_{i=1}^{n}\sum_{g=1}^{G} e_{ig} log f(x_i \mid \boldsymbol{\theta}_g). \qquad (2.7)
\end{aligned}
$$

The updated log likelihood follows from replacing z_{ig} by their expected values. This is equivalent to replacing z_{ig} by its estimate $\hat{z}_{ig}$.

$$
\hat{z}_{ig} := \frac{\hat{\pi}_g f(x_i \mid \hat{\boldsymbol{\theta}}_g)}{\sum_{h=1}^{G} \hat{\pi}_h f(x_i \mid \hat{\boldsymbol{\theta}}_h)} \qquad (2.8)
$$

For each M-step, the maximum likelihood estimates of the model parameters are updated based on maximising the expected value of the complete-data log-likelihood ($\ell\ell_c(\boldsymbol{\Theta} \mid x, \mathbf{z})$) from the preceding E-step with respect to $\boldsymbol{\Theta}$. Then by using Lagrange multipliers (λ),

$$
\frac{\partial}{\partial \pi_g}\left(\sum_{i=1}^{n}\sum_{g=1}^{G}(\hat{z}_{ig} log\pi_g) + \lambda[\sum_{i=1}^{G}\pi_g - 1]\right) = 0
$$

and

$$
\frac{\partial}{\partial \lambda}\left(\sum_{i=1}^{n}\sum_{g=1}^{G}(\hat{z}_{ig} log\pi_g) + \lambda[\sum_{i=1}^{G}\pi_g - 1]\right) = 0,
$$

thus $\lambda = -n$.

Then the updated estimates $\hat{\pi}_g$ are obtained as

$$
\hat{\pi}_g = \frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G}\hat{z}_{ig}, \qquad (2.9)
$$

The updated estimates $\hat{\boldsymbol{\theta}}_g$ are obtained as solutions of

$$
\frac{\partial}{\partial \boldsymbol{\theta}_g}\left(\sum_{i=1}^{n}\sum_{g=1}^{G}\hat{z}_{ig}(log f(x_i \mid \boldsymbol{\theta}_g))\right) = 0. \qquad (2.10)
$$

Where there are no analytical solutions to equation 2.10 numeric techniques are used to find updated estimates $\hat{\boldsymbol{\theta}}_g$. There are several iterative techniques that are used to calculate numeric solutions to problems in optimizations when analytical solutions are not plausible. These include quasi-Newton methods, modified Newton methods and Newton Raphson method.

2.2.2 The Newton Raphson Algorithm

The Newton Raphson method explained by Little and Rubin (2014) is a technique that uses the linear Taylor series expansion to find maximum likelihood estimates. In the context of finite mixtures, a system of partial derivatives of the complete log likelihood is represented as a Jacobian matrix denoted by $J(\mathbf{\Theta} \mid x, \mathbf{z})$. Thus the Newton Raphson method finds solutions to

$$J(\mathbf{\Theta} \mid x, \mathbf{z}) = \frac{\partial \ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial \boldsymbol{\theta}_g} = \mathbf{0}.$$

The gradient of $J(\mathbf{\Theta} \mid x)$ is approximated by a linear Taylor series expansion about the current estimate $\mathbf{\Theta}^{(s)}$ of the parameter $\mathbf{\Theta}$. The gradient of $J(\mathbf{\Theta} \mid x, \mathbf{z})$ is a Hessian matrix denoted by $H(\mathbf{\Theta} \mid x, \mathbf{z})$ and defined by

$$H(\mathbf{\Theta} \mid x, \mathbf{z}) = \frac{\partial^2 \ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial \boldsymbol{\theta}_g^2}.$$

Hence,

$$J(x, \mathbf{z} \mid \mathbf{\Theta}) \approx J(x, \mathbf{z} \mid \mathbf{\Theta}^{(s)}) - H(\mathbf{\Theta}^{(s)} \mid x, \mathbf{z})(\mathbf{\Theta} - \mathbf{\Theta}^{(s)}).$$

The new estimate $\mathbf{\Theta}^{(s+1)}$ is found by rearranging the equation above.

$$\mathbf{\Theta}^{(s+1)} \approx \mathbf{\Theta}^{(s)} + H^{-1}(\mathbf{\Theta}^{(s)} \mid x, \mathbf{z}) J(x, \mathbf{z} \mid \mathbf{\Theta}).$$

In this thesis the M step in the EM algorithm will be executed using the Newton Raphson algorithm.

The EM algorithm will alternate between these two steps until a convergence criterion is satisfied. Conversely the EM algorithm is stopped when the likelihood function is not improving significantly for consecutive iterations. That is when $\ell^{(k+1)} - \ell^{(k)} < \epsilon$, where ϵ is a small value indicating the allowed margin of error.

Over the years several authors such as Aitken (1926), Bohning et al., (1994), Lindsay (1995) and McNicholas (2010) have made suggestions on the convergence criterion of the EM algorithm. In this thesis the convergence criterion suggested by McNicholas (2010) will be used. This criterion is given by

$$\ell_\infty^{(k+1)} - \ell^{(k)} < \epsilon$$

where $\ell_\infty^{(k+1)} = \ell^{(k)} + \frac{1}{1-a^{(k)}}(\ell^{(k+1)} - \ell^{(k)})$ and $a^{(k)} = \frac{\ell^{(k+1)} - \ell^{(k)}}{\ell^{(k)} - \ell^{(k-1)}}$

2.2.3 The Bayesian Information Criterion

The Bayesian Information Criterion (BIC), proposed by Schwarz (1978) is given by

$$BIC = 2\ell\ell(\hat{\boldsymbol{\Theta}} \mid x) - plog(n),$$

where $\hat{\boldsymbol{\Theta}}$ is the maximum likelihood estimate of $\boldsymbol{\Theta}$, $\ell\ell(\hat{\boldsymbol{\Theta}} \mid x)$ is the maximised log likelihood and p is the number of free parameters. The larger the value of BIC the better the model.

According to Keribin (2000) and Dasgupta and Raftery (1998) the BIC is commonly used to choose the number of components in mixture models. Even though the BIC is popularly used, McNicholas (2016) argues that it does not necessarily give the best classification performance amongst candidate models. In this thesis the BIC will be used for model selection because it penalises for the number of parameters used in the mixture model.

2.3 Model Based Clustering

Model based clustering uses finite mixture models to perform clustering. This method supposes that the data follow a certain probability distribution model and it allocates observations following the same distribution into the same cluster. This is achieved by finding the maximum likelihood estimates of the mixture model via the EM algorithm and obtaining the partitions by using posterior probability estimation.

The predicted clustering results are given by a posterior probability defined by

$$\hat{z}_{ig} := \frac{\hat{\pi}_g f(x_i \mid \hat{\boldsymbol{\theta}}_g)}{\sum_{h=1}^{G} \hat{\pi}_h f(x_i \mid \hat{\boldsymbol{\theta}}_h)}. \tag{2.11}$$

The estimate $\hat{\pi}_g$ can be viewed as the prior probability that x_i belongs to the g^{th} sub-population. Then the posterior probability of z_{ig} given the observed value of x_i will be central for clustering purposes.

According to Kaufman and Rousseeuw (2009) and Bezdek (2013) the assignment of points to clusters is *soft*, if the membership of a data point in a cluster is given as a probability. This allows the observations to belong to several clusters simultaneously, with different degrees of membership. The assignment of points to clusters is *hard* if a strict regulation that each observation belongs to exactly one cluster is imposed. The assignment of labels can be hardened by the maximum a posterior (MAP), MAP$\{\hat{z}_{ig}\} = 1$ if g= arg max$_h$ $\{\hat{z}_{ih}\}$ occurs at component h and MAP$\{\hat{z}_{ig}\} = 0$ otherwise. This allows each observation to belong to exactly one cluster. Each observation is assigned to a cluster where the degree of membership of that observation is maximum. In this thesis the model based clustering with hard posterior will be used so that each observation is assigned to exactly one cluster.

2.4 Performance Diagnostics in Clustering

2.4.1 The Adjusted Rand Index

The Adjusted Rand Index suggested by Hubert and Arabie (1985) is a correction of the Rand index for the number of pairwise agreements that would be expected to occur if the observations were classified at random. The similarity evaluation is defined by

$$ARI = \frac{RI - expectedRI}{maximumRI - expectedRI}.$$

where RI is the rand index.

The Rand Index (RI) by Rand (1971) is an evaluation measure for a clustering problem based on the number of pairwise class agreement (A) and the number of pairwise class disagreement(D) between object pairs thus making a comparison of the true class labels and the predicted class labels. The similarity evaluation is defined by

$$RI = \frac{A}{A + D} \ ,$$

where RI $\in [0, 1]$, and RI=1 indicates perfect class agreement. The interpretation of the ARI is similar to that of the RI except that an ARI value of 0 indicates classification results that would be expected under random classification. In this thesis, the ARI will be used to cross validate the labels assigned by model based clustering and mixture discriminant analysis techniques since the ARI is corrected for random allocation.

2.5 Mixture Discriminant Analysis

Mixture discriminant analysis is based on the use of finite mixtures to perform discriminant analysis. This method as explained by Hastie and Tibshirani (1996) , allocates observations following the same distribution into the same group.

Suppose there are n observations where only the first k observations have a prior class label. Within the mixture discriminant analysis paradigm only these k observations are used to estimate the model parameters via the EM algorithm. The class labels of the $n-k$ unlabelled observations are then predicted by posterior probability defined by

$$\hat{z}_{jg} := \frac{\hat{\pi}_g f(x_j \mid \hat{\boldsymbol{\theta}}_g)}{\sum_{h=1}^{G} \hat{\pi}_h f(x_j \mid \hat{\boldsymbol{\theta}}_h)}. \tag{2.12}$$

for $j = k+1, ..., n$.

These $\hat{z}_{jg}$ play the role of a discriminant rule. Similar to model based clustering the posterior predicted classifications are soft and they can be hardened to MAP classifications.

2.6 The Beta Weibull G family of Distributions

Beta Weibull G Family of Distributions (BWG) was proposed by Fagbamigbe et al., (2018). It generalizes the Weibull distribution which is widely used in survival analysis and reliability studies.

Let $X = \{x_1, x_2, ..., x_n\}$ be a set of observations coming from a BWG family of distributions. Then $x \in X$ has a probability density function (pdf) given by;

$$\begin{aligned} f_{BWG}(x; a, b, \alpha, \beta, \upsilon) &= \frac{\alpha\beta}{B(a,b)} m(x;\upsilon) \frac{M(x;\upsilon)^{\beta-1}}{\bar{M}(x;\upsilon)^{\beta+1}} exp\left\{-\alpha b \left[\frac{M(x;\upsilon)}{\bar{M}(x;\upsilon)}\right]^{\beta}\right\} \\ &\times \left[1 - exp\left\{-\alpha \left[\frac{M(x;\upsilon)}{\bar{M}(x;\upsilon)}\right]^{\beta}\right\}\right]^{a-1}, \end{aligned} \tag{2.13}$$

where $x \geqslant 0$, $a > 0$, $b > 0$, α is a scale parameter such that $\alpha > 0$, β is a shape parameter such that $\beta > 0$, υ is the parameter of the underlying baseline distribution such that $\upsilon \geqslant 0$. Moreover, $m(x;\upsilon)$ is the probability density function of the baseline distribution, $M(x;\upsilon)$ is the cumulative distribution function of the baseline distribution and $\bar{M}(x;\upsilon)$ is the survival function of the baseline distribution.

The BWG family of distributions models complex data with non monotone failure rates such as bath tube and up side down bath tube. It is thus more flexible than the Weibull distributions. Fagbamigbe et al., (2018) used the BWG family of distributions to fit the time to failure of yarn, taking the underlying baseline distribution to be the log logistic distribution. In this thesis the BWG distribution will be used as a component density for mixtures of reliability data with non monotone failure rates.

Chapter 3

Mixture of the Beta Weibull G Family of Distributions

This chapter begins with the construction of mixtures of the Beta Weibull G family of distributions. As previously discussed in the introduction, mixtures of Beta Weibull G Family of distributions are used in a heterogeneous population that has several causes of failure where each cause of failure follows a Beta Weibull G distribution. The chapter proceeds by developing an EM algorithm that will be used for parameter estimation in mixtures of Beta Weibull G Family of distributions. We further develop model based clustering and mixture discriminant analysis techniques that are tailored for mixtures of Beta Weibull G Family of Distributions. The chapter ends by giving special cases of mixtures of the Beta Weibull G family of distribution being the mixture of Beta Weibull log logistic (BWLLoG) distribution and the mixture of Beta Weibull Exponential (BWE) distribution.

A BWG mixture follows from equation 2.1 where the component distributions follow from equation 2.13. Thus the mixture will be defined as ;

$$\begin{aligned} f(x \mid \mathbf{\Theta}) &= \sum_{g=1}^{G} \pi_g f_{BWG}(x; a_g, b_g, \alpha_g, \beta_g, \upsilon_g) \\ &= \sum_{g=1}^{G} \frac{\pi_g \alpha_g \beta_g}{B(a_g, b_g)} m(x; \upsilon_g) \frac{M(x; \upsilon_g)^{\beta_g - 1}}{\bar{M}(x; \upsilon_g)^{\beta_g + 1}} exp\left\{ -\alpha_g b_g \left[\frac{M(x; \upsilon_g)}{\bar{M}(x; \upsilon_g)} \right]^{\beta_g} \right\} \\ &\times \left[1 - exp\left\{ -\alpha_g \left[\frac{M(x; \upsilon_g)}{\bar{M}(x; \upsilon_g)} \right]^{\beta_g} \right\} \right]^{a_g - 1} \end{aligned} \tag{3.1}$$

The parameters in the above mixture distribution are unknown and they are estimated based on the complete log likelihood function.

3.1 Parametric Estimation of BWG Mixtures

The complete log likelihood function of the mixture of BWG follow from equation 2.4, thus defined as;

$$\begin{aligned}\ell\ell_c(\boldsymbol{\Theta} \mid x, \mathbf{z}) &:= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}[log\pi_g + logf_{BWG}(x_i; a_g, b_g, \alpha_g, \beta_g, \upsilon_g)] \\ &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(log\frac{\pi_g\alpha_g\beta_g}{B(a_g, b_g)} + log[m(x_i;\upsilon_g)] + log\left[\frac{M(x_i;\upsilon_g)^{\beta_g-1}}{\bar{M}(x_i;\upsilon_g)^{\beta_g+1}}\right] - \alpha_g b_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} \\ &+ (a_g-1)log\left[1 - exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]\Big) + \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}log(\pi_g)\end{aligned} \tag{3.2}$$

The afore log likelihood function is used in the Expected Maximisation for BWG mixtures to estimate the unknown parameters.

The EM has been explained in chapter 2 as an extension of maximum likelihood estimation. Identifiability is a necessary requirement for asymptotic theory to hold for the maximum likelihood estimation of parameters in mixture models, as such, the identifiability of BWG mixtures has to be proven before using the EM algorithm.

3.1.1 Identifiability of BWG mixture models

We prove the identifiability of BWG mixture models following an approach by Chandra (1977) based on Theorems 5.4.1 and 5.4.2 of the appendix.

Let T be a random variable with a BWG pdf. Since moments functions are unique for each distribution, define $\Phi_g(s)$ as the s^{th} moment of the g^{th} mixture component. Then

$$\Phi_g(s) = \sum_{g=1}^{G}\sum_{i,j,k=0}^{\infty} \frac{\alpha\beta(-1)^{i+j}[\alpha(b+i)]^j\Gamma(\alpha)\Gamma(\beta j+\beta+k+1)}{B(a,b)i!j!k!\Gamma(\alpha-i)\Gamma(\beta j+\beta+1)(\beta j+\beta+k)}\int_0^{\infty} x^s f_{\beta j+\beta+k}(x)dx \tag{3.3}$$

for g=1,2 and $i, j, k \in \mathbb{Z}^+$. Note that $D_\Phi = (-\beta, \infty)$ for $\beta > 0$ and if $\beta_1 < \beta_2$ then $D_{\Phi_1} \subseteq D_{\Phi_2}$.

Let t be a value T, then from theorem 5.4.1 $F_1 < F_2$ when $\beta_1 < \beta_2$ and $\alpha_1 = \alpha_2$ with $\alpha_2 < t$ or $\beta = 2$ but $\alpha_1 > \alpha_2$.

Let $s_1 = -\beta_1 + \frac{1}{n}$ where $s_1 \in D_{\Phi_1}$. When $s_1 \to -\beta$ then $n \to \infty$. Thus,

$$\lim_{s \to \beta_1} \Phi_1(s) = \lim_{n \to \infty} \alpha_1^{-\beta_1 + \frac{1}{n}} \Gamma(1 + \frac{-\beta_1 + \frac{1}{n}}{\beta_1}) = \alpha_1^{\beta_1} \Gamma(0^+) \to +\infty. \tag{3.4}$$

On the other hand when $\beta_1 < \beta_2$ and $\alpha_1 = \alpha_2 < t$ we have

$$\lim_{s \to -\beta_1} \Phi_2(s) = \alpha_1^{-\beta_1} \Gamma(1 - \frac{\beta_1}{\beta_2}) > 0. \tag{3.5}$$

thus

$$\lim_{s \to \beta_1} \frac{\Phi_2(s)}{\Phi_1(s)} = \frac{\alpha_1^{-\beta_1} \Gamma(1 - \frac{\beta_1}{\beta_2}) > 0}{+\infty} = 0 \tag{3.6}$$

Hence the BWG mixtures are identifiable and the EM algorithm and its extensions can be used.

3.1.2 The EM Algorithm for BWG Mixtures

A suitable approach to parameter estimation in mixture models is via the EM algorithm. The expected step of the EM algorithm used for BWG mixtures follows from equation 2.5 and the estimation of mixing proportions follow from equation 2.9.

$$\begin{aligned} E(\ell\ell_c(\boldsymbol{\Theta} \mid x, \mathbf{z})) &= E[\ell\ell_c(\boldsymbol{\Theta}, \mathbf{z}) \mid x] \\ &= \sum_{i=1}^{n} \sum_{g=1}^{G} log(\pi_g f_{BWG}(x_i; a_g, b_g, \alpha_g, \beta_g, \upsilon_g)) E[z_{ig} \mid x_i], \end{aligned} \tag{3.7}$$

where

$$\begin{aligned} E[z_{ig} \mid x_i] &= P(z_{ig} = 1 \mid x_i) \\ &= \frac{\pi_g f_{BWG}(x_i \mid a_g, b_g, \alpha_g, \beta_g, \upsilon_g)}{\sum_{h=1}^{G} \pi_h f_{BWG}(x_i \mid a_h, b_h, \alpha_h, \beta_h, \upsilon_h)} \\ &= e_{ig} \end{aligned} \tag{3.8}$$

thus,

$$\begin{aligned} E(\ell\ell_c(\boldsymbol{\Theta} \mid x, \mathbf{z})) &= \sum_{i=1}^{n}\sum_{g=1}^{G} log(\pi_g f_{BWG}(x_i \mid a_g, b_g, \alpha_g, \beta_g, \upsilon_g))e_{ig} \\ &= \sum_{i=1}^{n}\sum_{g=1}^{G}(e_{ig} log\pi_g) + \sum_{i=1}^{n}\sum_{g=1}^{G} e_{ig} log f_{BWG}(x_i \mid a_g, b_g, \alpha_g, \beta_g, \upsilon_g). \end{aligned} \tag{3.9}$$

The log likelihood of BWG mixtures is updated by replacing the z_{ig} by their expected values. This is equivalent to replacing z_{ig} by

$$\hat{z}_{ig} := \frac{\hat{\pi}_g f_{BWG}(x_i \mid \hat{a}_g, \hat{b}_g, \hat{\alpha}_g, \hat{\beta}_g, \hat{\upsilon}_g)}{\sum_{h=1}^{G} \hat{\pi}_h f_{BWG}(x_i \mid \hat{a}_h, \hat{b}_h, \hat{\alpha}_h, \hat{\beta}_h, \hat{\upsilon}_h)}. \tag{3.10}$$

The M step of the EM algorithm will be carried out by maximising the updated complete log likelihood of the BWG mixture. This requires solving a system of equations of partial derivatives of the complete log likelihood.

These partial derivatives are calculated in appendix 5.1 and given as

$$\begin{aligned} \frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \alpha_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\alpha_g} - b_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} + (a_g - 1)\left[1 - exp\left\{-\alpha_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]^{-1} \\ &\times \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} exp\left\{-\alpha_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\Big), \end{aligned} \tag{3.11}$$

$$\begin{aligned} \frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \beta_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\beta_g} + log\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right] - \alpha_g b_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} log\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right] \\ &+ (a_g - 1)\alpha_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} log\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right] \\ &\times exp\left\{-\alpha_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\left[1 - exp\left\{-\alpha_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]^{-1}\Big), \end{aligned} \tag{3.12}$$

$$\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial a_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big([\varphi_0(a_g + b_g) - \varphi_0(a_g)] + log\left[1 - exp\left\{-\alpha_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]\Big), \tag{3.13}$$

$$\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial b_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big([\varphi_0(a_g + b_g) - \varphi_0(b_g)] - \alpha_g \left[\frac{M(x_i; \upsilon_g)}{\bar{M}(x_i; \upsilon_g)}\right]^{\beta_g}\Big) \tag{3.14}$$

and

$$\begin{aligned}\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \upsilon_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{m(x_i, \upsilon_g)}\frac{\partial}{\partial \upsilon_g}m(x_i, \upsilon_g) + \frac{(\beta_g - 1)}{M(x_i, \upsilon_g)}\frac{\partial}{\partial \upsilon_g}M(x_i, \upsilon_g) - \frac{(\beta_g + 1)}{\bar{M}(x_i, \upsilon_g)}\frac{\partial}{\partial \upsilon_g}\bar{M}(x_i, \upsilon_g) \\ &- \alpha_g b_g \beta_g \left(\frac{M(x_i, \upsilon_g}{\bar{M}(x_i, \upsilon_g}\right)^{\beta_g - 1}\left(\frac{1}{\bar{M}(x_i, \upsilon_g)}\frac{\partial}{\partial \upsilon_g}M(x_i, \upsilon_g) - \frac{M(x_i, \upsilon_g)}{[\bar{M}(x_i, \upsilon_g)]^2}\frac{\partial}{\partial \upsilon_g}\bar{M}(x_i, \upsilon_g)\right) \\ &+ (a_g - 1)\left[1 - exp\left\{-\alpha_g \left[\frac{M(x_i; \upsilon_g)}{\bar{M}(x_i; \upsilon_g)}\right]^{\beta_g}\right\}\right]^{-1}\left[exp\left\{-\alpha_g\left[\frac{M(x_i; \upsilon_g)}{\bar{M}(x_i; \upsilon_g)}\right]^{\beta_g}\right\}\right] \\ &\left[\alpha_g \beta_g \left(\frac{M(x_i, \upsilon_g}{\bar{M}(x_i, \upsilon_g}\right)^{\beta_g - 1}\left(\frac{1}{\bar{M}(x_i, \upsilon_g)}\frac{\partial}{\partial \upsilon_g}M(x_i, \upsilon_g) - \frac{M(x_i, \upsilon_g)}{[\bar{M}(x_i, \upsilon_g)]^2}\frac{\partial}{\partial \upsilon_g}\bar{M}(x_i, \upsilon_g)\right)\right]\Big).\end{aligned} \tag{3.15}$$

The updated estimate for $\hat{\pi}_g$ is obtained as

$$\hat{\pi}_g = \frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G} \hat{z}_{ig}, \tag{3.16}$$

The partial derivatives with respect to $a,\ b,\ \alpha,\ \beta$ and υ do not have analytic solutions. They will be solved by the Newton Raphson algorithm which was explained in chapter 2.

The Jacobian matrix is set up as

$$J(\boldsymbol{\theta}_g \mid x, \mathbf{z}) = \frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \boldsymbol{\theta}_g} = \mathbf{0}.$$

$$J(\boldsymbol{\theta}_g \mid x, \mathbf{z}) = \begin{pmatrix} \frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \alpha_g} \\ \frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \beta_g} \\ \frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial a_g} \\ \frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial b_g} \\ \frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \upsilon_g} \end{pmatrix}$$

and the Hessian matrix as

$$H(\boldsymbol{\theta}_g \mid x, \mathbf{z}) = \begin{pmatrix} \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \alpha_g^2} & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \alpha_g \partial \beta_g} & \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \beta_g^2} & \cdots & \cdots & \cdots \\ \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \alpha_g \partial a_g} & \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \beta_g \partial a_g} & \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial a_g^2} & \cdots & \cdots \\ \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \alpha_g \partial b_g} & \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \beta_g \partial b_g} & \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial a_g \partial b_g} & \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial b_g^2} & \cdots \\ \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \alpha_g \partial v_g} & \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \beta_g \partial v_g} & \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial a_g \partial v_g} & \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial b_g \partial v_g} & \frac{\partial \ell\ell_c^2(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial v_g^2} \end{pmatrix}.$$

These gives updates to the parameter vector $\hat{\boldsymbol{\theta}}_g$ as

$$\boldsymbol{\theta}_g^{(s+1)} \approx \boldsymbol{\theta}_g^{(s)} + H^{-1}(\boldsymbol{\theta}_g^{(s)} \mid x, \mathbf{z}) J(x, \mathbf{z} \mid \boldsymbol{\theta}_g^{(s)}) = \begin{pmatrix} \alpha_g^{(s+1)} \\ \beta_g^{(s+1)} \\ a_g^{(s+1)} \\ b_g^{(s+1)} \\ v_g^{(s+1)} \end{pmatrix}.$$

The updated parameter estimates are used to re estimate the latent variable z_{ig} on the E step in the next iteration. The algorithm will alternate between the E step and M step until it reaches convergence.

A full BWG mixture has $(G \times p)$ parameters where G is the number of component densities and p is the number of parameters in each component density. It is often advisable to explain data with a simple model that has an optimal number of parameters. Too many parameters over fit the data and lower explanatory predictive power. Using very few parameters will under fit the data and lower predictive power. In finding the optimal number of parameters to be used in the model some parameters are constrained. This will yield simple models that have a higher predictive power .

3.2 Model Based Clustering with BWG Mixtures

Model based clustering has been identified in chapter 2 as a superior approach in clustering when the underlying data is characterised by a mixture distribution. The parameters estimated from the EM algorithm will be used to carry out a model based clustering technique.

The posterior probability of an observation x_i belonging to the g^{th} group in the BWG mixture will follow from equation 2.11 , thus,

$$\hat{z}_{ig} = \frac{\hat{\pi}_g f_{BWG}(x_i \mid \hat{a}_g, \hat{b}_g, \hat{\alpha}_g, \hat{\beta}_g, \hat{\upsilon}_g)}{\sum_{h=1}^{G} \hat{\pi}_h f_{BWG}(x_i \mid \hat{a}_h, \hat{b}_h, \hat{\alpha}_h, \hat{\beta}_h, \hat{\upsilon}_h)}. \tag{3.17}$$

An observation will be assigned membership to a group in which it scores the highest $\hat{z}_{ig}$.

3.3 Mixture Discriminant Analysis with BWG Mixtures

The labelled observations will be used for parameter estimation via the EM algorithm. The resulting estimates will be used to carry out a mixture discriminant analysis. The posterior probability of an observation x_i belonging to the g^{th} group in the mixture will follow from equation 2.12 ,thus,

$$\hat{z}_{jg} = \frac{\hat{\pi}_g f_{BWG}(x_j \mid \hat{a}_g, \hat{b}_g, \hat{\alpha}_g, \hat{\beta}_g, \hat{\upsilon}_g)}{\sum_{h=1}^{G} \hat{\pi}_h f_{BWG}(x_j \mid \hat{a}_h, \hat{b}_h, \hat{\alpha}_h, \hat{\beta}_h, \hat{\upsilon}_h)}. \tag{3.18}$$

An observation will be assigned membership to a group in which it scores the highest $\hat{z}_{jg}$.

3.4 Mixture of the Beta Weibull log logistic Distribution

In this subsection the baseline distribution of the BWG is specified to be the log logistic distribution, resulting in a Beta Weibull log logistic distribution (BWLLoG). Different density plots for BWLLoG with varying parameter values are given. Subsequently, density plots of mixtures of BWLLoG are given to depict heterogeneous reliability data with non monotone failure rate functions. The EM algorithm for BWLLoG mixtures is developed together with the model based clustering technique and mixture discriminant analysis rule suitable for BWLLoG mixtures.

Consider a mixture of the Beta Weibull G family of distributions when the baseline distribution is chosen to be the log logistic distribution. The pdf, cdf and survival function of the log logistic distribution are given by

$$m(x,c) = cx^{c-1}(1+x^c)^{-2}, M(x,c) = 1-(1+x^c)^{-1} \; and \; \bar{M}(x,c) = 1-[1-(1+x^c)^{-1}] = (1+x^c)^{-1}$$

respectively, where c is the shape parameter such that $c \geqslant 0$.

The pdf of the Beta Weibull log logistic is defined by;

$$f_{BWLLoG}(x; a, b, \alpha, \beta, c) = \frac{\alpha\beta c}{B(a,b)} x^{c-1}(1+x^c)^{-2} \frac{(1-(1+x^c)^{-1})^{\beta-1}}{((1+x^c)^{-1})^{\beta+1}} exp\left\{-\alpha b x^{c\beta}\right\} \times \left[1 - exp\left\{-\alpha x^{c\beta}\right\}\right]^{a-1}, \tag{3.19}$$

for $x \geqslant 0, a > 0, b > 0, \alpha > 0, \beta > 0, c \geqslant 0$.

The density plots for Beta Weibull log logistic with varying parameter values is given in figure 3.1.

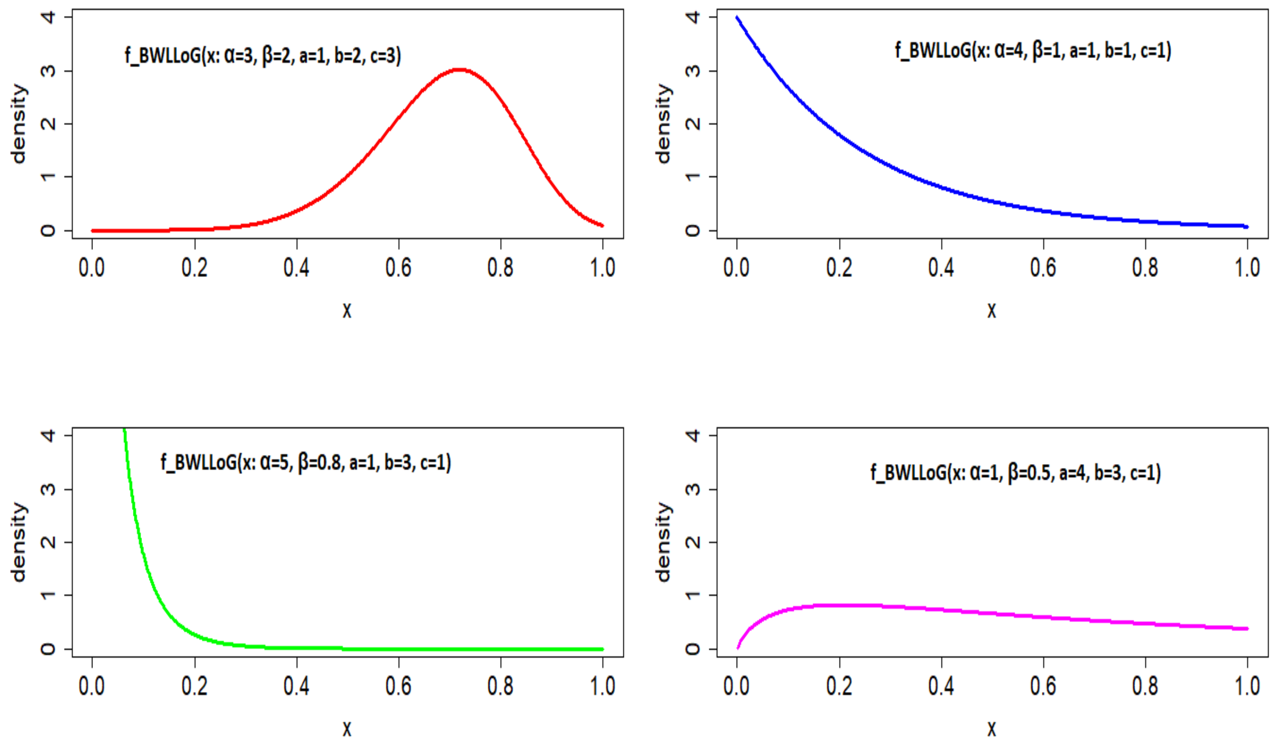

Figure 3.1: Density plots of the Beta Weibull log logistic distribution with different parameter values

A mixture of the Beta Weibull log logistic distribution will have a probability density function given by

$$
\begin{aligned}
f_{MBWLLoG}(x; a_g, b_g, \alpha_g, \beta_g, c_g) &= \sum_{g=1}^{G} \pi_g f_{BWLLoG}(x; a_g, b_g, \alpha_g, \beta_g, c_g) \\
&= \sum_{g=1}^{G} \frac{\pi_g \alpha_g \beta_g c_g}{B(a_g, b_g)} x^{c_g-1}(1+x^{c_g})^{-2} \frac{(1-(1+x^{c_g})^{-1})^{\beta_g-1}}{((1+x^{c_g})^{-1})^{\beta_g+1}} exp\left\{-\alpha_g b_g x^{c_g\beta_g}\right\} \\
&\times \left[1 - exp\left\{-\alpha_g x^{c_g\beta_g}\right\}\right]^{a_g-1}.
\end{aligned}
$$

The density plots of mixtures of the Beta Weibull log logistic distribution is given in figure 3.2.

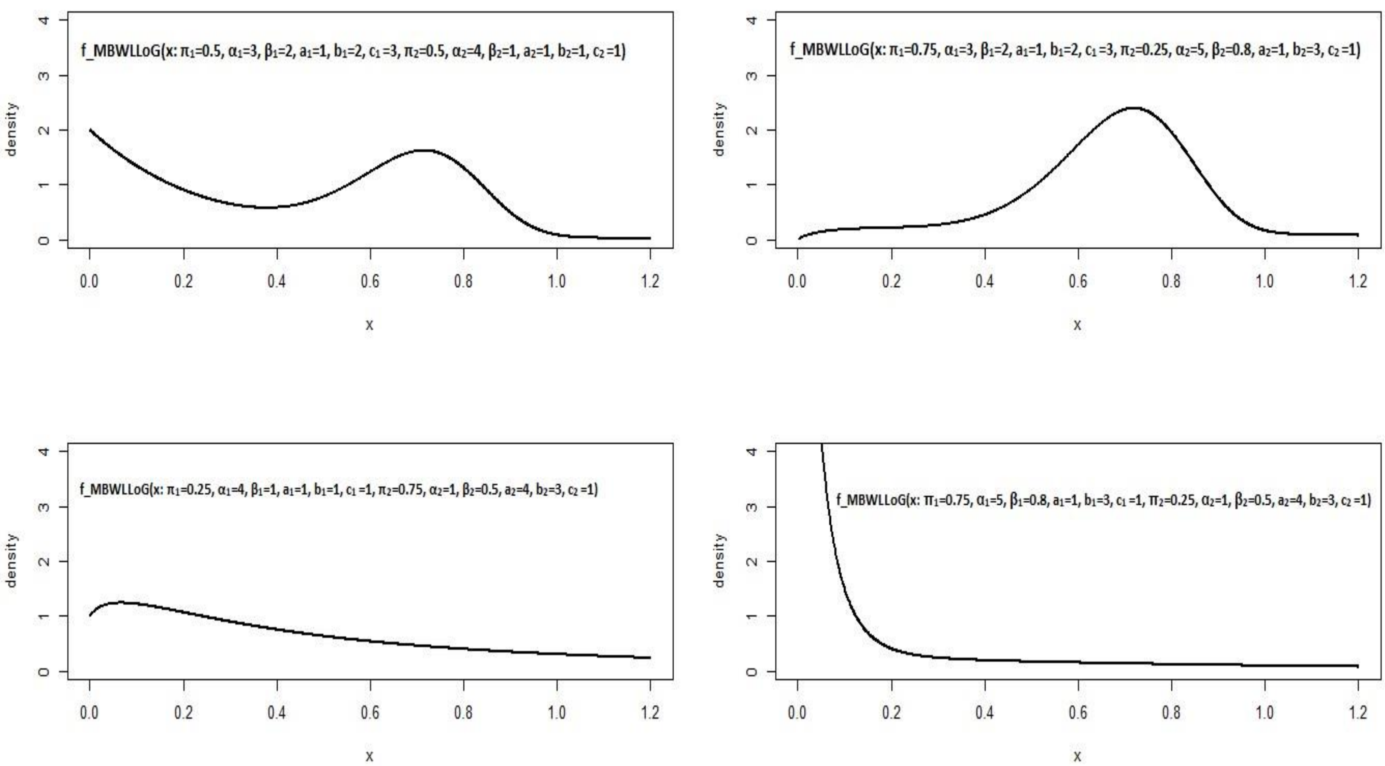

Figure 3.2: Density plots of mixtures of the Beta Weibull log logistic distribution with different parameter values

The corresponding complete log likelihood function of the BWLLoG mixture is defined by;

$$
\begin{aligned}
\ell\ell_c(\mathbf{\Theta} \mid x, \mathbf{z}) &:= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}[log\pi_g + log f_{BWLLoG}(x_i; a_g, b_g, \alpha_g, \beta_g, c_g)] \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(log\left(\frac{\alpha_g\beta_g}{B(a_g, b_g)}\right) + log(c_g x_i^{c_g-1}(1+x_i^{c_g})^{-2}) + log\left(\frac{(1-(1+x_i^{c_g})^{-1})^{\beta_g-1}}{((1+x_i^{c_g})^{-1})^{\beta_g+1}}\right) \\
&- \alpha_g b_g x_i^{c_g\beta_g} + (a_g-1)log\left[1-exp\left\{-\alpha_g x_i^{c_g\beta_g}\right\}\right]\Big) + \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig} log(\pi_g).
\end{aligned} \tag{3.20}
$$

The partial derivatives of the complete log likelihood equation of the BWLLoG mixture are calculated in appendix 5.2 and given as

$$
\frac{\partial\ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial\alpha_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\alpha_g} - b_g x_i^{c_g\beta_g} + (a_g-1)[1-exp\left\{-\alpha_g x_i^{c_g\beta_g}\right\}]^{-1} x_i^{c_g\beta_g} exp\left\{-\alpha_g x_i^{c_g\beta_g}\right\}\Big), \tag{3.21}
$$

$$
\begin{aligned}
\frac{\partial\ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial\beta_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\beta_g} + log(x_i^{c_g}) - \alpha_g b_g x_i^{c_g\beta_g} log(x_i^{c_g}) \\
&+ (a_g-1)\alpha_g x_i^{c_g\beta_g} log(x_i^{c_g}) exp\left\{-\alpha_g(x_i^{c_g\beta_g})\right\}[1-exp\left\{-\alpha_g x_i^{c_g\beta_g}\right\}]^{-1}\Big),
\end{aligned} \tag{3.22}
$$

$$
\frac{\partial\ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial a_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(a_g) + log\left[1-exp\left\{-\alpha_g x_i^{c_g\beta_g}\right\}\right]\Big), \tag{3.23}
$$

$$
\frac{\partial\ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial b_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(b_g) - \alpha_g x_i^{c_g\beta_g}\Big), \tag{3.24}
$$

$$
\begin{aligned}
\frac{\partial\ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial c_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\left[(b_g\alpha_g\beta_g x_i^{\beta_g c_g} log(x_i) - \beta_g log(x_i))c_g - 1\right] exp\left\{\alpha_g x_i^{c_g\beta_g}\right\} \\
&+ ((-b_g - a_g + 1)\alpha_g\beta_g x_i^{c_g\beta_g} log(x_i) + \beta_g log(x_i))c_g + 1)(c_g\left[1-exp\left\{\alpha_g x_i^{c_g\beta_g}\right\}\right])^{-1}\Big),
\end{aligned} \tag{3.25}
$$

$$
\frac{\partial\ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial\pi_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} \frac{z_{ig}}{\pi_g} \tag{3.26}
$$

and

$$\frac{\partial \ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial z_g} = \sum_{i=1}^{n}\sum_{g=1}^{G}\Big(log\left(\frac{\alpha_g\beta_g}{B(a_g,b_g)}\right) + log(c_g x_i^{c_g-1}(1+x_i^{c_g})^{-2}) + log\left(\frac{(1-(1+x_i^{c_g})^{-1})^{\beta_g-1}}{((1+x_i^{c_g})^{-1})^{\beta_g+1}}\right)$$
$$- \alpha_g b_g x_i^{c_g\beta_g} + (a_g-1)log\left[1-exp\left\{-\alpha_g x^{c_g\beta_g}\right\}\right]\Big) + \sum_{i=1}^{n}\sum_{g=1}^{G} log(\pi_g). \tag{3.27}$$

The partial derivatives yield parameter estimates for the BWLLoG mixture via the EM algorithm given in algorithm 1.

Notation for full models

The model $BWLLoG$ should be interpreted as a full mixture model of BWLLoG distributions with parameters α, β , a , b and c. These parameters are unconstrained. Model $Weibull$ should be interpreted as a full mixture model of Weibull distributions with parameters α and β. It should be noted that the Weibull mixtures do not have parameters a , b and c hence $-, -, -$.

Notation for constrained models

Model UC_2U should be interpreted as a model that is unconstrained for α, constrained for β (such that we have the value for β being calculated from group2 and fixed across all component distributions) and unconstrained for c.

Model C_1UC_2 should be interpreted as a model that is constrained for α (such that we have the value for α being calculated from group1 and fixed across all component distributions), unconstrained for β and constrained for c (such that we have the value for c being calculated from group2 and fixed across all component distributions). The parameters a and b are fixed to a constant value hence f. The afore notations will be used in chapter 4.

Model Based Clustering with BWLLoG Mixtures

The parameters estimated from the EM algorithm for BWLLoG mixtures will be used to carry out a model based clustering technique. The posterior probability of an observation x_i belonging

Algorithm 1 EM algorithm for BWLLoG mixtures

Input $x = \{x_1, x_2, ..., x_n\}$
Set $k = 1$ and $\ell^{(0)} = 0$

Set

$$\ell^{(s)} = \ell\ell_c(a_g^{(s)}, b_g^{(s)}, \alpha_g^{(s)}, \beta_g^{(s)}, c_g^{(s)} \mid x, \mathbf{z}) = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}[log\pi_g^{(s)} + logf_{BWLLoG}(x_i; a_g^{(s)}, b_g^{(s)}, \alpha_g^{(s)}, \beta_g^{(s)}, c_g^{(s)})]$$

Initialize $a_g^{(s)}, b_g^{(s)}, \alpha_g^{(s)}, \beta_g^{(s)}, c_g^{(s)}$ and $\pi_g^{(s)}$ via k means algorithm and Newton Raphson algorithm.

Repeat
E step;

$$z_{ig}^{(s)} := \frac{\hat{\pi}_g^{(s)} f_{BWLLoG}(x_i \mid a_g^{(s)}, b_g^{(s)}, \alpha_g^{(s)}, \beta_g^{(s)}, c_g^{(s)})}{\sum_{h=1}^{G} \hat{\pi}_h^{(s)} f_{BWLLoG}(x_i \mid a_h^{(s)}, b_h^{(s)}, \alpha_h^{(s)}, \beta_h^{(s)}, c_h^{(s)})}.$$

M step; Update π_g as

$$\pi_g^{(s+1)} = \frac{1}{n}\sum \hat{z}_{ig}^{(s)},$$

Set

$$\ell\ell_c(a_g^{(s+1)}, b_g^{(s+1)}, \alpha_g^{(s+1)}, \beta_g^{(s+1)}, c_g^{(s+1)} \mid x, \mathbf{z}) =$$
$$\sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}^{(s)}[log\pi_g^{(s+1)} + logf_{BWLLoG}(x_i; a_g^{(s+1)}, b_g^{(s+1)}, \alpha_g^{(s+1)}, \beta_g^{(s+1)}, c_g^{(s+1)})]$$

Re estimate the parameters $a_g, b_g, \alpha_g, \beta_g, c_g$ with current $z_{ig}^{(s)}$ and $\pi_{ig}^{(s+1)}$ by Newton Raphson algorithm. This gives updated parameter estimates as $a_g^{(s+1)}, b_g^{(s+1)}, \alpha_g^{(s+1)}, \beta_g^{(s+1)}, c_g^{(s+1)}$

Set

$$\ell\ell_c(a_g^{(s+1)}, b_g^{(s+1)}, \alpha_g^{(s+1)}, \beta_g^{(s+1)}, c_g^{(s+1)} \mid x, \mathbf{z}) =$$
$$\sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}^{(s)}[log\pi_g^{(s+1)} + logf_{BWLLoG}(x_i; a_g^{(s+1)}, b_g^{(s+1)}, \alpha_g^{(s+1)}, \beta_g^{(s+1)}, c_g^{(s+1)})]$$

Calculate $\ell_\infty^{(s+1)} = \ell^{(s)} + \frac{1}{1-a^{(s)}}(\ell^{(s+1)} - \ell^{(s)})$ where $a^{(s)} = \frac{\ell^{(s+1)} - \ell^{(s)}}{\ell^{(s)} - \ell^{(s-1)}}$

Until $\ell_\infty^{(s+1)} - \ell^{(s)} < \epsilon$

Set $s = s + 1$

to the g^{th} group in the mixture will take calculated as

$$\hat{z}_{ig} = \frac{\hat{\pi}_g f_{BWLLoG}(x_i \mid \hat{a}_g, \hat{b}_g, \hat{\alpha}_g, \hat{\beta}_g, \hat{c}_g)}{\sum_{h=1}^{G} \hat{\pi}_h f_{BWLLoG}(x_i \mid \hat{a}_h, \hat{b}_h, \hat{\alpha}_h, \hat{\beta}_h, \hat{c}_h)}. \tag{3.28}$$

An observation will be assigned membership to a group in which it scores the highest $\hat{z}_{ig}$.

Mixture Discriminant Analysis with BWLLoG Mixtures

Similarly the parameters estimated by the EM algorithm for BWLLoG mixtures will be used to carry out a mixture discriminant analysis technique. The posterior probability of an unlabelled observation x_j belonging to the g^{th} group in the mixture will calculated as

$$\hat{z}_{jg} = \frac{\hat{\pi}_g f_{BWLLoG}(x_j \mid \hat{a}_g, \hat{b}_g, \hat{\alpha}_g, \hat{\beta}_g, \hat{c}_g)}{\sum_{h=1}^{G} \hat{\pi}_h f_{BWLLoG}(x_j \mid \hat{a}_h, \hat{b}_h, \hat{\alpha}_h, \hat{\beta}_h, \hat{c}_h)}. \tag{3.29}$$

An unlabelled observation will be assigned membership to a group in which it scores the highest $\hat{z}_{jg}$.

3.5 Mixture of the Beta Weibull Exponential Distribution

In this subsection the baseline distribution of the BWG is specified to be exponential distribution, resulting in a Beta Weibull exponential distribution (BWE). Different density plots for BWE with varying parameter values are given. Subsequently, density plots of mixtures of BWE are given to depict heterogeneous reliability data with non monotone failure rate functions. The EM algorithm for BWE mixtures is developed together with the model based clustering technique and mixture discriminant analysis rule suitable for BWE mixtures.

Consider a mixture of the Beta Weibull G family of distributions when the baseline distribution is chosen to be exponential distribution. The as pdf, cdf and survival function of the exponential distribution are given by

$$m(x, \omega) = \omega e^{-\omega x}, M(x, \omega) = 1 - e^{-\omega x} \ and \ \bar{M}(x, \omega) = 1 - (1 - e^{-\omega x}) = e^{-\omega x}$$

respectively, where ω is the rate parameter such that $\omega > 0$.

The pdf of the Beta Weibull Exponential is defined by

$$f_{BWE}(x;a,b,\alpha,\beta,\upsilon) = \frac{\alpha\beta\omega}{B(a,b)} \frac{(1-e^{-\omega x})^{\beta-1}}{(e^{-\omega x})^{\beta}} exp\left\{-\alpha b(e^{\omega x}-1)^{\beta}\right\}[1-exp\left\{-\alpha(e^{\omega x}-1)^{\beta}\right\}]^{a-1} \tag{3.30}$$

for , $x \geqslant 0, a > 0, b > 0, \alpha > 0, \beta > 0, \omega \geqslant 0$.

The density plots for Beta Weibull Exponential distribution with varying parameter values is given in figure 3.3.

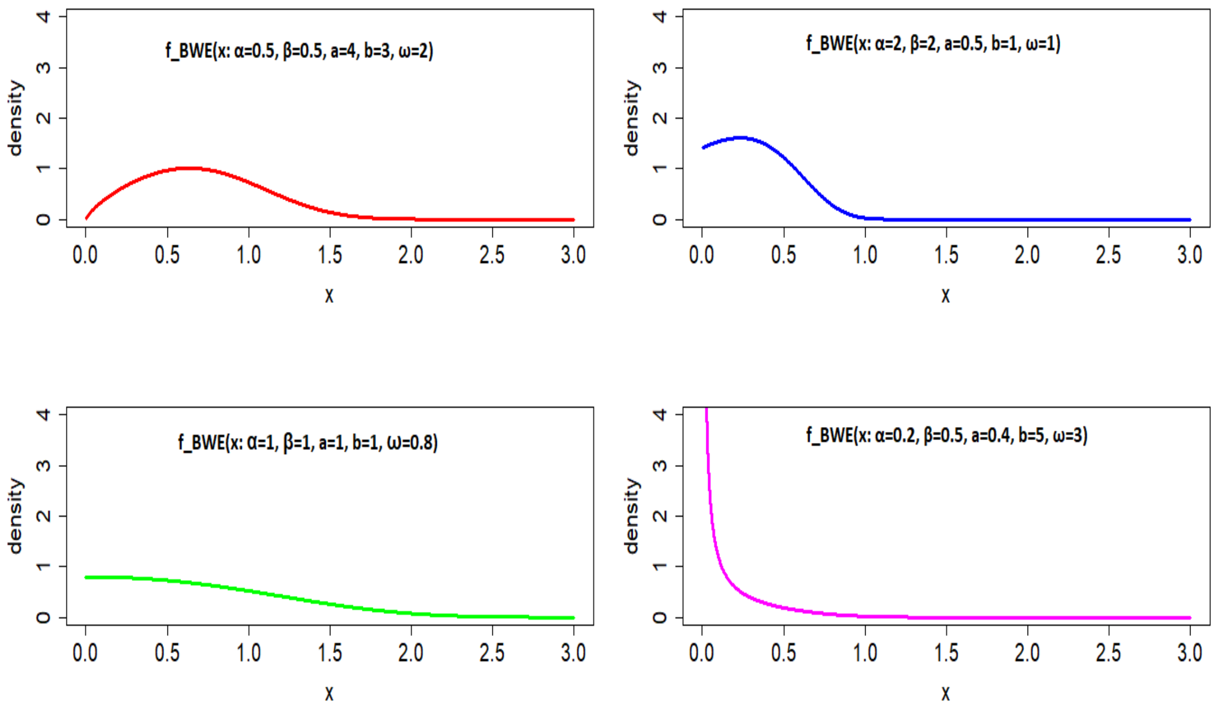

Figure 3.3: Density plots of the Beta Weibull Exponential distribution with different parameter values

A mixture of the Beta Weibull Exponential distribution will have a pdf defined by

$$
\begin{aligned}
f_{MBWE}(x; a, b, \alpha, \beta, \omega) &= \sum_{g=1}^{G} \pi_g f_{BWE}(x; a_g, b_g, \alpha_g, \beta_g, \omega_g) \\
&= \sum_{g=1}^{G} \frac{\pi_g \alpha_g \beta_g \omega_g}{B(a_g, b_g)} \frac{(1 - e^{-\omega_g x})^{\beta_g - 1}}{(e^{-\omega_g x})^{\beta_g}} exp\left\{-\alpha_g b_g (e^{\omega_g x} - 1)^{\beta_g}\right\} \\
&\times [1 - exp\left\{-\alpha_g (e^{\omega_g x} - 1)^{\beta_g}\right\}]^{a_g - 1}.
\end{aligned} \tag{3.31}
$$

The density plots for mixtures of the Beta Weibull Exponential distribution is given in figure 3.4.

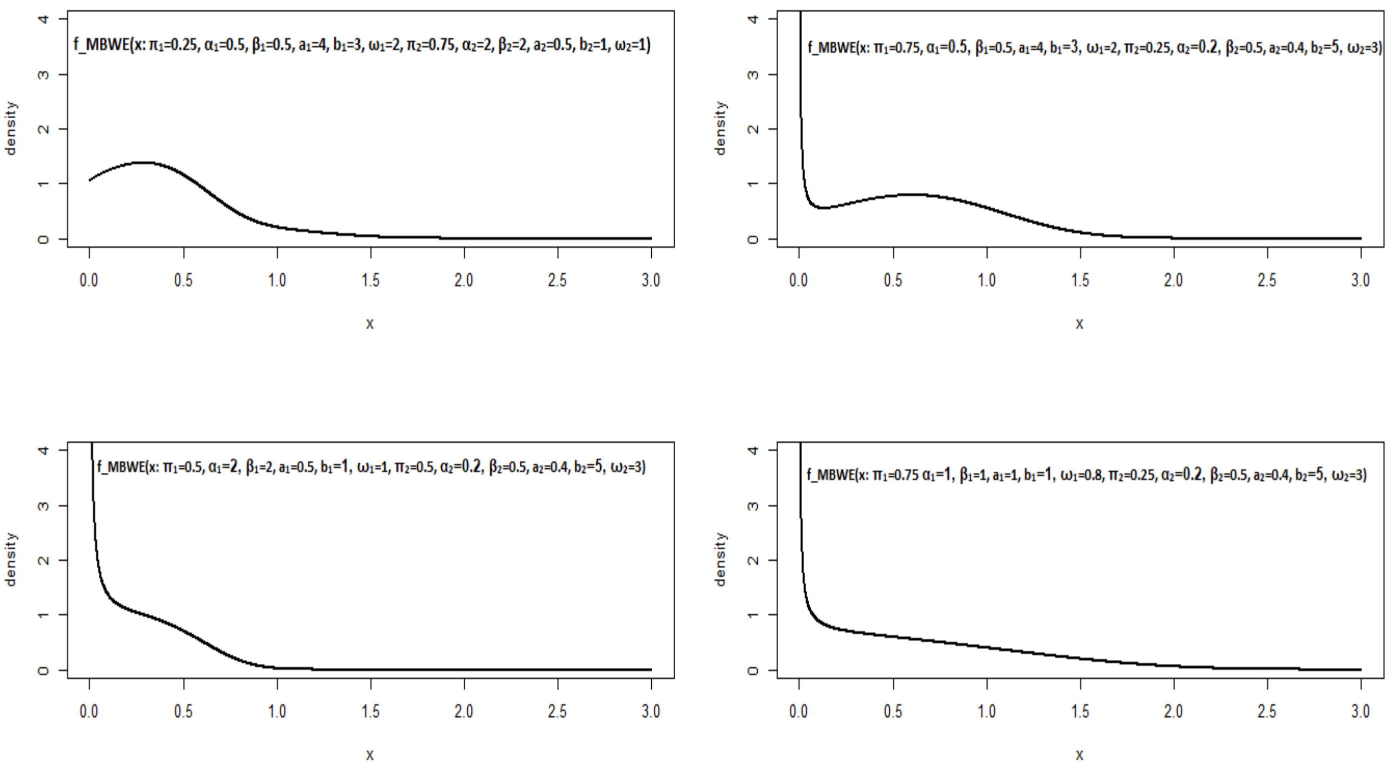

Figure 3.4: Density plots of mixtures of the Beta Weibull Exponential with different parameter values

The corresponding complete log likelihood function of the BWE mixture is defined by;

$$\ell\ell_c(\mathbf{\Theta} \mid x, \mathbf{z}) := \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}[log\pi_g + logf_{BWE}(x_i; a_g, b_g, \alpha_g, \beta_g, \omega_g)]$$
$$= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(log\left[\frac{\pi_g\alpha_g\beta_g\omega_g}{B(a_g, b_g)}\right] + log\left[\frac{(1-e^{-\omega_g x_i})^{\beta_g-1}}{(e^{-\omega_g x_i})^{\beta_g}}\right] - \alpha_g b_g(e^{\omega_g x_i} - 1)^{\beta_g}$$
$$+ (a_g - 1)log[1 - exp\{-\alpha_g(e^{\omega_g x_i} - 1)^{\beta_g}\}]\Big) + \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}log(\pi_g) \tag{3.32}$$

The partial derivatives of the complete log likelihood equation of the BWE mixture are calculated in appendix 5.3 and defined as

$$\frac{\partial\ell\ell_c(\mathbf{\Theta} \mid x, \mathbf{z})}{\partial\alpha_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\alpha_g} - b_g\left[e^{\omega_g x_i} - 1\right]^{\beta_g} + (a_g - 1)\left[1 - exp\left\{-\alpha_g\left[e^{\omega_g x_i} - 1\right]^{\beta_g}\right\}\right]^{-1}$$
$$\times \left[e^{\omega_g x_i} - 1\right]^{\beta_g} exp\left\{-\alpha_g\left[e^{\omega_g x_i} - 1\right]^{\beta_g}\right\}\Big), \tag{3.33}$$

$$\frac{\partial\ell\ell_c(\mathbf{\Theta} \mid x, \mathbf{z})}{\partial\beta_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\beta_g} + log\left[e^{\omega_g x_i} - 1\right] - \alpha_g b_g\left[e^{\omega_g x_i} - 1\right]^{\beta_g} log\left[e^{\omega_g x_i} - 1\right]$$
$$+ (a_g - 1)\alpha_g\left[e^{\omega_g x_i} - 1\right]^{\beta_g} log\left[e^{\omega_g x_i} - 1\right] exp\left\{-\alpha_g\left[e^{\omega_g x_i} - 1\right]^{\beta_g}\right\}\left[1 - exp\left\{-\alpha_g\left[e^{\omega_g x_i} - 1\right]^{\beta_g}\right\}\right]^{-1}\Big), \tag{3.34}$$

$$\frac{\partial\ell\ell_c(\mathbf{\Theta} \mid x, \mathbf{z})}{\partial a_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g + b_g) - \varphi_0(a_g) + log\left[1 - exp\{-\alpha_g[e^{\omega_g x_i} - 1]^{\beta_g}\}\right]\Big),$$

$$\frac{\partial\ell\ell_c(\mathbf{\Theta} \mid x, \mathbf{z})}{\partial b_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g + b_g) - \varphi_0(b_g) - \alpha_g[e^{\omega_g x_i} - 1]^{\beta_g}\Big),$$

$$\frac{\partial\ell\ell_c(\mathbf{\Theta} \mid x, \mathbf{z})}{\partial\omega_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1 - x_i\omega_g}{\omega_g} + \frac{(\beta_g - 1)x_i}{(e^{\omega_g x_i} - 1)} + (\beta_g + 1)x_i - \alpha_g b_g\beta_g(e^{\omega_g x_i} - 1)^{\beta_g - 1}(x_i + e^{\omega_g x_i} - 1)$$
$$+ (a_g - 1)[1 - exp\left\{-\alpha_g(e^{\omega_g x_i} - 1)^{\beta_g}\right\}]^{-1}[exp\left\{-\alpha_g(e^{\omega_g x_i} - 1)^{\beta_g}\right\}]$$
$$\times \alpha_g\beta_g(e^{\omega_g x_i} - 1)^{\beta_g - 1}[x_i + x_i(e^{\omega_g x_i} - 1)]\Big), \tag{3.35}$$

$$\frac{\partial \ell\ell_c(\boldsymbol{\Theta} \mid x, \mathbf{z})}{\partial \pi_g} = \sum_{i=1}^{n}\sum_{g=1}^{G}\frac{1}{\pi_g} + \sum_{i=1}^{n}\sum_{g=1}^{G}\frac{z_g}{\pi_g} \tag{3.36}$$

and

$$\begin{aligned}\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial z_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} log\Big(\frac{\alpha_g\beta_g}{B(a_g, b_g)}\omega_g e^{-\omega_g x_i}\frac{(1-e^{-\omega_g x_i})^{\beta_g - 1}}{(e^{-\omega_g x_i})^{\beta_g+1}} exp\left\{-\alpha_g b_g \left[e^{\omega_g x_i} - 1\right]^{\beta_g}\right\} \\ &\times \left[1 - exp\left\{-\alpha_g \left[e^{\omega_g x_i} - 1\right]^{\beta_g}\right\}\right]^{a_g - 1}\Big) + \sum_{i=1}^{n}\sum_{g=1}^{G} log(\pi_g). \end{aligned} \tag{3.37}$$

The partial derivatives will yield parameter estimates for the BWE mixture via the EM algorithm in algorithm 2.

Model Based Clustering with BWE Mixtures

The parameters estimated from the EM algorithm for BWE mixtures will be used to carry out a model based clustering technique. The posterior probability of an observation x_i belonging to the g^{th} group in the mixture will be calculated as

$$\hat{z}_{ig} = \frac{\hat{\pi}_g f_{BWE}(x_i \mid \hat{a}_g, \hat{b}_g, \hat{\alpha}_g, \hat{\beta}_g, \hat{\omega}_g)}{\sum_{h=1}^{G}\hat{\pi}_h f_{BWE}(x_i \mid \hat{a}_h, \hat{b}_h, \hat{\alpha}_h, \hat{\beta}_h, \hat{\omega}_h)}. \tag{3.38}$$

An observation will be assigned membership to a group in which it scores the highest $\hat{z}_{ig}$.

Mixture Discriminant Analysis with BWE Mixtures

Similarly the parameter estimated via the hybrid EM algorithm for BWE mixtures will be used to carry out a mixture discriminant analysis technique. The posterior probability of an labelled observation x_j belonging to the g^{th} group in the mixture will be calculated as

$$\hat{z}_{jg} = \frac{\hat{\pi}_g f_{BWE}(x_j \mid \hat{a}_g, \hat{b}_g, \hat{\alpha}_g, \hat{\beta}_g, \hat{\omega}_g)}{\sum_{h=1}^{G}\hat{\pi}_h f_{BWE}(x_j \mid \hat{a}_h, \hat{b}_h, \hat{\alpha}_h, \hat{\beta}_h, \hat{\omega}_h)}. \tag{3.39}$$

An observation will be assigned membership to a group in which it scores the highest $\hat{z}_{jg}$.

Algorithm 2 EM algorithm for BWE mixtures

Input $x = \{x_1, x_2, ..., x_n\}$
Set $k = 1$ and $\ell^{(0)} = 0$

Set

$$\ell^{(s)} = \ell\ell_c(a_g^{(s)}, b_g^{(s)}, \alpha_g^{(s)}, \beta_g^{(s)}, \omega_g^{(s)} \mid x, \mathbf{z}) = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}[log\pi_g^{(s)} + logf_{BWE}(x_i; a_g^{(s)}, b_g^{(s)}, \alpha_g^{(s)}, \beta_g^{(s)}, \omega_g^{(s)})]$$

Initialize $a_g^{(s)}, b_g^{(s)}, \alpha_g^{(s)}, \beta_g^{(s)}, \omega_g^{(s)}$ and $\pi_g^{(s)}$ via k means algorithm and Newton Raphson algorithm.

Repeat
E step;

$$z_{ig}^{(s)} := \frac{\hat{\pi}_g^{(s)} f_{BWE}(x_i \mid a_g^{(s)}, b_g^{(s)}, \alpha_g^{(s)}, \beta_g^{(s)}, \omega_g^{(s)})}{\sum_{h=1}^{G} \hat{\pi}_h^{(s)} f_{BWE}(x_i \mid a_h^{(s)}, b_h^{(s)}, \alpha_h^{(s)}, \beta_h^{(s)}, \omega_h^{(s)})}.$$

M step;
Update π_g as

$$\pi_g^{(s+1)} = \frac{1}{n}\sum \hat{z}_{ig}^{(s)},$$

Set

$$\ell\ell_c(a_g^{(s+1)}, b_g^{(s+1)}, \alpha_g^{(s+1)}, \beta_g^{(s+1)}, \omega_g^{(s+1)} \mid x, \mathbf{z}) = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}^{(s)}[log\pi_g^{(s+1)} + logf_{BWE}(x_i; a_g^{(s+1)}, b_g^{(s+1)}, \alpha_g^{(s+1)}, \beta_g^{(s+1)}, \omega_g^{(s+1)})]$$

Re estimate the parameters $a_g, b_g, \alpha_g, \beta_g, \omega_g$ with current $z_{ig}^{(s)}$ and $\pi_{ig}^{(s+1)}$ by Newton Raphson algorithm. This gives updated parameter estimates as $a_g^{(s+1)}, b_g^{(s+1)}, \alpha_g^{(s+1)}, \beta_g^{(s+1)}, \omega_g^{(s+1)}$

Set

$$\ell\ell_c(a_g^{(s+1)}, b_g^{(s+1)}, \alpha_g^{(s+1)}, \beta_g^{(s+1)}, \omega_g^{(s+1)} \mid x, \mathbf{z}) = \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}^{(s)}[log\pi_g^{(s+1)} + logf_{BWE}(x_i; a_g^{(s+1)}, b_g^{(s+1)}, \alpha_g^{(s+1)}, \beta_g^{(s+1)}, \omega_g^{(s+1)})]$$

Calculate $\ell_\infty^{(s+1)} = \ell^{(s)} + \frac{1}{1-a^{(s)}}(\ell^{(s+1)} - \ell^{(s)})$ where $a^{(s)} = \frac{\ell^{(s+1)} - \ell^{(s)}}{\ell^{(s)} - \ell^{(s-1)}}$

Until $\ell_\infty^{(s+1)} - \ell^{(s)} < \epsilon$

Set $s = s + 1$

Chapter 4

Discussion of results

In this chapter six data sets are simulated to mimic reliability data with non monotone failure rates functions. These data sets were used to mimic heterogeneous reliability data with two classes. Additionally, one real life reliability data set is provided for analyses. Results for fitting the full BWLLoG and BWE mixture models discussed in chapter 3 are provided. Corresponding results for fitting constrained mixture models of the BWLLoG and BWE family of distributions are also presented. The utility of the proposed model based clustering and mixture discriminant analysis approaches is demonstrated.

Notations

U represents an unconstrained parameter.
C_1 represents a parameter constrained to the density associated with the first component.
C_2 represents a parameter constrained to the density associated with the second component.
f represents a parameter that is fixed to a constant value.

4.1 Mixture models for simulated BWLLoG distributions (A-B data)

The acceptance rejection method was used to simulate 100 data points from the BWLLoG distribution with parameters $\alpha = 5, \beta = 3, a = 2, b = 4, c = 2$. Similarly another set of 100 data points from the BWLLoG distribution with parameters $\alpha = 2$, $\beta = 1, a = 3, b = 2, c = 1$

were simulated. These data points were assigned to group A and B respectively. A random sample of n=50 was selected from each group, forming a heterogeneous population (A & B). A two component mixture of the BWLLoG distribution was fitted to the heterogeneous population. The mixing proportions for group A and group B were 0.5. The EM algorithm for BWLLoG mixtures was used for parameter estimation. The data points from the random sample (A & B) were given class labels using the model based clustering technique for BWLLoG mixtures. A two component Weibull mixture model was also fitted to the data points and class labels were assigned using the model based clustering technique for Weibull mixtures.

The heterogeneous population (A & B) was used to carry out mixture discriminant analysis. The training to test ratios were chosen to be 70:30, 80:20 and 90:10 respectively. The data points in the training sets were used to infer class labels of the data points in the test sets using mixture discriminant analysis technique for the BWLLoG mixtures and mixture discriminant analysis technique for Weibull mixtures.

Table 4.1 shows the BIC, ARI for model based clustering and ARI for mixture discriminant analysis for two component full mixture of Weibull and BWLLoG distribution where the data is from a sample of heterogeneous simulated Beta Weibull log logistic distributions (A & B). The two component full mixture of BWLLoG distributions has a higher BIC value as compared to the two component mixture of Weibull distributions. This is an indication that the two component full BWLLoG mixture fits the A & B data better than the Weibull mixture. The two component full mixture of BWLLoG distributions has a higher value of the ARI for both MBC and MDA as compared to the two component mixture of Weibull distributions. This indicates that the two component full mixture of BWLLoG distributions performs better than the two component mixture of Weibull distributions in model based classification of the A & B data.

Table 4.1: Two component full mixtures for BWLLoG (A & B)

MODEL	α	β	a	b	c	BIC	ARI for MBC	ARI for MDA 70:30	ARI for MDA 80:20	ARI for MDA 90:10
Weibull	U	U	-	-	-	113.08	0.57	0.60	0.66	0.81
BWLLoG	U	U	U	U	U	118.01	0.86	0.58	0.84	1.00

4.2 Constrained mixture models for simulated BWLLoG distributions (A-B data)

To curb the problem of identifiability, some constraints were set on parameters a and b such that $a = b = 2.5$. The remaining parameters were constrained or unconstrained to come up with possible parsimonious models.

Table 4.2 shows the best parsimonious models based on the BIC, ARI for model based clustering and ARI for mixture discriminant analysis where the data is from the simulated Beta Weibull log logistic distributions (A & B). The two component constrained mixtures of BWLLoG distributions also have higher values of the BIC than the two component mixture of Weibull distributions. These indicate that the two component constrained mixtures of BWLLoG distributions fit the A & B data better than the two component mixture of Weibull distributions. The two component constrained mixtures of BWLLoG distributions also perform better in model based classification of the A & B data than the two component mixture of Weibull distributions. This is indicated by high values of the ARI in both MBC and MDA using the two component constrained mixtures of BWLLoG distributions as compared to using the two component mixture of Weibull distributions.

Table 4.2: Two component constrained mixtures for BWLLoG (A & B)

MODEL	α	β	a	b	c	BIC	ARI for MBC	ARI for MDA 70:30	ARI for MDA 80:20	ARI for MDA 90:10
UC_2U	U	C_2	f	f	U	116.5	0.74	0.54	0.67	0.92
UUC_2	U	U	f	f	C_2	196.1	0.70	0.55	0.74	0.83
C_1UC_2	C_1	U	f	f	C_2	198.5	0.60	0.58	0.71	0.86

4.3 Mixture models for simulated BWLLoG distributions (B-C data)

The acceptance rejection method was used to simulate 100 data points from the BWLLoG distribution with parameters $\alpha = 2$, $\beta = 1, a = 3, b = 2, c = 1$. Similarly another set of 100 data points from the BWLLoG distribution with parameters $\alpha = 5$, $\beta = 3, a = 2, b = 4, c = 2$ were simulated. These data points were assigned to group B and C respectively. A random sample of n=75 was selected from each group, forming a heterogeneous population (B & C). A two component mixture of the BWLLoG distribution was fitted to the heterogeneous population. The mixing proportions for group B and group C were 0.5. The EM algorithm for BWLLoG mixtures was used for parameter estimation. The data points from the random sample (B & C) were given class labels using the model based clustering technique for BWLLoG mixtures. These data points were also be fitted to the Weibull mixtures and assigned class labels with model based clustering technique for Weibull mixtures.

The heterogeneous population (B & C) was used to carry out mixture discriminant analysis. The training to test ratios were chosen to be 70:30, 80:20 and 90:10 respectively. The data points in the training sets were used to infer class labels of the data points in the test sets using mixture discriminant analysis technique for the BWLLoG mixtures and mixture discriminant analysis technique for Weibull mixtures.

Table 4.3 shows the BIC, ARI for model based clustering and ARI for mixture discriminant analysis for two component full mixtures of Weibull and BWLLoG distributions where the data is from a sample of heterogeneous simulated Beta Weibull log logistic distributions (B & C). The two component full mixture of BWLLoG distributions has a higher BIC value as compared to the two component mixture of Weibull distributions. This is an indication that the BWLLoG mixture fits the B & C data better than the two component Weibull mixture. The two component full mixture of BWLLoG distributions has a higher value of the ARI for both MBC and MDA as compared to the two component full mixture of Weibull distributions. This indicates that the two component full mixture of BWLLoG distributions performs better than the two component mixture of Weibull distributions in model based classification of the B & C data.

Table 4.3: Two component full mixtures for BWLLoG (B & C)

MODEL	α	β	a	b	c	BIC	ARI for MBC	ARI for MDA 70:30	ARI for MDA 80:20	ARI for MDA 90:10
Weibull	U	U	-	-	-	156.16	0.65	0.72	0.74	0.82
BWLLoG	U	U	U	U	U	133.04	0.78	0.66	0.85	0.86

Table 4.4: Two component constrained mixtures for BWLLoG (B & C)

MODEL	α	β	a	b	c	BIC	ARI for MBC	ARI for MDA 70:30	ARI for MDA 80:20	ARI for MDA 90:10
UC_2U	U	C_2	f	f	U	110.7	0.72	0.66	0.78	0.92
UUC_2	U	U	f	f	C_2	160.7	0.68	0.60	0.79	0.82
C_1UC_2	C_1	U	f	f	C_2	160.3	0.70	0.62	0.76	0.89

4.4 Constrained mixture models for simulated BWLLoG distributions (B-C data)

To curb the problem of identifiability we set constraints to the parameters a and b such that $a = b = 2.5$. The remaining parameters were constrained or unconstrained to come up with possible parsimonious mixture models.

Table 4.4 shows the best parsimonious models based on the BIC, ARI for model based clustering and ARI for mixture discriminant analysis where the data is from the simulated Beta Weibull log logistic distributions (B & C). The two component constrained mixtures of BWLLoG distributions also have higher values of the BIC than the two component mixture of Weibull distributions. This indicates that the two component constrained mixtures of BWLLoG distributions fit the B & C data better than the two component mixture of the Weibull distributions. The two component constrained mixtures of BWLLoG distributions also perform better in model based classification of the B & C data than the two component mixture of Weibull distributions. This is indicated by high values of the ARI for both MBC and MDA using the two component constrained mixtures of BWLLoG distributions as compared to using the two component mixture of Weibull distributions.

4.5 Mixture models for BWLLoG distributions (Yarn data)

The data set used in this analysis is from Shanker, Fasshaye and Selvaraj (2015). It is a reliability data set corresponding to time to failure of polyester yarn in a textile experiment for testing the tensile fatigue characteristics of yarn. It has 100 observations of yarns exposed to 2.5 % strain level. The k-means algorithm was used to split the yarn data set into two sub-populations being sub-population 1 and sub-population 2.

A two component mixture of BWLLoG distribution was fitted. Parameter estimation was carried out by the EM algorithm for BWLLoG mixtures and the observations in the random sample were assigned class labels using the model based clustering technique for BWLLoG mixtures. These data points were also fitted to Weibull mixtures and assigned class labels with model based clustering technique for Weibull mixtures.

The yarn data composing of sub-population 1 and sub-population 2 was used to carry out mixture discriminant analysis. The training to test ratios ere chosen to be 70:30, 80:20 and 90:10 respectively. The data points in the training sets were used to infer class labels of the data points in the test sets using mixture discriminant analysis technique for the BWLLoG mixtures and mixture discriminant analysis technique for Weibull mixtures.

Table 4.5 shows the BIC, ARI for model based clustering and ARI for mixture discriminant analysis for two component full mixture of Weibull and BWLLoG distributions where the yarn data used is from Shanker, Fasshaye and Selvaraj (2015). The two component full mixture of BWLLoG distributions has a higher BIC value as compared to the two component mixture of Weibull distributions. This is an indication that the BWLLoG mixture fits the yarn data better than the Weibull mixture. The two component full mixture of BWLLoG distributions has a higher value of the ARI for both MBC and MDA as compared to the two component full mixture of Weibull distributions. This indicates that the two component full mixture of BWLLoG distributions performs better than the two component mixture of Weibull distributions in model based classification of the yarn data.

Table 4.5: Two component full mixtures for BWLLoG (Yarn data)

MODEL	α	β	a	b	c	BIC	ARI for MBC	ARI for MDA 70:30	ARI for MDA 80:20	ARI for MDA 90:10
Weibull	U	U	-	-	-	1053.09	0.57	0.64	0.68	0.73
BWLLoG	U	U	U	U	U	1077.94	0.87	0.57	0.60	0.79

4.6 Constrained mixture models for BWLLoG distributions (Yarn data)

To curb the problem of identifiability constraints were set on parameters a and b such that $a = b = 2.5$. The remaining parameters were constrained or unconstrained to come up with possible parsimonious models.

Table 4.6 shows the best parsimonious models based on the BIC, ARI for model based clustering and ARI for mixture discriminant analysis where the yarn data used is from Shanker, Fasshaye and Selvaraj (2015). The two component constrained mixtures of BWLLoG distributions also have higher values of the BIC than the two component mixture of Weibull distributions. This indicates that the two component constrained mixtures of BWLLoG distributions fit the yarn data better than the two component mixture of Weibull distributions. The two component constrained mixtures of BWLLoG distributions also perform better in model based classification of the yarn data than the two component mixture of the Weibull distributions. This is indicated by high values of the ARI for both MBC and MDA using the two component constrained mixtures of BWLLoG distributions as compared to using the two component mixture of Weibull distributions.

Table 4.6: Two component constrained mixtures for BWLLoG (Yarn data)

MODEL	α	β	a	b	c	BIC	ARI for MBC	ARI for MDA 70:30	ARI for MDA 80:20	ARI for MDA 90:10
UC_1C_1	U	C_1	f	f	C_1	1448.3	0.76	0.55	0.60	0.76
C_1UC_1	C_1	U	f	f	C_1	1409.2	0.60	0.58	0.64	0.79
C_2C_1U	C_2	C_1	f	f	U	2220.5	0.70	0.53	0.77	0.81

4.7 Mixture models for simulated BWE distributions (A-B data)

The acceptance rejection method was used to simulate 100 data points from the BWE distribution with parameters $\alpha = 1, \beta = 2, a = 1, b = 2, \omega = 2$. Similarly another set of 100 data points from the BWE distribution with parameters $\alpha = 3$, $\beta = 1, a = 2, b = 1, \omega = 1$ were simulated. These data points were assigned to group A and B respectively. A random sample of n=50 was selected from each group, forming a heterogeneous population (A & B). A two component mixture of the BWE distribution was fitted to the heterogeneous population. The mixing proportions for group A and group B were 0.5. The EM algorithm for BWE mixtures was used for parameter estimation. The data points from the random sample (A & B) were given class labels using the model based clustering technique for BWE mixtures. These data points were also fitted to Weibull mixtures and assigned class labels with model based clustering technique for Weibull mixtures.

The heterogeneous population (A & B) was used to carry out mixture discriminant analysis. The training to test ratios were chosen to be 70:30, 80:20 and 90:10 respectively. The data points in the training sets were used to infer class labels of the data points in the test sets using mixture discriminant analysis technique for the BWE mixtures and mixture discriminant analysis technique for Weibull mixtures.

Table 4.7 shows the BIC, ARI for model based clustering and ARI for mixture discriminant analysis for two component full mixtures of Weibull and BWE distributions where the data is from a sample of heterogeneous simulated Beta Weibull Exponential distributions (A & B). The two component full mixture of BWE distributions has a higher BIC value as compared to the two component mixture of Weibull distributions. This is an indication that the BWE mixture fits the A & B data better than the Weibull mixture. The two component full mixture of BWE distributions has a higher value of ARI for both MBC and MDA as compared to the two component mixture of Weibull distributions. This indicates that the two component mixture of BWE distributions performs better than the two component mixture of Weibull distributions in model based classification of the A & B data.

Table 4.7: Two component full mixtures for BWE (A & B)

MODEL	α	β	a	b	ω	BIC	ARI for MBC	ARI for MDA 70:30	ARI for MDA 80:20	ARI for MDA 90:10
Weibull	U	U	-	-	-	30.36	0.55	0.60	0.71	0.82
BWE	U	U	U	U	U	39.38	0.64	0.53	0.73	0.91

4.8 Constrained mixture models for simulated BWE distributions (A-B data)

To curb the problem of identifiability constraints were set on parameters a and $b$such that $a = b = 2.5$. The remaining parameters were constrained or unconstrained to come up with possible parsimonious models.

Table 4.8 shows the best parsimonious models based on the BIC, ARI for model based clustering and ARI for mixture discriminant analysis where the data is from a sample of heterogeneous simulated Beta Weibull Exponential distributions (A & B). The two component constrained mixtures of BWE distributions also have higher values of the BIC than the two component mixture of Weibull distributions. This indicates that the two component constrained mixtures of BWE distributions fit the A & B data better than the two component mixture of Weibull distributions. The two component constrained mixtures of BWE distributions also perform better in model based classification of the A & B data than the two component mixture of Weibull distributions. This is indicated by high values of the ARI for both MBC and MDA using the two component constrained mixtures of BWE distributions as compared to using the two component mixture of Weibull distributions.

Table 4.8: Two component constrained mixtures for BWE (A & B)

MODEL	α	β	a	b	ω	BIC	ARI for MBC	ARI for MDA 70:30	ARI for MDA 80:20	ARI for MDA 90:10
UUC_1	U	U	f	f	C_1	30.5	0.73	0.54	0.86	0.85
C_1UC_1	C_1	U	f	f	C_1	41.9	0.60	0.59	0.83	0.95

4.9 Mixture models for simulated BWE distributions (B-C data)

The acceptance rejection method was used to simulate 100 data points from the BWE distribution with parameters $\alpha = 3$, $\beta = 1, a = 2, b = 1, \omega = 1$. Similarly another set of 100 data points from the BWE distribution with parameters $\alpha = 3$, $\beta = 1, a = 2, b = 1, \omega = 1$ were simulated. These data points were assigned to group B and C respectively. A random sample of n=75 was selected from each group, forming a heterogeneous population (B & C). A two component mixture of the BWE distributions was fitted to the heterogeneous population. The mixing proportions for group A and group B were 0.5. The EM algorithm for BWE mixtures was used for parameter estimation. The data points from the random sample (B & C) were given class labels using the model based clustering technique for BWE mixtures. These data points were also fitted to Weibull mixtures and assigned class labels with model based clustering technique for Weibull mixtures.

The heterogeneous population (B & C) was used to carry out mixture discriminant analysis. The training to test ratios were chosen to be 70:30, 80:20 and 90:10 respectively. The data points in the training sets were used to infer class labels of the data points in the test sets using mixture discriminant analysis technique for the BWLLoG mixtures and mixture discriminant analysis technique for Weibull mixtures.

Table 4.9 shows the BIC, ARI for model based clustering and ARI for mixture discriminant analysis for two component full mixtures of Weibull and BWE distributions where the data is from a sample of heterogeneous simulated Beta Weibull Exponential distributions (B & C). The two component full mixture of BWE distributions has a higher BIC value as compared to the two component mixture of Weibull distributions. This is an indication that the two component full BWE mixture fits the B & C data better than the Weibull mixture. The two component full mixture of BWE distributions has a higher ARI for both MBC and MDA as compared to the two component mixture of Weibull distributions. This indicates that the two component full mixture of BWE distributions performs better than the two component mixture of Weibull distributions in model based classification of the B & C data.

Table 4.9: Two component full mixtures for BWE (B & C)

MODEL	α	β	a	b	ω	BIC	ARI for MBC	ARI for MDA 70:30	ARI for MDA 80:20	ARI for MDA 90:10
Weibull	U	U	-	-	-	11.9	0.57	0.72	0.68	0.85
BWE	U	U	U	U	U	14.15	0.78	0.56	0.79	0.91

4.10 Constrained mixture models for simulated BWE distributions (B-C data)

To curb the problem of identifiability some constraints were set on parameters a and b such that $a = b = 2.5$. The remaining parameters where constrained or unconstrained to come up with possible parsimonious models.

Table 4.10 shows the best parsimonious models based on the BIC, ARI for model based clustering and ARI for mixture discriminant analysis were the data is from a sample of heterogeneous simulated Beta Weibull Exponential distributions (B & C). The two component constrained mixtures of BWE distributions also have higher values of the BIC than the two component mixture of Weibull distributions. This indicates that the two component constrained mixtures of BWE distributions fit the B & C data better than the two component mixture of the Weibull distributions. The two component constrained mixtures of BWE distributions also perform better in model based classification of the B & C data than the two component mixture of Weibull distributions. This is indicated by high values of the ARI for both MBC and MDA using the two component constrained mixtures of BWE distributions as compared to using the two component mixture of Weibull distributions.

Table 4.10: Two component constrained mixtures for BWE (B & C)

MODEL	α	β	a	b	ω	BIC	ARI for MBC	ARI for MDA 70:30	ARI for MDA 80:20	ARI for MDA 90:10
C_2C_1U	C_2	C_1	f	f	U	29.97	0.64	0.57	0.78	0.93
UC_2C_1	U	C_2	f	f	C_1	21.46	0.97	0.58	0.83	0.91
C_2UC_1	C_2	U	f	f	C_1	28.33	0.64	0.61	0.78	0.87

Chapter 5

Conclusion

This chapter highlights the conclusions reached from this work. It further makes suggestions for future work.

In this thesis, mixtures of the Beta Weibull G family of distributions were introduced. These mixtures were motivated by scarce literature on mixtures of reliability data. They fill the gap particularly when the underlying reliability data has hazard rates that are non monotone. A detailed account on how parameter estimation was carried out by EM algorithm was given. Model based clustering and mixture discriminant analysis within the framework of mixtures of the Beta Weibull G Family of Distributions was discussed in detail.

In chapter 3, some special cases of mixtures of the BWG family of distributions being the mixtures of Beta Weibull log logistic distribution and mixtures of Beta Weibull Exponential distribution were developed. For each special case, an EM algorithm that could be used for parameter estimation was developed. A detailed explanation on how model based clustering and mixture discriminant analysis would be carried out in these special cases was also given.

In chapter 4, different data sets were simulated to mimic different populations from the BWG family of distributions. A contrast between BWG mixtures, constrained BWG mixtures and Weibull mixtures was made based on the BIC and the ARI.

The findings of this thesis demonstrate that mixtures of the BWG family of distributions fit heterogeneous population with non monotone hazard rates better than mixtures of the Weibull distributions as evidenced by higher values of BIC for BWG mixtures. The BWG mixtures performed better than Weibull mixtures in both model based clustering and mixture discriminant analysis as evidenced by high values of the ARI.

More work can be done in comparing mixtures of generalisations of Weibull distributions to mixtures of the Weibull distributions. These could include mixtures of Weibull G family of distributions and mixtures of generalised modified Weibull distributions which model reliability data with fewer parameters than the BWG family of distributions. Other maximisation algorithms such as the Broyden-Fletcher-Goldfard-Shanno (BFGS) algorithm or it′s variations could be considered for carrying out the M-step in the EM algorithms. The BFGS is known to be more robust than Newton Raphson algorithm, so it might give better results. It could also be interesting to develop a multivariate BWG distribution and explore its utility in model based techniques.

Bibliography

[1] Aitken, A. C. (1926). A Series Formula for the Roots of Algebric Transcendental Equations. Proceedings of the Royal Society of Edinburgh, 46:14-22.

[2] Almalki, S. J. and Nadarajah, S. (2015). Comparing the Exponentiated and Generalized Modified Weibull Distributions. Journal of Data Science, 13(4):713-731.

[3] Bezdek, J. C. (2013). Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Springer Science Business Media.

[4] Bohning, D., Dietz, E., Schaub, R., Schlattmann, P., and Lindsay, B. G. (1994). The distribution of the likelihood ratio for mixtures of densities from the one-parameter Exponential family. Annals of the Institute of Statistical Mathematics, 46(2):373-388.

[5] Bucar, T., Nagode, M., and Fajdiga, M. (2004). Reliability approxi- mation using finite weibull mixture distributions. Reliability Engineering System Safety, 84(3):241-251.

[6] Chandra, S. (1977). On the mixtures of probability distributions. Scandinavian Journal of Statistics, pages 105-112.

[7] Dasgupta, A. and Raftery, A. E. (1998). Detecting features in spatial point processes with clutter via model-based clustering. Journal of the American Statistical Association, 93(441):294-302.

[8] Day, N. E. (1969). Estimating the components of a mixture of Normal distributions. Biometrika, 56(3):463-474.

[9] 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), 39(1):1-38.

[10] Fagbamigbe, A. F., Makubate, B., Huang, S., Oluyede, B. O., and Motobetso, G. (2018). The Beta Weibull-G Family of Distributions: Model, Properties and Application. International Journal of Statistics and Probability, 7(2):12-32.

[11] Hastie, T. and Tibshirani, R. (1996). Discriminant analysis by Gaussian mixtures. Journal of the Royal Statistical Society: Series B (Methodological), 58(1):155-176.

[12] Holgersson, M. and Jorner, U. (1978). Decomposition of a mixture into Normal components: A review. International Journal of Bio-Medical Computing, 9(5):367392. Hubert, L. and Arabie, P. (1985). Comparing partitions. Journal of Classification, 2(1):193-218.

[13] Kaufman, L. and Rousseeuw, P. J. (2009). Finding groups in data: An introduction to cluster analysis, volume 344. John Wiley Sons.

[14] Keribin, C. (2000). Consistent estimation of the order of mixture models. Sankhya: The Indian Journal of Statistics, Series A, pages 49-66.

[15] Ko, Y., Zhai, C., and Rodriguez-Zas, S. L. (2007). Inference of gene pathways using gaussian mixture models. In 2007 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007), pages 362367. IEEE.

[16] Lindsay, B. G. (1995). Mixture models: Theory, geometry and applications. In NSF-CBMS Regional Conference Series in Probability and Statistics, pages i-163. JSTOR.

[17] Little, R. J. and Rubin, D. B. (2014). Statistical analysis with missing data, volume 333. John Wiley Sons.

[18] McLachlan, G. J. and Peel, D. (2000). Finite Mixture Models. Wiley Series in Probability and Statistics.

[19] McNicholas, P. D. (2010). Model-Based Classification using latent Gaussian mixture models. Journal of Statistical Planning and Inference, 140(5):1175- 1181.

[20] McNicholas, P. D. (2016). Mixture Model-Based Classification. Boca Raton: Chapman and Hall/CRC.

[21] Meeker, W. Q. and Escobar, L. A. (2014). Statistical methods for reliability data. John Wiley Sons.

[22] Merovci, F. and Elbatal, I. (2015). A New Generalized of Exponentiated Modified Weibull Distribution. Journal of Data Science, 13(2):213-238.

[23] Nelson, W. (1985). Weibull Analysis of Reliability Data with few or no failures. Journal of Quality Technology, 17(3):140-146.

[24] Oluyede, B. (2018). The Gamma-Weibull-G Family of Distributions with Applications. Austrian Journal of Statistics, 47(1):45-76.

[25] Oluyede, B. O., Huang, S., and Yang, T. (2015). A New Class of Generalized Modified Weibull Distribution with Applications. Austrian Journal of Statistics, 44(3):43.

[26] Pearson, K. (1894). Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London. A:Mathematical, Physical and Engineering Sciences, 185:71-110.

[27] Qin, X., Zhang, J.-s., and Yan, X.-d. (2012). Two improved mixture weibull models for the analysis of wind speed data. Journal of applied meteorology and climatology, 51(7):1321-1332.

[28] Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association, 66(336):846-850.

[29] Razali, A. M. and Al-Wakeel, A. A. (2013). Mixture Weibull Distributions for fitting failure times data. Applied Mathematics and Computation, 58 219(24):11358-11364.

[30] Rinne, H. (2008). The Weibull distribution: a handbook. CRC press.

[31] Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2):461-464.

[32] Shin, J.-Y., Ouarda, T. B., and Lee, T. (2016). Heterogeneous mixture distributions for modeling wind speed, application to the uae. Renewable Energy, 91:40-52.

[33] Tan, W. and Chang, W. (1972). Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of two Normal densities. Journal of the American Statistical Association, 67(339):702-708.

[34] Titterington, D. M., Smith, A. F., and Makov, U. E. (1985). Statistical Analysis of Finite Mixture Distributions. Chichester: John Wiley Sons,.

[35] Weldon, W. F. R. (1892). Certain correlated variations in crangon vulgaris. Proceedings of the Royal Society of London, 51(308-314):1-21.

Appendix

5.1 The partial derivatives of the complete log likelihood function of BWG mixture

$$
\begin{aligned}
\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \alpha_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{\frac{\pi_g \beta_g}{B(a_g,b_g)}}{\frac{\pi_g \alpha_g \beta_g}{B(a_g,b_g)}} + 0 + 0 - b_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} \\
&+ \frac{(a_g-1)\left[1-exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]^{(a_g-2)}\left[-\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}}{\left[1-exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]^{(a_g-1)}} + 0\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\alpha_g} - b_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} + (a_g-1)\left[1-exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]^{-1} \\
&\times \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\Big). \qquad (5.1)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \beta_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{\frac{\pi_g \alpha_g}{B(a_g,b_g)}}{\frac{\pi_g \alpha_g \beta_g}{B(a_g,b_g)}} + 0 + log\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right] - \alpha_g b_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} log\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right] \\
&+ \frac{(a_g-1)\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} log\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right] exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}}{\left[1-exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]} + 0\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\beta_g} + log\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right] - \alpha_g b_g \left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} log\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right] \\
&+ (a_g-1)\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g} log\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right] \\
&\times exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\left[1-exp\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\}\right]^{-1}\Big). \qquad (5.2)
\end{aligned}
$$

Note that $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ *and* $\varphi_0(a) = \dfrac{\Gamma'(a)}{\Gamma(a)}$, *thus,*

$$\frac{\partial B(a,b)}{\partial a}=\frac{\partial}{\partial a}\Big(\frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}\Big)=B(a,b)\Big(\frac{\Gamma'(a)}{\Gamma(a)}-\frac{\Gamma'(a+b)}{\Gamma(a+b)}\Big)=B(a,b)[\varphi_0(a)-\varphi_0(a+b)],\ hence,$$

$$\frac{\partial}{\partial a}log\Big(\frac{\pi\alpha\beta}{B(a,b)}\Big)=\frac{\Big(\frac{\partial}{\partial a}\frac{\pi\alpha\beta}{B(a,b)}\Big)}{\Big(\frac{\pi\alpha\beta}{B(a,b)}\Big)}=\frac{\Big(\frac{-\pi\alpha\beta B(a,b)[\varphi_0(a)-\varphi_0(a+b)]}{B^2(a,b)}\Big)}{\Big(\frac{\pi\alpha\beta}{B(a,b)}\Big)}=\frac{\Big(\frac{-\pi\alpha\beta[\varphi_0(a)-\varphi_0(a+b)]}{B(a,b)}\Big)}{\Big(\frac{\pi\alpha\beta}{B(a,b)}\Big)}.$$

$$\begin{aligned}\frac{\partial\ell\ell_c(\mathbf{\Theta},\mathbf{z}\mid x)}{\partial a_g}&=\sum_{i=1}^{n}\sum_{g=1}^{G}z_{ig}\Big(\frac{\Big(\frac{-\pi_g\alpha_g\beta_g[\varphi_0(a_g)-\varphi_0(a_g+b_g)]}{B(a_g,b_g)}\Big)}{\Big(\frac{\pi_g\alpha_g\beta_g}{B(a_g,b_g)}\Big)}+0+0+0+log\left[1-exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]\Big)\\&=\sum_{i=1}^{n}\sum_{g=1}^{G}z_{ig}\Big(-[\varphi_0(a_g)-\varphi_0(a_g+b_g)]+log\left[1-exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]\Big)\\&=\sum_{i=1}^{n}\sum_{g=1}^{G}z_{ig}\Big(\varphi_0(a_g+b_g)-\varphi_0(a_g)+log\left[1-exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]\Big).\end{aligned}\tag{5.3}$$

$$\begin{aligned}\frac{\partial\ell\ell_c(\mathbf{\Theta},\mathbf{z}\mid x)}{\partial b_g}&=\sum_{i=1}^{n}\sum_{g=1}^{G}z_{ig}\Big(\frac{\frac{-\pi_g\alpha_g\beta_g[\varphi_0(b_g)-\varphi_0(a_g+b_g)]}{B(a_g,b_g)}}{\frac{\pi_g\alpha_g\beta_g}{B(a_g,b_g)}}+0+0-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}+0\Big)+0\\&=\sum_{i=1}^{n}\sum_{g=1}^{G}z_{ig}\Big(\varphi_0(a_g+b_g)-\varphi_0(b_g)-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\Big).\end{aligned}\tag{5.4}$$

$$
\begin{aligned}
\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \upsilon_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(0 + \frac{1}{m(x_i,\upsilon_g)}\frac{\partial}{\partial \upsilon_g} m(x_i,\upsilon_g) + \frac{(\beta_g-1)}{M(x_i,\upsilon_g)}\frac{\partial}{\partial \upsilon_g} M(x_i,\upsilon_g) - \frac{(\beta_g+1)}{\bar{M}(x_i,\upsilon_g)}\frac{\partial}{\partial \upsilon_g}\bar{M}(x_i,\upsilon_g) \\
&- \alpha_g b_g \beta_g \left(\frac{M(x_i,\upsilon_g}{\bar{M}(x_i,\upsilon_g}\right)^{\beta_g-1}\left(\frac{1}{\bar{M}(x_i,\upsilon_g)}\frac{\partial}{\partial \upsilon_g}M(x_i,\upsilon_g) - \frac{M(x_i,\upsilon_g)}{[\bar{M}(x_i,\upsilon_g)]^2}\frac{\partial}{\partial \upsilon_g}\bar{M}(x_i,\upsilon_g)\right) \\
&+ (a_g-1)\left[1 - exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]^{-1}\left[exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right] \\
&\times \left[\alpha_g\beta_g\left(\frac{M(x_i,\upsilon_g}{\bar{M}(x_i,\upsilon_g}\right)^{\beta_g-1}\left(\frac{1}{\bar{M}(x_i,\upsilon_g)}\frac{\partial}{\partial \upsilon_g}M(x_i,\upsilon_g) - \frac{M(x_i,\upsilon_g)}{[\bar{M}(x_i,\upsilon_g)]^2}\frac{\partial}{\partial \upsilon_g}\bar{M}(x_i,\upsilon_g)\right)\right]\Big) + 0 \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{m(x_i,\upsilon_g)}\frac{\partial}{\partial \upsilon_g} m(x_i,\upsilon_g) + \frac{(\beta_g-1)}{M(x_i,\upsilon_g)}\frac{\partial}{\partial \upsilon_g} M(x_i,\upsilon_g) - \frac{(\beta_g+1)}{\bar{M}(x_i,\upsilon_g)}\frac{\partial}{\partial \upsilon_g}\bar{M}(x_i,\upsilon_g) \\
&- \alpha_g b_g \beta_g \left(\frac{M(x_i,\upsilon_g}{\bar{M}(x_i,\upsilon_g}\right)^{\beta_g-1}\left(\frac{1}{\bar{M}(x_i,\upsilon_g)}\frac{\partial}{\partial \upsilon_g}M(x_i,\upsilon_g) - \frac{M(x_i,\upsilon_g)}{[\bar{M}(x_i,\upsilon_g)]^2}\frac{\partial}{\partial \upsilon_g}\bar{M}(x_i,\upsilon_g)\right) \\
&+ (a_g-1)\left[1 - exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]^{-1}\left[exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right] \\
&\times \left[\alpha_g\beta_g\left(\frac{M(x_i,\upsilon_g}{\bar{M}(x_i,\upsilon_g}\right)^{\beta_g-1}\left(\frac{1}{\bar{M}(x_i,\upsilon_g)}\frac{\partial}{\partial \upsilon_g}M(x_i,\upsilon_g) - \frac{M(x_i,\upsilon_g)}{[\bar{M}(x_i,\upsilon_g)]^2}\frac{\partial}{\partial \upsilon_g}\bar{M}(x_i,\upsilon_g)\right)\right]\Big).
\end{aligned}
\tag{5.5}
$$

$$
\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \pi_g} = \sum_{i=1}^{n}\sum_{g=1}^{G} \frac{z_{ig}}{\pi_g}.
$$

$$
\begin{aligned}
\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial z_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} log\Big(\frac{\alpha_g\beta_g}{B(a_g,b_g)} m(x_i;\upsilon_g)\frac{M(x_i;\upsilon_g)^{\beta_g-1}}{\bar{M}(x_i;\upsilon_g)^{\beta_g+1}} exp\left\{-\alpha_g b_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\} \\
&\times \left[1 - exp\left\{-\alpha_g\left[\frac{M(x_i;\upsilon_g)}{\bar{M}(x_i;\upsilon_g)}\right]^{\beta_g}\right\}\right]^{a_g-1}\Big) + \sum_{i=1}^{n}\sum_{g=1}^{G} log(\pi_g).
\end{aligned}
\tag{5.6}
$$

5.2 The partial derivatives of the complete log likelihood of the BWLLoG mixture

We have that

$$m(x_i, c_g) = c_g x_i^{c_g-1}(1+x_i^{c_g})^{-2}, M(x_i, c_g) = 1-(1+x_i^{c_g})^{-1} \tag{5.7}$$
$$and\ \bar{M}(x_i, c_g) = 1-[1-(1+x_i^{c_g})^{-1}] = (1+x_i^{c_g})^{-1}.$$

Then

$$\frac{\partial}{\partial c_g} m(x_i, c_g) = [1 + c_g x_i^{c_g} ln(x_i)][x_i^{(c_g-1)}(1+x_i^{c_g})^{-2}],\ \frac{\partial}{\partial c_g} M(x_i, c_g) = x_i^{c_g}(1+x_i^{c_g})^{-2} ln(x_i),$$
$$\frac{\partial}{\partial c_g}\bar{M}(x_i, c_g) = -x_i^{c_g}(1+x_i^{c_g})^{-2} ln(x_i)\ and\ \frac{M(x_i, c_g)}{\bar{M}(x_i, c_g)} = \frac{1-(1+x_i^{c_g})^{-1}}{(1+x_i^{c_g})^{-1}} = x_i^{c_g}$$

$$\begin{aligned}
\frac{\partial \ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial \alpha_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Bigg(\frac{1}{\alpha_g} - b_g\left[\frac{M(x_i; c_g)}{\bar{M}(x_i; c_g)}\right]^{\beta_g} + (a_g-1)\left[1 - exp\left\{-\alpha_g\left[\frac{M(x_i; c_g)}{\bar{M}(x_i; c_g)}\right]^{\beta_g}\right\}\right]^{-1} \\
&\times \left[\frac{M(x_i; c_g)}{\bar{M}(x_i; c_g)}\right]^{\beta_g} exp\left\{-\alpha_g\left[\frac{M(x_i; c_g)}{\bar{M}(x_i; \upsilon_g)}\right]^{\beta_g}\right\}\Bigg) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Bigg(\frac{1}{\alpha_g} - b_g\left[x_i^{c_g}\right]^{\beta_g} + (a_g-1)\left[1 - exp\left\{-\alpha_g\left[x_i^{c_g}\right]^{\beta_g}\right\}\right]^{-1} \\
&\times \left[x_i^{c_g}\right]^{\beta_g} exp\left\{-\alpha_g\left[x_i^{c_g}\right]^{\beta_g}\right\}\Bigg) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Bigg(\frac{1}{\alpha_g} - b_g x_i^{c_g\beta_g} + (a_g-1)[1 - exp\left\{-\alpha_g x_i^{c_g\beta_g}\right\}]^{-1} x_i^{c_g\beta_g} exp\left\{-\alpha_g x_i^{c_g\beta_g}\right\}\Bigg).
\end{aligned} \tag{5.8}$$

$$
\begin{aligned}
\frac{\partial \ell\ell_c(\mathbf{\Theta},\mathbf{z}\mid x)}{\partial \beta_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\beta_g} + log\left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right] - \alpha_g b_g \left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right]^{\beta_g} log\left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right] \\
&+ (a_g-1)\alpha_g \left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right]^{\beta_g} log\left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right] \\
&\times exp\left\{-\alpha_g \left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right]^{\beta_g}\right\}\left[1-exp\left\{-\alpha_g \left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right]^{\beta_g}\right\}\right]^{-1}\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\beta_g} + log\left[x_i^{c_g}\right] - \alpha_g b_g \left[x_i^{c_g}\right]^{\beta_g} log\left[x_i^{c_g}\right] \\
&+ (a_g-1)\alpha_g \left[x_i^{c_g}\right]^{\beta_g} log\left[x_i^{c_g}\right] exp\left\{-\alpha_g\left[x_i^{c_g}\right]^{\beta_g}\right\}\left[1-exp\left\{-\alpha_g\left[x_i^{c_g}\right]^{\beta_g}\right\}\right]^{-1}\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\beta_g} + log(x_i^{c_g}) - \alpha_g b_g x_i^{c_g\beta_g} log(x_i^{c_g}) + \\
&(a_g-1)\alpha_g x_i^{c_g\beta_g} log(x_i^{c_g}) exp\left\{-\alpha_g(x_i^{c_g\beta_g})\right\}[1-exp\left\{-\alpha_g x_i^{c_g\beta_g}\right\}]^{-1}\Big). \qquad (5.9)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \ell\ell_c(\mathbf{\Theta},\mathbf{z}\mid x)}{\partial a_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(a_g) + log\left[1-exp\left\{-\alpha_g\left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right]^{\beta_g}\right\}\right]\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(a_g) + log\left[1-exp\left\{-\alpha_g\left[x_i^{c_g}\right]^{\beta_g}\right\}\right]\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(a_g) + log\left[1-exp\left\{-\alpha_g x_i^{c_g\beta_g}\right\}\right]\Big). \qquad (5.10)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \ell\ell_c(\mathbf{\Theta},\mathbf{z}\mid x)}{\partial b_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(b_g) - \alpha_g\left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right]^{\beta_g}\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(b_g) - \alpha_g\left[x^{c_g}\right]^{\beta_g}\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(b_g) - \alpha_g x_i^{c_g\beta_g}\Big). \qquad (5.11)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial c_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{m(x_i,c_g)}\frac{\partial}{\partial c_g}m(x_i,c_g) + \frac{(\beta_g-1)}{M(x_i,c_g)}\frac{\partial}{\partial c_g}M(x_i,c_g) - \frac{(\beta_g+1)}{\bar{M}(x_i,c_g)}\frac{\partial}{\partial c_g}\bar{M}(x_i,c_g) \\
&- \alpha_g b_g \beta_g \left(\frac{M(x_i,c_g}{\bar{M}(x_i,c_g}\right)^{\beta_g-1}\left(\frac{1}{\bar{M}(x_i,c_g)}\frac{\partial}{\partial c_g}M(x_i,c_g) - \frac{M(x_i,c_g)}{[\bar{M}(x_i,c_g)]^2}\frac{\partial}{\partial c_g}\bar{M}(x_i,c_g)\right) \\
&+ (a_g-1)\left[1 - exp\left\{-\alpha_g\left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right]^{\beta_g}\right\}\right]^{-1}\left[exp\left\{-\alpha_g\left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right]^{\beta_g}\right\}\right] \\
&\times \left[\alpha_g\beta_g\left(\frac{M(x_i,c_g}{\bar{M}(x_i,c_g}\right)^{\beta_g-1}\left(\frac{1}{\bar{M}(x_i,c_g)}\frac{\partial}{\partial c_g}M(x_i,c_g) - \frac{M(x_i,c_g)}{[\bar{M}(x_i,c_g)]^2}\frac{\partial}{\partial c_g}\bar{M}(x_i,c_g)\right)\right]\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{c_g x_i^{c_g-1}(1+x_i^{c_g})^{-2}}[1+c_g x_i^{c_g} ln(x_i)][x_i^{(c_g-1)}(1+x_i^{c_g})^{-2}] \\
&+ \frac{(\beta_g-1)}{1-(1+x_i^{c_g})^{-1}}x_i^{c_g}(1+x_i^{c_g})^{-2}ln(x_i) - \frac{(\beta_g+1)}{(1+x_i^{c_g})^{-1}}(-x_i^{c_g}(1+x_i^{c_g})^{-2}ln(x_i) \\
&- \alpha_g b_g\beta_g\left(x_i^{c_g}\right)^{\beta_g-1}\left(\frac{1}{(1+x_i^{c_g})^{-1}}x_i^{c_g}(1+x_i^{c_g})^{-2}ln(x_i) - \frac{1-(1+x_i^{c_g})^{-1}}{[(1+x_i^{c_g})^{-1}]^2}(-x_i^{c_g}(1+x_i^{c_g})^{-2}ln(x_i))\right) \\
&+ (a_g-1)\left[1-exp\left\{-\alpha_g\left[x_i^{c_g}\right]^{\beta_g}\right\}\right]^{-1}\left[exp\left\{-\alpha_g\left[x_i^{c_g}\right]^{\beta_g}\right\}\right] \\
&\times\left[\alpha_g\beta_g\left(x_i^{c_g}\right)^{\beta_g-1}\left(\frac{1}{(1+x_i^{c_g})^{-1}}x_i^{c_g}(1+x_i^{c_g})^{-2}ln(x_i) - \frac{1-(1+x_i^{c_g})^{-1}}{[(1+x_i^{c_g})^{-1}]^2}(-x_i^{c_g}(1+x_i^{c_g})^{-2}ln(x_i))\right)\right]\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\left[(b_g\alpha_g\beta_g x_i^{\beta_g c_g}log(x_i) - \beta_g log(x_i))c_g - 1\right]exp\left\{\alpha_g x_i^{c_g\beta_g}\right\} \\
&+ ((-b_g-a_g+1)\alpha_g\beta_g x_i^{c_g\beta_g}log(x_i) + \beta_g log(x_i))c_g + 1)(c_g\left[1-exp\left\{\alpha_g x_i^{c_g\beta_g}\right\}\right])^{-1}\Big).
\end{aligned} \tag{5.12}
$$

$$
\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \pi_g} = \sum_{i=1}^{n}\sum_{g=1}^{G}\frac{z_{ig}}{\pi_g}. \tag{5.13}
$$

$$
\begin{aligned}
\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial z_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} log\Big(\frac{\alpha_g\beta_g}{B(a_g,b_g)} m(x_i;c_g)\frac{M(x_i;c_g)^{\beta_g-1}}{\bar{M}(x_i;c_g)^{\beta_g+1}} exp\left\{-\alpha_g b_g\left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right]^{\beta_g}\right\} \\
&\times \left[1-exp\left\{-\alpha_g\left[\frac{M(x_i;c_g)}{\bar{M}(x_i;c_g)}\right]^{\beta_g}\right\}\right]^{a_g-1}\Big) + \sum_{i=1}^{n}\sum_{g=1}^{G} log(\pi_g) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} log\Big(\frac{\alpha_g\beta_g}{B(a_g,b_g)} c_g x_i^{c_g-1}(1+x_i^{c_g})^{-2}\frac{(1-(1+x_i^{c_g})^{-1})^{\beta_g-1}}{((1+x_i^{c_g})^{-1})^{\beta_g+1}} exp\left\{-\alpha_g b_g\left[x_i^{c_g}\right]^{\beta_g}\right\} \\
&\times \left[1-exp\left\{-\alpha_g\left[\frac{1-(1+x_i^{c_g})^{-1}}{(1+x_i^{c_g})^{-1}}\right]^{\beta_g}\right\}\right]^{a_g-1}\Big) + \sum_{i=1}^{n}\sum_{g=1}^{G} log(\pi_g) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G}\left(log\left(\frac{\alpha_g\beta_g}{B(a_g,b_g)}\right) + log(c_g x_i^{c_g-1}(1+x_i^{c_g})^{-2}) + log\left(\frac{(1-(1+x_i^{c_g})^{-1})^{\beta_g-1}}{((1+x_i^{c_g})^{-1})^{\beta_g+1}}\right)\right. \\
&\left. -\alpha_g b_g x_i^{c_g\beta_g} + (a_g-1)log\left[1-exp\left\{-\alpha_g x_i^{c_g\beta_g}\right\}\right]\right) + \sum_{i=1}^{n}\sum_{g=1}^{G} log(\pi_g).
\end{aligned}
\tag{5.14}
$$

5.3 The partial derivatives of the complete log likelihood of the BWE mixture

$m(x_i;\omega_g) = \omega_g e^{-\omega_g x_i}, M(x_i;\omega_g) = 1-e^{-\omega_g x_i}$ *and* $\bar{M}(x_i;\omega_g) = 1-(1-e^{-\omega_g x_i}) = e^{-\omega_g x_i}$, *thus,*

$\frac{\partial}{\partial\omega_g} m(x_i;\omega_g) = e^{-\omega_g x_i}(1-x_i\omega_g)$, $\frac{\partial}{\partial\omega_g} M(x_i;\omega_g) = x_i e^{-\omega_g x_i}$, $\frac{\partial}{\partial\omega_g}\bar{M}(x_i;\omega_g) = -x_i e^{-\omega_g x_i}$ *and*
$\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)} = \frac{1-e^{-\omega_g x_i}}{e^{-\omega_g x_i}} = e^{\omega_g x_i}-1$

$$\begin{aligned}\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \alpha_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\alpha_g} - b_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g} + (a_g-1)\left[1 - exp\left\{-\alpha_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g}\right\}\right]^{-1} \\ &\times \left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g} exp\left\{-\alpha_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g}\right\}\Big) \\ &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\alpha_g} - b_g\left[e^{\omega_g x_i}-1\right]^{\beta_g} + (a_g-1)\left[1 - exp\left\{-\alpha_g\left[e^{\omega_g x_i}-1\right]^{\beta_g}\right\}\right]^{-1} \\ &\times \left[e^{\omega_g x_i}-1\right]^{\beta_g} exp\left\{-\alpha_g\left[e^{\omega_g x_i}-1\right]^{\beta_g}\right\}\Big). \end{aligned} \tag{5.15}$$

$$\begin{aligned}\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial \beta_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\beta_g} + log\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right] - \alpha_g b_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g} log\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right] \\ &+ (a_g-1)\alpha_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g} log\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right] \\ &\times exp\left\{-\alpha_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g}\right\}\left[1 - exp\left\{-\alpha_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g}\right\}\right]^{-1}\Big) \\ &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\beta_g} + log\left[e^{\omega_g x_i}-1\right] - \alpha_g b_g\left[e^{\omega_g x_i}-1\right]^{\beta_g} log\left[e^{\omega_g x_i}-1\right] \\ &+ (a_g-1)\alpha_g\left[e^{\omega_g x_i}-1\right]^{\beta_g} log\left[e^{\omega_g x_i}-1\right] exp\left\{-\alpha_g\left[e^{\omega_g x_i}-1\right]^{\beta_g}\right\}\left[1 - exp\left\{-\alpha_g\left[e^{\omega_g x_i}-1\right]^{\beta_g}\right\}\right]^{-1}\Big). \end{aligned} \tag{5.16}$$

$$\begin{aligned}\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial a_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(a_g) + log\left[1 - exp\left\{-\alpha_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g}\right\}\right]\Big) \\ &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(a_g) + log\left[1 - exp\left\{-\alpha_g\left[e^{\omega_g x_i}-1\right]^{\beta_g}\right\}\right]\Big). \end{aligned} \tag{5.17}$$

$$\begin{aligned}\frac{\partial \ell\ell_c(\boldsymbol{\Theta}, \mathbf{z} \mid x)}{\partial b_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(b_g) - \alpha_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g}\Big) \\ &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\varphi_0(a_g+b_g) - \varphi_0(b_g) - \alpha_g\left[e^{\omega_g x_i}-1\right]^{\beta_g}\Big). \end{aligned} \tag{5.18}$$

$$
\begin{aligned}
\frac{\partial \ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial \omega_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{m(x_i;\omega_g)}\frac{\partial}{\partial\omega_g}m(x_i;\omega_g) + \frac{(\beta_g-1)}{M(x_i;\omega_g)}\frac{\partial}{\partial\omega_g}M(x_i;\omega_g) - \frac{(\beta_g+1)}{\bar{M}(x_i;\omega_g)}\frac{\partial}{\partial\omega_g}\bar{M}(x_i;\omega_g) \\
&- \alpha_g b_g \beta_g \left(\frac{M(x_i;\omega_g}{\bar{M}(x_i;\omega_g}\right)^{\beta_g-1}\left(\frac{1}{\bar{M}(x_i;\omega_g)}\frac{\partial}{\partial\omega_g}M(x_i;\omega_g) - \frac{M(x_i;\omega_g)}{[\bar{M}(x_i;\omega_g)]^2}\frac{\partial}{\partial\omega_g}\bar{M}(x_i;\omega_g)\right) \\
&+ (a_g-1)\left[1 - exp\left\{-\alpha_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g}\right\}\right]^{-1}\left[exp\left\{-\alpha_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g}\right\}\right] \\
&\times \left[\alpha_g\beta_g\left(\frac{M(x_i;\omega_g}{\bar{M}(x_i;\omega_g}\right)^{\beta_g-1}\left(\frac{1}{\bar{M}(x_i;\omega_g)}\frac{\partial}{\partial\omega_g}M(x_i;\omega_g) - \frac{M(x_i;\omega_g)}{[\bar{M}(x_i;\omega_g)]^2}\frac{\partial}{\partial\omega_g}\bar{M}(x_i;\omega_g)\right)\right]\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1}{\omega_g e^{-\omega_g x_i}}e^{-\omega_g x}(1-x_i\omega_g) + \frac{(\beta_g-1)}{1-e^{-\omega_g x_i}}x_i e^{-\omega_g x_i} - \frac{(\beta_g+1)}{e^{-\omega_g x_i}}(-x_i e^{-\omega_g x_i}) \\
&- \alpha_g b_g \beta_g \left(e^{\omega_g x_i}-1\right)^{\beta_g-1}\left(\frac{1}{e^{-\omega_g x_i}}x_i e^{-\omega_g x_i} - \frac{1-e^{-\omega_g x_i}}{[e^{-\omega_g x_i}]^2}(-x_i e^{-\omega_g x_i})\right) \\
&+ (a_g-1)\left[1 - exp\left\{-\alpha_g\left[e^{\omega_g x_i}-1\right]^{\beta_g}\right\}\right]^{-1}\left[exp\left\{-\alpha_g\left[e^{\omega_g x_i}-1\right]^{\beta_g}\right\}\right] \\
&\times \left[\alpha_g\beta_g\left(e^{\omega_g x_i}-1\right)^{\beta_g-1}\left(\frac{1}{e^{-\omega_g x_i}}x_i e^{-\omega_g x_i} - \frac{1-e^{-\omega_g x_i}}{[e^{-\omega_g x_i}]^2}(-x_i e^{-\omega_g x_i})\right)\right]\Big) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\frac{1-x_i\omega_g}{\omega_g} + \frac{(\beta_g-1)x_i}{(e^{\omega_g x_i}-1)} + (\beta_g+1)x_i - \alpha_g b_g\beta_g(e^{\omega_g x_i}-1)^{\beta_g-1}(x_i + e^{\omega_g x_i}-1) \\
&+ (a_g-1)[1 - exp\left\{-\alpha_g(e^{\omega_g x_i}-1)^{\beta_g}\right\}]^{-1}[exp\left\{-\alpha_g(e^{\omega_g x_i}-1)^{\beta_g}\right\}] \\
&\times \alpha_g\beta_g(e^{\omega_g x_i}-1)^{\beta_g-1}[x_i + x_i(e^{\omega_g x_i}-1)]\Big). \qquad \textbf{(5.19)}
\end{aligned}
$$

$$
\frac{\partial \ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial \pi_g} = \sum_{i=1}^{n}\sum_{g=1}^{G}\frac{1}{\pi_g} + \sum_{i=1}^{n}\sum_{g=1}^{G}\frac{z_g}{\pi_g}. \qquad \textbf{(5.20)}
$$

$$
\begin{aligned}
\frac{\partial \ell\ell_c(\mathbf{\Theta}, \mathbf{z} \mid x)}{\partial z_g} &= \sum_{i=1}^{n}\sum_{g=1}^{G} log\Big(\frac{\alpha_g\beta_g}{B(a_g,b_g)}m(x_i;\omega_g)\frac{M(x_i;\omega_g)^{\beta_g-1}}{\bar{M}(x_i;\omega_g)^{\beta_g+1}}exp\left\{-\alpha_g b_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g}\right\} \\
&\times \left[1 - exp\left\{-\alpha_g\left[\frac{M(x_i;\omega_g)}{\bar{M}(x_i;\omega_g)}\right]^{\beta_g}\right\}\right]^{a_g-1}\Big) + \sum_{i=1}^{n}\sum_{g=1}^{G} log(\pi_g) \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} log\Big(\frac{\alpha_g\beta_g}{B(a_g,b_g)}\omega_g e^{-\omega_g x_i}\frac{(1-e^{-\omega_g x_i})^{\beta_g-1}}{(e^{-\omega_g x_i})^{\beta_g+1}}exp\left\{-\alpha_g b_g\left[e^{\omega_g x_i}-1\right]^{\beta_g}\right\} \\
&\times \left[1 - exp\left\{-\alpha_g\left[e^{\omega_g x_i}-1\right]^{\beta_g}\right\}\right]^{a_g-1}\Big) + \sum_{i=1}^{n}\sum_{g=1}^{G} log(\pi_g). \qquad \textbf{(5.21)}
\end{aligned}
$$

5.4 Identifiability of mixture models

Chandra (1977) proposed the following theorems to prove weak condition for identifiability of mixture models.

5.4.1 Theorem

Let there be a transform Φ_g associated with each $F_g \in \Phi$ having the domain of definition D_{Φ_g}, and suppose that the mapping $M : F_g \to \Phi_g$ is linear. Suppose also that there exists a total ordering $(\leq)$ in Φ such that

i. $F_1 \leq F_2, (F1, F2 \in \Phi)$ implies $D_{\Phi_1} \subseteq D_{\Phi_2}$

ii. For each $F_1 \in \Phi$, there exists some t_1 in the closure of $T_1 = t : \Phi_1(t) \neq 0$ such that

$$\lim_{\substack{t \to t_1 \\ t \in T_1}} \frac{\Phi_2(t)}{\Phi_1(t)} = 0$$

for each $F_1 < F_2, (F1, F2 \in \Phi)$. Then the class Λ of all finite mixing distributions is identifiable relative to Φ.

5.4.2 Theorem

Let $\boldsymbol{F}$ be a family of distributions. Let M be a linear mapping which transforms any $F \in \boldsymbol{F}$ into a real function Φ_F with domain $D(F) \subset \mathbb{R}^d$. Let $D_0(F) = \{s \in D(F) : \Phi_F(s) \neq 0\}$. Suppose that there exists a total order $\prec$ on $\boldsymbol{F}$, such that for any $F \in \boldsymbol{F}$ there exists $s(F) \in D_0(F)'$ verifying:

a) If $F_1, F_2, ..., F_m \in \boldsymbol{F}$ with $F_1 \prec F_g$ for $2 \leq g \leq m$, then

$$s(F_1) \in [D_0(F_1) \cap (\bigcap_{g=2}^{m} D(F_g))]'.$$

b) If $F_1 \prec F_2$, then $\lim_{s \to s(F_1)} \frac{\Phi_{F_2}(s)}{\Phi_{F_1}(s)} = 0$. Then the class $\boldsymbol{M}$ of all finite mixture distribution of $\boldsymbol{F}$ is identifiable.