BIVARIATE MIXTURE CURE RATE MODEL WITH MORAN-DOWNTON WEIBULL DISTRIBUTION AND ASSOCIATED EM ALGORITHM IMPLEMENTATION

BIVARIATE MIXTURE CURE RATE MODEL WITH MORAN-DOWNTON WEIBULL DISTRIBUTION AND ASSOCIATED EM ALGORITHM IMPLEMENTATION

By MATILDA PITT

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2023)

Hamilton, Ontario, Canada (Mathematics and Statistics)

TITLE:	Bivariate Mixture Cure Rate Model with Moran-Downton Weibull Distribution and Associated EM Algorithm Implementation
AUTHOR:	Matilda Pitt
SUPERVISOR:	Prof. N. Balakrishnan, Prof. Katherine Davies
NUMBER OF PAGES:	viii, 87

Abstract

This thesis introduces a new bivariate cure rate model and develops an Expectation-Maximization (EM) algorithm in R to fit the model. Within survival analysis, cure rate models describe scenarios wherein part of the population is cured and therefore would never experience the event of interest. Under this set-up, bivariate cure rate models are needed when there is a pair of events of interest. Here, a Moran-Downton bivariate Weibull distribution is used to model the paired event times of the susceptible individuals. An EM algorithm is developed here and implemented in R for this parametric bivariate cure rate model. Simulation studies are then performed to evaluate the performance of the developed model-fitting methods and finally the algorithm is applied to a real life dataset on diabetic retinopathy.

Acknowledgements

I would like thank my supervisors, Professor N. Balakrishnan and Professor Katherine Davies for their encouragement, kindness and wisdom they shared with me. I am truly grateful for all the help and guidance they have provided me with and I could not have completed this thesis without them.

I would like to thank all of the friends I have made since arriving in Canada for helping it feel like a home.

Last, but not least, I would like to thank my parents for their constant support and encouragement. I could not succeed without them.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Survival analysis is the area of statistics concerned with the time until an event happens, where traditionally it is assumed that the event will happen for certain to every subject in the population. However, in some situations this assumption may not hold and the event will not happen to every subject in the population. For example, not everyone who is diagnosed with cancer will die of cancer and not all prisoners will re-offend, and for this reason a model is needed that allows for part of the population to not experience the event of interest. This is referred to as a cure rate model and is usually modelled via a mixture model made up of subjects that are susceptible to the event and subjects that are cured of the event. Data that fits this scenario can be seen to have a survival function that plateaus substantially above 0 on a Kaplan-Meier curve. More generally, there may be a pair of events of interest for each subject having some dependency. For example, modelling blindness in individuals where the two events of interest may correspond to going blind in each eye. In this case, a bivariate cure rate model would be useful enabling the dependency between the paired events to be modelled. Here an extension of the univariate mixture

model with a Moran-Downton bivariate Weibull distribution as the joint distribution for the subjects who are susceptible to both events, is developed. Due to the lack of knowledge on which part of the population is cured and which part is susceptible, the fitting of the considered model becomes a challenging task. An Expectation-Maximization (EM) algorithm is developed here for the first time to the bivariate cure rate model with the unobserved latent variable being taken as the cure status. No previous attempts in published literature of an EM algorithm implementation could be found. We have implemented the developed EM algorithm in R and have applied it to simulated and real life datasets.

In Chapter 2, some introductory notions and results are provided. Chapter 3 details the specifics of the Moran-Downton bivariate Weibull cure rate model. Chapter 4 develops the EM algorithm for fitting the model, while Chapter 5 presents the results of an empirical study. In Chapter 6, the algorithm is applied to a real life dataset on diabetic retinopathy. Finally, Chapter 7 provides some concluding remarks and suggestions for future work.

Chapter 2

Preliminaries

2.1 Survival Analysis

Survival analysis is the area of statistics in which we are interested in the time until an event takes place [41]. It is also referred to as reliability theory in engineering literature, where it has been utilized to study aeroplane engine failure [38], and as duration modelling in economic literature [57]. The time to event is often referred to as lifetime or failure time, although the event doesn't have to correspond to failure or death of a unit/individual. Events of interest can include death, a mechanical failure, or even a positive such as time until hospital discharge following a treatment. Traditionally, only one event will occur at a specific time and once that event has occurred, no more data is collected on that subject. Survival analysis models these times to event through various methods with the aim of interpreting the models, comparing models between outcomes as well as for inferring relationships between predictor variables and the outcomes.

2.1.1 Univariate Lifetime Distributions

In survival analysis, there are some distributions that are particularly useful. Also, there are many different ways of representing these distributions based on different characteristics of the distributions [44].

A standard method of describing a distribution is through its probability density function (pdf). The probability density function, $f(x)$, is such that:

$$\int_a^b f(x)dx = P(a < X < b), \tag{2.1.1}$$

where X is the random variable representing the lifetime. In survival analysis, the support of the lifetime will be $[0, \infty)$. This means that $\int_0^\infty f(x)dx = 1$ and $f(x) \geq 0$ for all $x \geq 0$.

Another key function of interest is the survival function, defined as the probability that a subject has not experienced the event of interest before a given time. It is given by

$$S(t) = P(X > t) = \int_t^\infty f(x)dx. \tag{2.1.2}$$

The following relationship is known between the survival function and the cumulative distribution function (cdf):

$$S(t) = 1 - P(X \leq t) = 1 - F(t), \tag{2.1.3}$$

where $F(t)$ denotes the cdf. Another characteristic of importance is the hazard rate. The hazard rate is the instantaneous failure rate at time t, and is defined as:

$$h(t) = \lim_{\Delta t \to 0} \left[\frac{P(t < X < t + \Delta t | X > t)}{\Delta t} \right]. \tag{2.1.4}$$

If there are covariates present in the data, the hazard rate can be modified to incorporate the covariates to get the Cox proportional hazards model [15], with the hazard rate function now being of the form:

$$h(t) = h_0(t) \exp(\boldsymbol{\beta}\mathbf{z}), \ t \in [0, \infty), \tag{2.1.5}$$

where $h_0(t)$ is the baseline hazard, $\mathbf{z}$ is the vector of covariates, and $\boldsymbol{\beta}$ is a vector of coefficients.

If we assume a constant hazard rate irrespective of time, t, then this leads to the exponential distribution. The exponential distribution has hazard rate:

$$h(t) = \lambda, \tag{2.1.6}$$

with λ being the rate parameter. The corresponding pdf is

$$f(x) = \lambda \exp(\lambda x), \ x \in [0, \infty), \tag{2.1.7}$$

and the survival function is

$$S(x) = \exp(-\lambda x), \ x \in [0, \infty). \tag{2.1.8}$$

Due to the constant hazard rate, the exponential distribution has the memoryless property which means that, when waiting for an event to occur after some initial time, the distribution of the remaining time until the event is the same as the original distribution; that is $P(X > x + y|X > x) = P(X > y), y > 0$ [4]. This means the exponential distribution will be a useful model when the risk does not change over time, such as modelling death within an extremely ill population [53].

If we take Y as the sum of n independent lifetimes from an exponential distribution, then Y is distributed as gamma.

The gamma distribution with shape parameter α and rate parameter β has its pdf as

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x),\ x \in [0, \infty), \tag{2.1.9}$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x)dx$ is the Gamma function [4]. It has its hazard rate as

$$h(t) = \frac{\beta^\alpha t^{\alpha-1} \exp(-\beta t)}{\Gamma(\alpha) - \Gamma(\alpha, \beta t)},\ t \in [0, \infty) \tag{2.1.10}$$

where $\Gamma(\alpha, \beta t) = \int_0^{\beta t} x^{\alpha-1} \exp(-x)dx$ is the incomplete Gamma function [4]. The survival function is

$$S(x) = 1 - \frac{\Gamma(\alpha, x\beta)}{\Gamma(\alpha)},\ x \in [0, \infty). \tag{2.1.11}$$

The gamma distribution is a very popular model in survival analysis as the shape parameter α allows for a variety of distributional shapes. In particular, it has been used in modelling cancer occurence after several carcinogenic events [6].

Finally, another useful lifetime model is the Weibull distribution, which is the one we focus on in this thesis. The Weibull distribution was introduced in 1939 by Weibull

and has the following pdf [60]:

$$f(x) = \frac{\beta}{\alpha} \exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right)\left(\frac{x}{\alpha}\right)^{\beta-1},\ x \in [0,\infty), \tag{2.1.12}$$

where α is the scale parameter and β is the shape parameter, with $\alpha, \beta \in (0,\infty)$. The hazard rate is given by

$$h(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1},\ t \in [0,\infty)\ [33], \tag{2.1.13}$$

and the survival function is given by

$$S(x) = \exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right),\ x \in [0,\infty). \tag{2.1.14}$$

The Weibull distribution is popular within the medical community and also extensively used in reliability theory [43]. Furthermore, the Weibull distribution is an extension of the exponential distribution, and in fact exponential variables can be readily transformed into Weibull variables through a power transformation: for Weibull random variable W and standard exponential random variable, X, $W = \alpha X^{\frac{1}{\beta}}$ with α the scale and β the shape parameter.

2.1.2 Bivariate Lifetime Modelling

As mentioned earlier, there are situations where there are two lifetimes which are dependent on each other. In such a case, a bivariate distribution is needed to model their lifetimes. Fréchet showed that for any two given marginal distributions, there are infinitely many bivariate distributions [32]. There are a variety of methods for

constructing such distributions with different interpretations and/or motivations.

One method we use is through frailty models [13]. Frailty models are an extension of Cox proportional hazard models and they assume there are two causes that lead to variability in lifetimes, one due to observed covariates and the other due to an unobserved random effect [36]. The hazard function is therefore made up of the baseline hazard, a covariate function if applicable and the frailty variable which is unobserved and has a different value for each subject in the population. In the bivariate case, in shared frailty models, the value of the frailty variable is shared by the two event times [59]. In correlated frailty models, the paired events have separate frailty random variables, but they have a joint distribution which allows for dependency [61]. The frailty random variable comes from a given distribution and the gamma distribution in particular is often used in the shared frailty model and various bivariate gamma distributions are used in the correlated frailty model [59, 62].

Copulas offer another method. Copulas are multivariate cumulative distribution functions with marginal uniform functions on [0,1] [23]. In the case of two lifetimes, one has to consider bivariate copulas. If $(x_1, x_2, ..., x_n)$ is a sample of observations from a distribution with pdf $f(x)$ and cdf $F(x)$, then by probability integral transform, $(F(x_1), F(x_2), ..., F(x_n))$ is uniformly distributed on [0,1] [23]. This means we can apply a copula to any marginal distributions. Furthermore, Sklar's Theorem states that any multivariate distribution can be written in terms of a copula and univariate marginal distributions [54]. There is a wide array of copulas which allow for a variety of dependence structures while modelling two or more lifetimes [4].

Finally, there are other bivariate distributions that do not arise from either copulas or frailty models. However, from Sklar's Theorem, we know there must be a copula that

describes them. Other methods include formulating from specific scenarios or using transformation methods to change the marginals of an existing bivariate distribution [4]. In particular, we are interested in the Moran-Downton bivariate distribution in this thesis.

The Moran-Downton bivariate exponential distribution was first developed by Moran [46] and applied in reliability analysis by Downton [24]. The distribution is constructed by assuming that "shocks" occur to each individual in a pair at independent random exponential intervals and after a set number of shocks, that individual fails and the time of failure is recorded as the lifetime. The number of shocks that is required for failure has a bivariate geometric distribution with joint probability generating function [4]:

$$P(z_1, z_2) = \frac{z_1 z_2}{1 + \alpha + \beta + \gamma - \alpha z_1 - \beta z_2 - \gamma z_1 z_2}. \tag{2.1.15}$$

As the number of shocks until failure is correlated within a pair, the lifetimes will be correlated as well which is supposed to explain the dependence between the individuals in a pair. The marginal distributions of both individuals lifetimes are exponential. The Moran-Downton bivariate exponential distribution thus devised has the following joint pdf [47]:

$$f(x, y) = \frac{1}{\alpha_1 \alpha_2 (1 - \rho)} \exp\left(-\frac{x}{\alpha_1 (1 - \rho)} - \frac{y}{\alpha_2 (1 - \rho)} \right) I_0 \left(\frac{2(xy\rho)^{1/2}}{(1 - \rho)(\alpha_1 \alpha_2)^{1/2}} \right),$$
$$x, y \in [0, \infty), \tag{2.1.16}$$

where

$$I_0(z) = \sum_{r=0}^{\infty} \frac{z^{2r}}{2^{2r}(r!)^2} \tag{2.1.17}$$

is the modified Bessel function. The parameters $\alpha_1, \alpha_2 \in (0, \infty)$ are the scale parameters from the two marginal exponential distributions. The correlation parameter, ρ, has support $[0, 1)$, meaning only a positive correlation between the outcomes can be modelled. We would generally expect a positive correlation in the cure context we are considering so this restriction seems quite reasonable [24]. The standard Moran-Downton bivariate exponential distribution, when $\alpha_1 = \alpha_2 = 1$, has the following pdf:

$$f(x, y) = \frac{1}{(1-\rho)} \exp\left(-\frac{x}{(1-\rho)} - \frac{y}{(1-\rho)}\right) I_0\left(\frac{2(xy\rho)^{1/2}}{(1-\rho)}\right), \; x, y \in [0, \infty), \tag{2.1.18}$$

and the corresponding joint cdf is given by [4]

$$F(x, y) = (1 - \exp(-x))(1 - \exp(-y)) + \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_j^{(1)}(x) L_j^{(1)}(y) xy \exp(-(x+y)),$$
$$x, y \in [0, \infty), \tag{2.1.19}$$

where $L_j^{\alpha}(x)$ are Laguerre polynomials and are defined as

$$L_j^{\alpha}(x) = \sum_{k=0}^{j} \binom{j+\alpha}{j-k} \frac{(-x)^k}{k!}. \tag{2.1.20}$$

Furthermore, they are orthogonal polynomials, meaning $\int L_j^\alpha(x)L_k^\alpha(x)dx = 0$ for $j \neq k$ with the following property [56]:

$$\frac{dL_n^\alpha(x)}{dx} = -L_{n-1}^{\alpha+1}(x). \tag{2.1.21}$$

2.1.3 Censoring

What makes survival analysis more complicated is that normally some kind of censoring occurs in the lifetimes. This means two values are returned for each subject. There is the observed time that will either be the lifetime or the censoring time and a censoring indicator that indicates whether the observed time is in fact a lifetime or a censored time. There are three main types of censoring that occur in practice.

Right Censoring

This is when we have a censoring time that is smaller than the lifetime, meaning the actual lifetime will never be observed. This could be due to an individual leaving the study early from unrelated causes, individuals being lost to follow-up after moving away and loss of contact, the study ending before the event has occurred, or a preliminary analysis being undertaken whilst the study is still underway [41]. The censoring time can be pre-fixed before the trial starts, or could be random. If we denote x_i for the lifetime, c_i for the censoring time and t_i for the observed lifetime for observation i then in the case of right censoring, we have $t_i = min(c_i, x_i)$ with $t_i \leq x_i$. If the probability of being censored does not affect the failure rate, then the censoring mechanism is said to be independent [39]. This is the most common type of censoring seen in survival analysis. There are two particular mechanisms that can

lead to right censoring.

Type-I Censoring: This is when all event times are recorded up to a specified time and after that, they get censored. This is often due to a trial ending at a specific timepoint [42];

Type-II Censoring: This is when the first r out of n event times are recorded and the remaining subjects get censored. This is often due to time and financial constraints, and so the trial could end early. The uncensored values are the first r order statistics out of n [2].

Left Censoring

This is when the lifetime must have occurred before a censoring time. In the case when the time of becoming infected is the event of interest, left censoring can occur when a subject tested positive at time t_i for being infected. This means the subject must have gotten infected before they tested positive with the exact time of being infected unknown. Alternatively, the event could have occurred even before the trial started [41].

Interval Censoring

This is when the event occurs within a specific time interval, but the true lifetime is unknown. In the case when becoming infected is the event of interest with infection status known by means of a test, this can occur when an individual is tested at regular intervals. This means the true lifetime is between the last negative test and the first positive test [41].

2.2 Likelihood-Based Inference

Standard parametric methods for fitting models to datasets assume the data observed is from a specific distribution and the parameter values that maximize the likelihood of observing that dataset are calculated.

2.2.1 Maximum Likelihood Inference

In the simplest case, we assume that all observations are independent and identically distributed (i.i.d.) from a distribution with parameters $\boldsymbol{\theta}$ and pdf $f(x|\boldsymbol{\theta})$. The likelihood function, denoted by $\mathcal{L}(\boldsymbol{\theta}|\mathbf{X})$, is the likelihood of observing that dataset, $\mathbf{X}$. If the data is complete, it is simply the joint pdf, and in this case it is given by

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{X}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}). \tag{2.2.1}$$

Suppose we have observations that have been Type I censored on the right. Then, we cannot use the pdf for the censored values, but must use the survival function instead. This is because we do not know the exact time the event will occur but all we know is that it occurs after the censoring time. The likelihood function becomes:

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{X}) = \prod_{i=1}^{n-m} f(x_i|\boldsymbol{\theta}) \prod_{j=1}^{m} S(y_j|\boldsymbol{\theta}), \tag{2.2.2}$$

where the x_i's are the $n-m$ uncensored values and the y_j's are the m censored values. We can maximise (2.2.2) to find the estimates of the parameters. Due to the product form and for ease of differentiation, it is common to take log and then differentiate with respect to each parameter and solve the system of equations. The

solutions found are referred to as the maximum likelihood estimates (MLE). In the case when the observations are not i.i.d., such as when Type-II censoring occurs and the observations are order statistics, the likelihood equation is still the joint pdf, but it takes on a more complicated form. The technique still remains the same.

2.2.2 EM Algorithm

The Expectation-Maximization (EM) algorithm was introduced by Dempster et al (1977) and is used to fit a model when there is an unobserved latent variable in the dataset [22]. It still uses the principles of maximum likelihood, but is more adaptable and can allow for greater inference about the unobserved variable. Some earlier works have used the EM algorithm for mixture models [11] and also for bivariate interval censored models [40].

The generic EM algorithm for a dataset, with unobserved latent variable z, can be summarized as follows:

1. Choose starting values for the parameters;
2. E-step: Calculate the conditional distribution of z, given the dataset, and using current parameter estimates. Then calculate the expected complete (log) likelihood function, termed as the Q-function;
3. M-step: Maximize the Q-function with respect to the parameters and update the parameter estimates;
4. Repeat Steps 2 and 3 until convergence is achieved.

There is no specific convergence criteria, but most implementations require either

subsequent parameter estimates to be sufficiently close or subsequent observed log-likelihoods with parameter estimates to be sufficiently close [35]. There are noted limitations in the EM algorithm. First, it is very sensitive to its starting values for the parameters and can converge to a local, but not global, maximum [8]. It can also be slow to converge [58]. Furthermore, it does not directly provide standard errors for the estimates obtained, but there have been some suggestions as to how to calculate the standard errors for the parameter estimates. Methods of calculating standard errors are related to observed information matrix, expected information matrix, or resampling methods [3]. One such resampling method is the well-known bootstrap [45].

2.2.3 Standard Error and Confidence Interval Calculations: Bootstrap

The bootstrap method to calculate the standard error of an estimate was first introduced by Efron (1979) [26]. In 1981, it was applied to right censored data [27] and in 1986 to bivariate data [29], but with the key principles of the method remaining the same. The method works for any statistic calculated from the dataset, $\mathbf{X}$. Let us denoted the statistic of interest by $d(\mathbf{X})$. In the bootstrap method there are two main types, namely, non-parametric and parametric.

Non-parametric Bootstrap

Non-parametric bootstrap was introduced first and it uses the following steps:

1. Sample with replacement observations from the original dataset to create a new dataset, a bootstrap sample $\mathbf{X}_i$;

2. Using $\mathbf{X}_i$, calculate the statistic, $d(\mathbf{X}_i)$, called the bootstrap replication;

3. Repeat Steps 1 and 2 to generate B bootstrap replications $d(\mathbf{X}_1), d(\mathbf{X}_2), ..., d(\mathbf{X}_B)$;

4. Calculate the standard deviation of the bootstrap replications.

The idea is that the variance of $d(\mathbf{X}_1), d(\mathbf{X}_2), ..., d(\mathbf{X}_B)$ is approximately the variance of the statistic $d(\mathbf{X})$. This means calculating the standard deviation of the bootstrap replications is an estimate for the standard error of the parameter estimate. In general, suitable values for B are from 25 to 200 [30].

Parametric Bootstrap

Parametric bootstrap uses new datasets that are generated by sampling from the assumed population distribution, instead of sampling from the original dataset [30]. Replications of the statistic are found for each dataset and then the standard deviation of the replications are found as in the non-parametric case. This leads to results that closely match analytical estimates of standard errors, but requires assumptions about the population distribution [30].

Percentile Bootstrap Confidence Intervals

The bootstrap method can also be used to find confidence intervals. One method is to use the empirical percentiles of the bootstrap replications for confidence intervals [17]. For confidence intervals, more than 200 replications are required for an accurate estimate, normally, at least 1000 [29]. For a percentile $100(1-\alpha)\%$ confidence interval using the parametric bootstrap, the steps are as follows:

1. Sample from the assumed distribution to create a new dataset, a bootstrap sample $\mathbf{X}_i$;

2. Calculate the statistic, $d(\mathbf{X}_i)$, using the bootstrap sample;

3. Repeat Steps 1 and 2 to generate B bootstrap replications $d(\mathbf{X}_1), d(\mathbf{X}_2), ..., d(\mathbf{X}_B)$;

4. Order the values of $d(\mathbf{X}_i)$ and select the $B \times \frac{\alpha}{2}$th and $B \times (1 - \frac{\alpha}{2})$th values.

This means for a 90% confidence interval, with $B = 1000$, the 50th and the 950th ordered replications would be used. The percentile interval is invariant under reparamatrization, but may not always perform well [17].

$\mathbf{BC_a}$ Bootstrap Confidence Intervals

The percentile method to find confidence intervals does not adapt to bias and skewness in the bootstrap replications meaning it can be inaccurate. An improved version of the percentile method that does adjusts for skewness and bias is the bias-corrected and accelerated bootstrap, referred to as the $\mathrm{BC_a}$ [28]. Here the interval is still defined by percentiles but not necessarily the same percentiles as the percentile method. The lower bound is the:

$$\Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})}\right) \tag{2.2.3}$$

percentile and the upper bound is the:

$$\Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})}\right) \tag{2.2.4}$$

percentile. Here, $\hat{z}_0$ is referred to as the bias-correction and is calculated by:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#d(\mathbf{X}_i) < d(\mathbf{X})}{B}\right), \tag{2.2.5}$$

with $d(\mathbf{X})$ the parameter estimate from the original dataset and $d(\mathbf{X}_i)$ the bootstrap replications. The acceleration parameter, $\hat{a}$ can be calculated by [30]:

$$\hat{a} = \frac{\sum_{i=1}^{n}(\bar{d}(\mathbf{X}) - d(\mathbf{X}_i^*))^3}{6(\sum_{i=1}^{n}(\bar{d}(\mathbf{X}) - d(\mathbf{X}_i^*))^2)^{3/2}}, \tag{2.2.6}$$

where $d(\mathbf{X}_i^*)$ are the jackknife values of the statistic of interest and $\bar{d}(\mathbf{X})$ is the mean of the jackknife estimates. In the case where $\hat{z}_0 = \hat{a} = 0$, then we have the usual percentile confidence intervals.

2.3 Cure Rate Models

2.3.1 Univariate Cure Rate Model

A cure rate model was first introduced by Boag (1949) [9] and is used extensively in survival analysis when part of the population is considered cured or long term survivors, and consequently would not experience the event of interest. Cure rate models are seen in many fields, such as economics where it is often called the split-population model [52]. There are two main types of cure rate models, namely, the mixture model and the less studied non-mixture model [5]. Here, the mixture model will be used meaning the population is considered to be made up of two types of subjects: subjects who are cured and would not experience the event and subjects who are susceptible and are likely to experience the event. Here, the time until the

event occurs for the subjects who are susceptible will be referred to as the lifetime. The survival function for a subject's lifetime X from a cure rate mixture model is therefore of the form

$$S(t) = P(X > t) = p + (1 - p)S_0(t), \tag{2.3.1}$$

where

- p is the probability of being cured, often known as the cure fraction;
- $S_0(t)$ is the survival function of the susceptible subjects.

As $t \to \infty$, $S(t)$ does not go to 0, and it actually goes to p. So, $S(t)$ is not a proper survival function. In cure rate mixture models, there is competing risk scenarios [14] wherein there are multiple causes for the event to occur and it is the first lifetime from these causes that gives rise to the lifetime. A special case of this is the Bernoulli cure model wherein there is only one cause of the event. This is the model that is considered here.

As part of the population is cured, cured subjects have no lifetime and so will only have a censored time. Furthermore, those in the susceptible part of the population may also be right censored. This means it is not easy to identify those cured and those not cured and therefore trying to fit a distribution to the dataset is an involved task [7]. In the univariate case, maximum likelihood methods [51, 49], the EM algorithm and adaptions of the EM algorithm [16] as well as Bayesian methods [18] have all been discussed for fitting cure rate models to observed data.

2.3.2 Bivariate Cure Rate Model

Chatterjee and Shih (2001) [12] introduced the bivariate mixture cure rate model for which further work has been carried out by Wienke [63, 64]. The model discussed here is the model described by de Oliveira Peres et. al. [21]. In bivariate cure rate models, there are paired events of interest for each subject with dependence within the pairs. For example, each observation could correspond to one person and the paired events could correspond to going blind in each eye [19, 20]. In a different setting, each subject could be a set of twins and the event of interest could be each twin getting breast cancer [12, 21, 63, 64].

Previous research on the bivariate cure rate model has considered copulas to link the two outcomes [12, 21] with some other research work using specific bivariate distributions [20]. In addition, correlated frailty models have also been used in this context [63, 64].

Let the random variable X be the lifetime of event 1 in the pair and the random variable Y be the lifetime of event 2 in the pair. Furthermore, let I and J be random indicator variables representing the cure status for each event:

$$I = \begin{cases} 0 \text{ if cured from event 1,} \\ 1 \text{ if susceptible to event 1.} \end{cases} \tag{2.3.2}$$

$$J = \begin{cases} 0 \text{ if cured from event 2,} \\ 1 \text{ if susceptible to event 2.} \end{cases} \tag{2.3.3}$$

The marginal distribution for each outcome is then the standard univariate cure rate model presented in (2.3.1). This means the marginal survival functions can be written

as:

$$\begin{aligned} S_X(x) &= p_X + (1-p_X)S_{10}(x|I=1),\ x \in [0,\infty), \\ S_Y(y) &= p_Y + (1-p_Y)S_{01}(y|J=1),\ y \in [0,\infty), \end{aligned} \tag{2.3.4}$$

where $p_X = P(I=0)$ is the probability of being cured of event 1 and $p_Y = P(J=0)$ is the probability of being cured of event 2. Then,

$$S_{10}(x|I=1) = P(X > x|I=1),\ x \in [0,\infty), \tag{2.3.5}$$

is the marginal survival function when subjects are susceptible to event 1 and

$$S_{01}(y|J=1) = P(Y > y|J=1),\ y \in [0,\infty), \tag{2.3.6}$$

is the marginal survival function when subjects are susceptible to event 2. There is then a bivariate distribution for which $S_{10}(x|I=1)$ and $S_{01}(y|J=1)$ are the marginals, and it models the dependence between the events within each pair of subjects when the subject is susceptible to both events. We will denote this bivariate survival function for subjects susceptible to both events by

$$S_{11}(x,y|I=1,J=1) = P(X > x, Y > y|I=1,J=1),\ x,y \in [0,\infty). \tag{2.3.7}$$

We would not expect the probability of being susceptible to each event within a pair to be independent, and so this dependence must be taken into consideration in the construction of the joint overall survival function. The joint overall survival function

is given by [63]:

$$\begin{aligned} S(x,y) =& P(X > x, Y > y) \\ =& \phi_{11} S_{11}(x, y | I = 1, J = 1) + \phi_{10} S_{10}(x | I = 1) + \phi_{01} S_{01}(y | J = 1) + \phi_{00}, \\ & x, y \in [0, \infty), \end{aligned} \tag{2.3.8}$$

where $\phi_{11}, \phi_{10}, \phi_{01}, \phi_{00}$ are defined as follows:

- $\phi_{11} = P(I = 1, J = 1) = (1 - p_X)(1 - p_Y) + \omega$, the probability of being susceptible to both events.
- $\phi_{10} = P(I = 1, J = 0) = (1 - p_X)p_Y - \omega$, the probability of being susceptible to event 1 and cured of event 2.
- $\phi_{01} = P(I = 0, J = 1) = p_X(1 - p_Y) - \omega$, the probability of being cured of event 1 and susceptible to event 2.
- $\phi_{00} = P(I = 0, J = 0) = p_X p_Y + \omega$, the probability of being cured of both events,

where $\omega = Cov(I, J)$. We also have the following constraints:

1. $\phi_{11} + \phi_{10} + \phi_{01} + \phi_{00} = 1$ as they are the mixture probabilities and so have to sum to one.
2. $\max\{(1 - p_X)p_Y - 1, p_X(1 - p_Y) - 1\} \leq \omega \leq \min\{(1 - p_X)p_Y, p_X(1 - p_Y)\}$ to ensure that $\phi_{00}, \phi_{01}, \phi_{10}, \phi_{11} \geq 0$ as they are all probabilities.

A value of $\omega = 0$ suggests the event of an individual being susceptible to event 1 is independent of the event of being susceptible to event 2. A positive value of ω

indicates a positive correlation between being susceptible to both and a negative ω suggests a negative correlation.

We can rewrite the marginal survival functions, as seen in (2.3.4), using the parameters $\phi_{00}, \phi_{01}, \phi_{10}, \phi_{11}$ as follows:

$$\begin{aligned} S_X(x) &= \phi_{00} + \phi_{01} + (\phi_{11} + \phi_{10})S_{10}(x|I = 1),\ x \in [0, \infty), \\ S_Y(y) &= \phi_{00} + \phi_{10} + (\phi_{11} + \phi_{01})S_{01}(y|J = 1),\ y \in [0, \infty). \end{aligned} \tag{2.3.9}$$

This means that $p_X = \phi_{00} + \phi_{01}$ and $p_Y = \phi_{00} + \phi_{10}$.

As in the univariate case, there will be right censoring present in the data when subjects have censored lifetimes before they experience one or both of the events. In this thesis, we assume there is one censoring mechanism for both events within a pair so that both events have the same censoring time. This can be seen when, for example, each subject is a person and the event of interest is going blind in each eye meaning that we would expect to loose contact with the subject or the trial to end for both eyes at the same time. Furthermore, we assume Type 1 right censoring. This means we do not have full data and do not know the cure status for all subjects. If the censoring time is c for both events 1 and 2 and the true lifetimes are x and y, respectively, then the observed lifetimes will be $t_x = min(x, c)$ and $t_y = min(y, c)$. As there is a pair of events, we have four possible censoring cases, with each pair in exactly one case:

- C_1: Both uncensored;
- C_2: Event 1 uncensored, event 2 censored;
- C_3: Event 1 censored, event 2 uncensored;

- C_4: Both censored.

From these four cases and with the $\phi_{00}, \phi_{01}, \phi_{10}, \phi_{11}$ being as defined earlier, we can write the joint observed likelihood function for the bivariate cure rate dataset as follows:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}|\mathbf{X}) = &\prod_{i\in C_1} \phi_{11} f_{11}(t_{xi}, t_{yi}|I=1, J=1) \\ &\times \prod_{i\in C_4} (\phi_{11} S_{11}(t_{xi}, t_{yi}|I=1, J=1) + \phi_{10} S_{10}(t_{xi}|I=1) + \phi_{01} S_{01}(t_{yi}|J=1) + \phi_{00}) \\ &\times \prod_{i\in C_2} (-\phi_{11} S'_{11x}(t_{xi}, t_{yi}|I=1, J=1) + \phi_{10} f_{10}(t_{xi}|I=1) \\ &\times \prod_{i\in C_3} (-\phi_{11} S'_{11y}(t_{xi}, t_{yi}|I=1, J=1) + \phi_{01} f_{01}(t_{yi}|J=1)).\end{aligned} \tag{2.3.10}$$

We also have

$$f_{11}(t_x, t_y|I=1, J=1) \tag{2.3.11}$$

to be the pdf for subjects susceptible to both events. We then have

$$S'_{11x}(t_x, t_y|I=1, J=1) = \left.\frac{\partial S_{11}(x,y)}{\partial x}\right|_{x=t_x, y=t_y} \tag{2.3.12}$$

as the joint survival function (of a subject susceptible to both events) differentiated with respect to x. Similarly, we have

$$S'_{11y}(t_x, t_y|I=1, J=1) = \left.\frac{\partial S_{11}(x,y)}{\partial y}\right|_{x=t_x, y=t_y} \tag{2.3.13}$$

as the joint survival function (of a subject susceptible to both events) differentiated with respect to y. Finally, we have

$$f_{10}(t_x|I=1) \tag{2.3.14}$$

as the marginal pdf for subjects susceptible to event 1 and

$$f_{01}(t_y|J=1) \tag{2.3.15}$$

as the marginal pdf for subjects susceptible to event 2. Moreover, $S_{11}(t_{xi}, t_{yi}|I = 1, J = 1)$, $S_{10}(t_{xi}|I = 1)$ and $S_{01}(t_{yi}|J = 1)$ are as defined in (2.3.7), (2.3.5) and (2.3.6), respectively.

We can explain the contribution of each term present in (2.3.10):

- $\phi_{11} f_{11}(t_x, t_y|I = 1, J = 1)$ is for the subjects in C_1 wherein both lifetimes are uncensored. Since these lifetimes are uncensored, the subjects must be susceptible to both events. Furthermore there is no censoring and so the joint pdf for subjects susceptible to both events is used and then multiplied by the overall probability of being susceptible to both events;

- $\phi_{11} S_{11}(t_x, t_y|I = 1, J = 1)$ is for the subjects in C_4. These lifetimes for both events are censored and this is the case for subjects who are susceptible to both events. As only censored lifetimes are observed, the joint survival function is used at the censoring times. It is then multiplied by the overall probability that a subject is susceptible to both events;

- $\phi_{10} S_{10}(t_x|I = 1) = P(x > t_x|I = 1, J = 0)$ is for the subjects in C_4. In this

case, the subjects are susceptible to event 1, but cured of event 2. We therefore multiply by ϕ_{10}, the overall probability this occurs, the marginal survival function for subjects susceptible to event 1. As x is censored, we use the survival function;

- $\phi_{01}S_{01}(t_y|J=1)$ is also for subjects in C_4. This corresponds to subjects being susceptible to event 2, but cured of event 1. We therefore multiply by ϕ_{01}, the overall probability this occurs, the marginal survival function for subjects susceptible to event 2. As y is censored, we use the survival function;

- $\phi_{00} = P(I=0, J=0)$ is the final term for case C_4 and is where we assume the subjects are cured of both events. This is therefore just ϕ_{00}, the overall probability that a subject is cured of both events;

- $-\phi_{11}S'_{11x}(t_x, t_y|I=1, J=1)$ is a term in case C_2, corresponding to event 1 being uncensored and event 2 being censored. This means the subjects must be susceptible to event 1 as there is an uncensored lifetime. This first term is wherein the subjects are also susceptible to event 2. As event 1 is uncensored but event 2 is censored, the survival function is differentiated with respect to x, but not y. This is then multiplied by the overall probability of a subject being susceptible to both events;

- $\phi_{10}f_{10}(t_x|I=1)$ is the other term in case C_2 corresponding to subjects being susceptible to event 1, but cured of event 2. Event 1 is uncensored and so the marginal pdf for subjects susceptible to event 1 is used and then multiplied by the overall probability of being susceptible to event 1, but cured of event 2;

- $-\phi_{11}S'_{11y}(t_x, t_y|I=1, J=1)$ is a term in the final case C_3 wherein subjects have

a censored event 1 lifetime and an uncensored event 2 lifetime. This means the subjects must be susceptible to event 2 as there is an uncensored lifetime. This term corresponds to when the subjects are also susceptible to event 1. As event 2 is uncensored but event 1 is censored, the survival function is differentiated with respect to y, but not x. This is then multiplied by the overall probability of being susceptible to both events;

- $\phi_{01} f_{01}(t_2|J=1)$ is the other term in case C_3 representing subjects being cured of event 1, but being susceptible to event 2. Event 2 is uncensored and so the marginal pdf for subjects susceptible to event 2 is used and then multiplied by the overall probability of being cured of event 1, but susceptible to event 2.

Chapter 3

Moran-Downton Bivariate Weibull Cure Rate Model

3.1 Distributional Form

In this thesis, a transformed extension of the the Moran-Downton bivariate distribution, as introduced earlier in Section 2.1.2, is used as the joint distribution of lifetimes for subjects who are susceptible to both events in the population. The Moran-Downton bivariate exponential distribution has exponential marginals, which is restrictive in shape characteristics. The Weibull distribution would offer more flexibility so is used instead for this thesis. Exponential random variables can be transformed into Weibull random variables to give rise to the Moran-Downton bivariate Weibull distribution. A standard bivariate exponential random variable, (X_1, Y_1), is transformed into the standard bivariate Weibull random variable, (X_2, Y_2), with

shape parameters β_1, β_2 by setting $X_2 = X_1^{\frac{1}{\beta_1}}$ and $Y_2 = Y_1^{\frac{1}{\beta_2}}$. Upon using this transformation, we obtain the standard Moran-Downton bivariate Weibull pdf as

$$f_{11}(x,y) = \frac{1}{(1-\rho)} \exp\left(-\frac{x^{\beta_1}}{(1-\rho)} - \frac{y^{\beta_2}}{(1-\rho)}\right) I_0\left(\frac{2(x^{\beta_1}y^{\beta_2}\rho)^{1/2}}{(1-\rho)}\right) \beta_1\beta_2 x^{\beta_1-1}y^{\beta_2-1}, \quad x, y \in [0,\infty). \tag{3.1.1}$$

We can then readily transform it into a bivariate Weibull, (X, Y), with scale parameters α_1 and α_2 by setting $X = \alpha_1 X_2$ and $Y = \alpha_2 Y_2$. This then gives the general Moran-Downton bivariate Weibull pdf as

$$f_{11}(x,y) = \frac{1}{\alpha_1\alpha_2(1-\rho)} \exp\left(-\frac{\left(\frac{x}{\alpha_1}\right)^{\beta_1} - \left(\frac{y}{\alpha_2}\right)^{\beta_2}}{(1-\rho)}\right) I_0\left(\frac{2\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\left(\frac{y}{\alpha_2}\right)^{\beta_2}\rho\right)^{1/2}}{(1-\rho)}\right) \times \beta_1\beta_2\left(\frac{x}{\alpha_1}\right)^{\beta_1-1}\left(\frac{y}{\alpha_2}\right)^{\beta_2-1}, \ x, y \in [0,\infty). \tag{3.1.2}$$

The joint cdf of (X, Y) can then be expressed as

$$\begin{aligned} F_{11}(x,y) = &\left(1 - \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right)\right)\left(1 - \exp\left(-\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right)\right) \\ &+ \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_j^{(1)}\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) L_j^{(1)}\left(\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right)\left(\frac{x}{\alpha_1}\right)^{\beta_1}\left(\frac{y}{\alpha_2}\right)^{\beta_2} \\ &\times \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1} - \left(\frac{y}{\alpha_2}\right)^{\beta_2}\right), \ x, y \in [0,\infty). \end{aligned} \tag{3.1.3}$$

The survival function of (X, Y) is then given by

$$\begin{aligned}
S_{11}(x,y) =& -1 + \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) + \exp\left(-\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) + \left(1 - \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right)\right) \\
&\times \left(1 - \exp\left(-\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right)\right) \\
&+ \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_j^{(1)}\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) L_j^{(1)}\left(\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \left(\frac{x}{\alpha_1}\right)^{\beta_1} \left(\frac{y}{\alpha_2}\right)^{\beta_2} \\
&\times \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1} - \left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
=& \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) \exp\left(-\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) + \left(\frac{x}{\alpha_1}\right)^{\beta_1} \left(\frac{y}{\alpha_2}\right)^{\beta_2} \\
&\times \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1} - \left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_j^{(1)}\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) L_j^{(1)}\left(\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right), \\
=& \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) \exp\left(-\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&\times \left(1 + \left(\frac{x}{\alpha_1}\right)^{\beta_1} \left(\frac{y}{\alpha_2}\right)^{\beta_2} \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_j^{(1)}\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) L_j^{(1)}\left(\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right)\right), \\
& x, y \in [0, \infty),
\end{aligned} \tag{3.1.4}$$

where $L_j^{(1)}$ are Laguerre polynomials. The marginal pdfs are given by

$$\begin{aligned}
f_{10}(x) &= \frac{\beta_1}{\alpha_1} \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) \left(\frac{x}{\alpha_1}\right)^{\beta_1 - 1}, \ x \in [0, \infty), \\
f_{01}(y) &= \frac{\beta_2}{\alpha_2} \exp\left(-\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \left(\frac{y}{\alpha_2}\right)^{\beta_2 - 1}, \ y \in [0, \infty),
\end{aligned} \tag{3.1.5}$$

which are both Weibull distributions. The marginal survival functions are given by

$$S_{10}(x) = \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right),\ x \in [0, \infty), \tag{3.1.6}$$

$$S_{01}(y) = \exp\left(-\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right),\ y \in [0, \infty). \tag{3.1.7}$$

3.1.1 Correlation ρ_w

Of interest in the model is the correlation between the event times when a subject is susceptible to both events, and in the bivariate exponential Moran-Downton distribution, it is the parameter ρ [24]. As the random variables have been transformed from exponential to Weibull, the parameter ρ, which is the correlation between the exponential random variables, is no longer the correlation between the Weibull random variables. Denoting ρ_w as the correlation between the Weibull random variables (X, Y) in the Moran-Downton bivariate Weibull distribution, ρ_w can be calculated as follows [34]:

$$\rho_w = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}, \tag{3.1.8}$$

where σ_X and σ_Y are the standard deviations of X and Y, and $E(X)$ and $E(Y)$ are the expectation of X and Y, given by [31]

$$E(X) = \alpha_1 \Gamma\left(1 + \frac{1}{\beta_1}\right), \tag{3.1.9}$$

$$E(Y) = \alpha_2 \Gamma\left(1 + \frac{1}{\beta_2}\right). \tag{3.1.10}$$

Also, σ_X^2 and σ_Y^2 are given by [31]

$$\sigma_X^2 = \alpha_1^2 \left[\Gamma\left(1 + \frac{2}{\beta_1}\right) - \left(\Gamma\left(1 + \frac{1}{\beta_1}\right)\right)^2 \right], \tag{3.1.11}$$

and

$$\sigma_Y^2 = \alpha_2^2 \left[\Gamma\left(1 + \frac{2}{\beta_2}\right) - \left(\Gamma\left(1 + \frac{1}{\beta_2}\right)\right)^2 \right]. \tag{3.1.12}$$

As α_1 and α_2 are scale parameters, ρ_w will be free of them, and so for simplicity we may take $\alpha_1 = \alpha_2 = 1$, to find $E(XY)$ as follows:

$$\begin{aligned}
E(XY) =& \int_0^\infty \int_0^\infty \frac{1}{1-\rho} \exp\left(\frac{-x^{\beta_1} - y^{\beta_2}}{1-\rho}\right) I_0\left(\frac{2(x^{\beta_1} y^{\beta_2} \rho)^{1/2}}{1-\rho}\right) \beta_1 x^{\beta_1 - 1} \beta_2 y^{\beta_2 - 1} xy dx dy \\
=& \int_0^\infty \int_0^\infty \frac{1}{1-\rho} \exp\left(\frac{-x^{\beta_1} - y^{\beta_2}}{1-\rho}\right) \sum_{r=0}^\infty \frac{\frac{2^{2r}(x^{\beta_1} y^{\beta_2} \rho)^r}{(1-\rho)^{2r}}}{(2)^{2r}(r!)^2} \beta_1 x^{\beta_1 - 1} \beta_2 y^{\beta_2 - 1} xy dx dy \\
=& \frac{1}{1-\rho} \sum_{r=0}^\infty \frac{2^{2r}\rho^r}{(2)^{2r}(r!)^2(1-\rho)^{2r}} \int_0^\infty \exp\left(\frac{-y^{\beta_2}}{1-\rho}\right) y^{r\beta_2} \beta_2 y^{\beta_2 - 1} y \\
& \times \int_0^\infty \exp\left(\frac{-x^{\beta_1}}{1-\rho}\right) x^{r\beta_1} \beta_1 x^{\beta_1 - 1} x dx dy.
\end{aligned}$$

By setting $w = x^{\beta_1}$, we get

$$\begin{aligned}
E(XY) =& \frac{1}{1-\rho} \sum_{r=0}^\infty \frac{2^{2r}\rho^r}{(2)^{2r}(r!)^2(1-\rho)^{2r}} \int_0^\infty \exp\left(\frac{-y^{\beta_2}}{1-\rho}\right) y^{r\beta_2} \beta_2 y^{\beta_2 - 1} y \\
& \times \int_0^\infty \exp\left(\frac{-w}{1-\rho}\right) w^r w^{\frac{1}{\beta_1} + 1 - 1} dw dy \\
=& \frac{1}{1-\rho} \sum_{r=0}^\infty \frac{2^{2r}\rho^r}{(2)^{2r}(r!)^2(1-\rho)^{2r}} \Gamma\left(r + \frac{1}{\beta_1} + 1\right) (1-\rho)^{r + \frac{1}{\beta_1} + 1} \\
& \times \int_0^\infty \exp\left(\frac{-y^{\beta_2}}{1-\rho}\right) y^{r\beta_2} \beta_2 y^{\beta_2 - 1} y dy.
\end{aligned}$$

Now, by setting $w = y^{\beta_2}$, we get

$$\begin{aligned}E(XY) =&\frac{1}{1-\rho}\sum_{r=0}^{\infty}\frac{2^{2r}\rho^r}{(2)^{2r}(r!)^2(1-\rho)^{2r}}\Gamma\left(r+\frac{1}{\beta_1}+1\right)(1-\rho)^{r+\frac{1}{\beta_1}+1}\Gamma\left(r+\frac{1}{\beta_2}+1\right)\\ &\times(1-\rho)^{r+\frac{1}{\beta_2}+1}\\ =&(1-\rho)^{\frac{1}{\beta_1}+\frac{1}{\beta_2}+1}\sum_{r=0}^{\infty}\frac{\rho^r}{(r!)^2}\Gamma\left(r+\frac{1}{\beta_1}+1\right)\Gamma\left(r+\frac{1}{\beta_2}+1\right).\end{aligned} \tag{3.1.13}$$

Thus, we finally obtain

$$\rho_w=\frac{(1-\rho)^{\frac{1}{\beta_1}+\frac{1}{\beta_2}+1}\sum_{r=0}^{\infty}\frac{\rho^r}{(r!)^2}\Gamma\left(r+\frac{1}{\beta_1}+1\right)\Gamma\left(r+\frac{1}{\beta_2}+1\right)-\Gamma\left(1+\frac{1}{\beta_1}\right)\Gamma\left(1+\frac{1}{\beta_2}\right)}{\left[\Gamma\left(1+\frac{2}{\beta_1}\right)-\left(\Gamma\left(1+\frac{1}{\beta_1}\right)\right)^2\right]^{\frac{1}{2}}\left[\Gamma\left(1+\frac{2}{\beta_2}\right)-\left(\Gamma\left(1+\frac{1}{\beta_2}\right)\right)^2\right]^{\frac{1}{2}}}. \tag{3.1.14}$$

The support for ρ_w is $[0,1)$. For $\beta_1=\beta_2=1$, the bivariate exponential case, this formula simplifies to $\rho_w=\rho$, as expected. For values of $\beta_1,\beta_2\neq 1$, we have $\rho_w<\rho$, as seen by Figure 3.1. Moreover, if $\rho=0$, we have $\rho_w=0$, as seen in (3.1.15).

$$
\begin{aligned}
\rho_w &= \frac{(1-0)^{\frac{1}{\beta_1}+\frac{1}{\beta_2}+1}\sum_{r=0}^{\infty}\frac{0^r}{(r!)^2}\Gamma\left(r+\frac{1}{\beta_1}+1\right)\Gamma\left(r+\frac{1}{\beta_2}+1\right)-\Gamma\left(1+\frac{1}{\beta_1}\right)\Gamma\left(1+\frac{1}{\beta_2}\right)}{\left[\Gamma\left(1+\frac{2}{\beta_1}\right)-\left(\Gamma\left(1+\frac{1}{\beta_1}\right)\right)^2\right]^{\frac{1}{2}}\left[\Gamma\left(1+\frac{2}{\beta_2}\right)-\left(\Gamma\left(1+\frac{1}{\beta_2}\right)\right)^2\right]^{\frac{1}{2}}} \\
&= \frac{\frac{0^0}{(0!)^2}\Gamma\left(0+\frac{1}{\beta_1}+1\right)\Gamma\left(0+\frac{1}{\beta_2}+1\right)-\Gamma\left(1+\frac{1}{\beta_1}\right)\Gamma\left(1+\frac{1}{\beta_2}\right)}{\left[\Gamma\left(1+\frac{2}{\beta_1}\right)-\left(\Gamma\left(1+\frac{1}{\beta_1}\right)\right)^2\right]^{\frac{1}{2}}\left[\Gamma\left(1+\frac{2}{\beta_2}\right)-\left(\Gamma\left(1+\frac{1}{\beta_2}\right)\right)^2\right]^{\frac{1}{2}}} \\
&= \frac{\Gamma\left(\frac{1}{\beta_1}+1\right)\Gamma\left(\frac{1}{\beta_2}+1\right)-\Gamma\left(1+\frac{1}{\beta_1}\right)\Gamma\left(1+\frac{1}{\beta_2}\right)}{\left[\Gamma\left(1+\frac{2}{\beta_1}\right)-\left(\Gamma\left(1+\frac{1}{\beta_1}\right)\right)^2\right]^{\frac{1}{2}}\left[\Gamma\left(1+\frac{2}{\beta_2}\right)-\left(\Gamma\left(1+\frac{1}{\beta_2}\right)\right)^2\right]^{\frac{1}{2}}} \\
&= 0
\end{aligned}
\tag{3.1.15}
$$

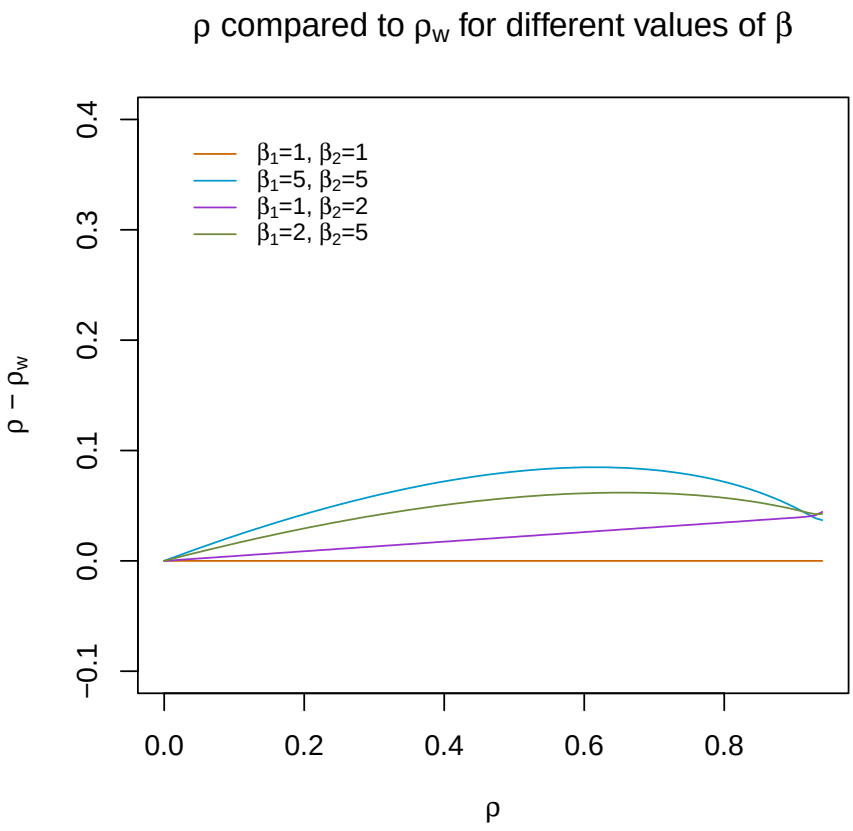

Figure 3.1: $\rho - \rho_w$ for Various Combinations of $\beta's$

Chapter 4

EM Algorithm

Given a dataset, to find the MLE for the parameters in the Moran-Downton bivariate Weibull cure rate model described in Chapter 3, an EM algorithm, like the one described earlier in Section 2.2.2, is derived and implemented here in this chapter.

4.1 EM Implementation

Expectation Step

Here, we treat cure status as the latent variable as it is unobserved in the censored cases. Assuming that we have a dataset with cure status included as well, we can introduce four indicator random variables corresponding to the bivariate cure status:

- $k_1 = \begin{cases} 1 \text{ if subject is cured of events 1 and 2,} \\ 0 \text{ otherwise;} \end{cases}$

- $k_2 = \begin{cases} 1 \text{ if subject is cured of event 1, but is susceptible to Event 2,} \\ 0 \text{ otherwise;} \end{cases}$
- $k_3 = \begin{cases} 1 \text{ if subject is susceptible to event 1, but is cured of event 2,} \\ 0 \text{ otherwise;} \end{cases}$
- $k_4 = \begin{cases} 1 \text{ if subject is susceptible to both events 1 and 2,} \\ 0 \text{ otherwise.} \end{cases}$

Then when we know the cure status, despite the censored values, the complete data likelihood function can be expressed as

$$\begin{aligned} C\mathcal{L}(\boldsymbol{\theta}|\mathbf{X}) = & \prod_{i \in C_1} (\phi_{11} f_{11}(t_{xi}, t_{yi}|I=1, J=1))^{k_{4i}} \\ & \times \prod_{i \in C_4} (\phi_{11} S_{11}(t_{xi}, t_{yi}|I=1, J=1))^{k_{4i}} (\phi_{10} S_{10}(t_{xi}|I=1))^{k_{3i}} (\phi_{01} S_{01}(t_{yi}|J=1))^{k_{2i}} \phi_{00}^{k_{1i}} \\ & \times \prod_{i \in C_2} (-\phi_{11} S'_{11x}(t_{xi}, t_{yi}|I=1, J=1))^{k_{4i}} (\phi_{10} f_{10}(t_{xi}|I=1))^{k_{3i}} \\ & \times \prod_{i \in C_3} (-\phi_{11} S'_{11y}(t_{xi}, t_{yi}|I=1, J=1))^{k_{4i}} (\phi_{01} f_{01}(t_{yi}|J=1))^{k_{2i}}. \end{aligned} \tag{4.1.1}$$

Then the complete data log-likelihood follows readily from (4.1.1) as

$$\begin{aligned} Log(\mathcal{CL}(\boldsymbol{\theta}|\mathbf{X})) = &\sum_{i\in C_1} k_{4i} log(\phi_{11} f_{11}(t_{xi}, t_{yi}|I=1, J=1) \\ &+ \sum_{i\in C_4} k_{4i} log(\phi_{11} S_{11}(t_{xi}, t_{yi}|I=1, J=1)) + k_{3i} log(\phi_{10} S_{10}(t_{xi}|I=1) \\ &\quad + k_{2i} log(\phi_{01} S_{01}(t_{yi}|J=1)) + k_{1i} log(\phi_{00})) \\ &+ \sum_{i\in C_2} k_{4i} log(-\phi_{11} S'_{11x}(t_{xi}, t_{yi}|I=1, J=1)) + k_{3i} log(\phi_{10} f_{10}(t_{xi}|I=1)) \\ &+ \sum_{i\in C_3} k_{4i} log(-\phi_{11} S'_{11y}(t_{xi}, t_{yi}|I=1, J=1)) + k_{2i} log(\phi_{01} f_{01}(t_{yi}|J=1)). \end{aligned} \tag{4.1.2}$$

This means the conditional expectation of the random variables $(k_{1i}, k_{2i}, k_{3i}, k_{4i})$, for each observation i, need to be calculated in the Expectation step. These conditional expectations will depend on the cases, C_1, C_2, C_3, C_4, as explained below.

As subjects in C_1 have both lifetimes uncensored, the subject is susceptible to both events, and so $E(k_{4i}|C_1) = 1$. Also, in this case, it is clear the subject cannot be cured of either of the events and so other expectations are all 0; that is

$$\begin{aligned} E(k_{1i}|C_1) &= 0, \\ E(k_{2i}|C_1) &= 0, \\ E(k_{3i}|C_1) &= 0, \\ E(k_{4i}|C_1) &= 1. \end{aligned} \tag{4.1.3}$$

In C_2, the subjects have an uncensored lifetime for event 1 and a censored lifetime for event 2. This means the subject must be susceptible to event 1 and cannot be cured

leading to conditional probabilities of 0 for $E(k_{1i}|C_2)$ and $E(k_{2i}|C_2)$. The conditional probability of being cured of event 2 while being susceptible to event 1 uses the overall probability of being susceptible to event 1 and cured of event 2 multiplied by the marginal pdf for the subjects susceptible to event 1. The conditional probability of being susceptible to both uses the overall probability of being susceptible to both multiplied by the survival function differentiated with respect to x as event 1 has been observed while event 2 is censored. As the conditional probabilities must sum to 1, the denominator is the sum of these conditional probabilities, and so we have

$$\begin{aligned} E(k_{1i}|C_2) &= 0, \\ E(k_{2i}|C_2) &= 0, \\ E(k_{3i}|C_2) &= \frac{\phi_{10} f_{10}(t_{xi})}{-\phi_{11} S'_{11x}(t_{xi}, t_{yi}) + \phi_{10} f_{10}(t_{xi})}, \\ E(k_{4i}|C_2) &= \frac{-\phi_{11} S'_{11x}(t_{xi}, t_{yi})}{-\phi_{11} S'_{11x}(t_{xi}, t_{yi}) + \phi_{10} f_{10}(t_{xi})}. \end{aligned} \tag{4.1.4}$$

In C_3, the lifetime for event 1 is censored and the lifetime for event 2 is uncensored. This means the subject must be susceptible to event 2 and cannot be cured leading to conditional probabilities of 0 for $E(k_{1i}|C_3)$ and $E(k_{3i}|C_3)$. For the conditional probability of being susceptible to both events, the joint survival function differentiated with respect to y is used as event 2 is observed while event 1 is censored. This is multiplied by the overall probability of a subject being susceptible to both events. For the conditional probability that the subject is cured of event 1 and susceptible to event 2, the pdf for subjects susceptible to event 2 is used and is then multiplied by the overall probability of being cured of event 1 and being susceptible to event 2. As before, the denominator is the sum of the conditional probabilities, and thus we have

$$
\begin{aligned}
E(k_{1i}|C_3) &= 0, \\
E(k_{2i}|C_3) &= \frac{\phi_{01} f_{01}(t_{yi})}{-\phi_{11} S'_{11y}(t_{xi}, t_{yi}) + \phi_{01} f_{01}(t_{yi})}, \\
E(k_{3i}|C_3) &= 0, \\
E(k_{4i}|C_3) &= \frac{-\phi_{11} S'_{11y}(t_{xi}, t_{yi})}{-\phi_{11} S'_{11y}(t_{xi}, t_{yi}) + \phi_{01} S_{01}(t_{yi})}.
\end{aligned}
\tag{4.1.5}
$$

If both lifetimes are censored, as in C_4, then any scenario of being cured or susceptible to either events is possible. The numerator for the conditional probability of being cured of both events is the overall probability of being cured of both events, ϕ_{00}. The numerator for the conditional probability of being cured of event 1 and susceptible to event 2 uses the overall probability of being cured of event 1 and susceptible to event 2 multiplied by the marginal survival function for a subject susceptible to event 2. The numerator for the conditional probability of being susceptible to event 1 and cured of event 2 uses the overall probability of being susceptible to event 1 and cured of event 2 multiplied by the marginal survival function for being susceptible to event 1. The final conditional probability of being susceptible to both uses the joint survival function at the censored values. As before, the denominator for all the conditional

probabilities is the sum of all the conditional probabilities, and thus we have

$$\begin{aligned}
E(k_{1i}|C_4) &= \frac{\phi_{00}}{\phi_{11}S_{11}(t_{xi},t_{yi}) + \phi_{10}S_{10}(t_{xi}) + \phi_{01}S_{01}(t_{yi}) + \phi_{00}}, \\
E(k_{2i}|C_4) &= \frac{\phi_{01}S_{01}(t_{yi})}{\phi_{11}S_{11}(t_{xi},t_{yi}) + \phi_{10}S_{10}(t_{xi}) + \phi_{01}S_{01}(t_{yi}) + \phi_{00}}, \\
E(k_{3i}|C_4) &= \frac{\phi_{10}S_{10}(t_{xi})}{\phi_{11}S_{11}(t_{xi},t_{yi}) + \phi_{10}S_{10}(t_{xi}) + \phi_{01}S_{01}(t_{yi}) + \phi_{00}}, \\
E(k_{4i}|C_4) &= \frac{\phi_{11}S_{11}(t_{xi},t_{yi})}{\phi_{11}S_{11}(t_{xi},t_{yi}) + \phi_{10}S_{10}(t_{xi}) + \phi_{01}S_{01}(t_{yi}) + \phi_{00}}.
\end{aligned} \tag{4.1.6}$$

Maximization Step

We then maximize the Q-function with respect to the model parameters, which in this case are $\alpha_1, \alpha_2, \beta_1, \beta_2, \rho, \phi_{00}, \phi_{01}, \phi_{10}, \phi_{11}$. Recall from Section 2.2.2 that the Q-function is the expected complete log-likelihood function using the conditional distribution of the cure status, given the current parameter estimates. From (4.1.2), the Q-function is as follows:

$$\begin{aligned}
Q(\boldsymbol{\theta}) = &\sum_{i\in C_1} E(k_{4i}|C_1) log(\phi_{11} f_{11}(t_{xi},t_{yi})) \\
&+ \sum_{i\in C_4} E(k_{4i}|C_4) log(\phi_{11}S_{11}(t_{xi},t_{yi})) + E(k_{3i}|C_4) log(\phi_{10}S_{10}(t_{xi})) \\
&\quad + E(k_{2i}|C_4) log(\phi_{01}S_{01}(t_{yi})) + E(k_{1i}|C_4) log(\phi_{00}) \\
&+ \sum_{i\in C_2} E(k_{4i}|C_2) log(-\phi_{11}S'_{11x}(t_{xi},t_{yi})) + E(k_{3i}|C_2) log(\phi_{10} f_{10}(t_{xi})) \\
&+ \sum_{i\in C_3} E(k_{4i}|C_3) log(-\phi_{11}S'_{11y}(t_{xi},t_{yi})) + E(k_{2i}|C_3) log(\phi_{01} f_{01}(t_{yi})).
\end{aligned} \tag{4.1.7}$$

The joint survival function, differentiated with respect to x, present in (4.1.7) has the following expression:

$$
\begin{aligned}
S'_{11x}(x,y) = &- \frac{\beta_1}{\alpha_1^{\beta_1}} x^{\beta_1-1} \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) \exp\left(-\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&- \frac{\beta_1 x^{\beta_1-1}}{\alpha_1^{\beta_1}} \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_j^{(1)}\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) L_j^{(1)}\left(\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \left(\frac{x}{\alpha_1}\right)^{\beta_1} \left(\frac{y}{\alpha_2}\right)^{\beta_2} \\
&\times \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1} - \left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) + \frac{\beta_1}{x} \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_j^{(1)}\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) L_j^{(1)}\left(\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&\times \left(\frac{x}{\alpha_1}\right)^{\beta_1} \left(\frac{y}{\alpha_2}\right)^{\beta_2} \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1} - \left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&- \frac{\beta_1 x^{\beta_1-1}}{x\alpha_1^{\beta_1}} \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_{j-1}^{(2)}\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) L_j^{(1)}\left(\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \left(\frac{x}{\alpha_1}\right)^{\beta_1} \left(\frac{y}{\alpha_2}\right)^{\beta_2} \\
&\times \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1} - \left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
= &- \frac{\beta_1}{\alpha_1^{\beta_1}} x^{\beta_1-1} \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) \exp\left(-\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&+ \left(\frac{\beta_1}{x} - \frac{\beta_1 x^{\beta_1-1}}{\alpha_1^{\beta_1}}\right) \left(\frac{x}{\alpha_1}\right)^{\beta_1} \left(\frac{y}{\alpha_2}\right)^{\beta_2} \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1} - \left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&\times \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_j^{(1)}\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) L_j^{(1)}\left(\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&- \frac{\beta_1 x^{\beta_1-1}}{\alpha_1^{\beta_1}} \left(\frac{x}{\alpha_1}\right)^{\beta_1} \left(\frac{y}{\alpha_2}\right)^{\beta_2} \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1} - \left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&\times \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_{j-1}^{(2)}\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) L_j^{(1)}\left(\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right).
\end{aligned}
\tag{4.1.8}
$$

By symmetry, the joint survival function differentiated with respect to y is simply given by

$$\begin{aligned}
S'_{11y}(x,y) = &- \frac{\beta_2}{\alpha_2^{\beta_2}} y^{\beta_2-1} \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) \exp\left(-\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&+ \left(\frac{\beta_2}{y} - \frac{\beta_2 y^{\beta_2-1}}{\alpha_2^{\beta_2}}\right) \left(\frac{x}{\alpha_1}\right)^{\beta_1} \left(\frac{y}{\alpha_2}\right)^{\beta_2} \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1} - \left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&\times \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_j^{(1)}\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) L_j^{(1)}\left(\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&- \frac{\beta_2 y^{\beta_2-1}}{\alpha_2^{\beta_2}} (\frac{x}{\alpha_1})^{\beta_1} \left(\frac{y}{\alpha_2}\right)^{\beta_2} \exp\left(-\left(\frac{x}{\alpha_1}\right)^{\beta_1} - \left(\frac{y}{\alpha_2}\right)^{\beta_2}\right) \\
&\times \sum_{j=0}^{\infty} \frac{\rho^{j+1}}{(j+1)^2} L_j^{(1)}\left(\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right) L_{j-1}^{(2)}\left(\left(\frac{y}{\alpha_2}\right)^{\beta_2}\right).
\end{aligned} \tag{4.1.9}$$

The other terms in (4.1.7) are as given earlier in (3.1.7), (3.1.6), (3.1.4), (2.3.4) and (2.3.9).

The Q-function can then be maximized with respect to each parameter, subject to the following constraints:

1. $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (0, \infty)$;
2. $\phi_{00}, \phi_{01}, \phi_{10}, \phi_{11} \in [0, 1]$ as they are probabilities;
3. $\rho \in [0, 1)$ as dictated by the Moran-Downton bivariate exponential distribution;
4. $\phi_{00} + \phi_{01} + \phi_{10} + \phi_{11} = 1$ as the probabilities have to sum to 1 as the subjects have to belong to one of the four cases.

Due to the complicated form of the Q-function, it is not practical to differentiate it with respect to each parameter and then solve the obtained system of equations. So,

numerical optimization is used. Here, optim in R with the L-BFGS-B method is used with no gradient given, and hence the optim function also estimates the gradient of the function via finite differences [50]. L-BFGS-B is a quasi-Newton algorithm that does not require the second derivatives of the function and instead estimates the Hessian [10]. However, the constraints still need to be satisfied and in particular, constraints 2 and 4 on $\phi_{00}, \phi_{01}, \phi_{10}, \phi_{11}$ are the most complicated to satisfy. We cannot just maximize the Q-function with respect to each parameter as it may not ensure that the sum of $\phi_{00}, \phi_{01}, \phi_{10}, \phi_{11}$ is equal to 1. If we write ϕ_{11} as $1 - \phi_{00} - \phi_{01} - \phi_{10}$ it will satisfy constraint 4, but can cause $\phi_{11} < 0$ and violate constraint 2. For the L-BFGS-B numerical optimization, as well as other numerical optimization, we can only specify bounds on individual parameters as opposed to combinations of parameters. This means we cannot implement constraint 4 in its current form. However, we may use new dummy parameters such that by placing bounds on them, it would enforce the constraints on the real parameters. This procedure then proceeds as follows:

- Let $a = \phi_{00}$ with the support of $a = (0, 1)$. This directly satisfies the constraint that $\phi_{00} \in (0, 1)$;
- Let $b = \frac{1-\phi_{00}-\phi_{01}}{1-\phi_{00}}$ with support of $b = (0, 1)$. This means that $0 < 1 - \phi_{00} - \phi_{01} \implies \phi_{00} + \phi_{01} < 1$ and $1 - \phi_{00} - \phi_{01} < 1 - \phi_{00} \implies 0 < \phi_{01}$. As $\phi_{00} \in (0, 1)$, this means that $\phi_{01} \in (0, 1)$ as well;
- Let $c = \frac{1-\phi_{00}-\phi_{01}-\phi_{10}}{1-\phi_{00}-\phi_{01}}$ with support of $c = (0, 1)$. This means that $0 < 1 - \phi_{00} - \phi_{01} - \phi_{10} \implies \phi_{00} + \phi_{01} + \phi_{10} < 1$ and $1 - \phi_{00} - \phi_{01} - \phi_{10} < 1 - \phi_{00} - \phi_{01} \implies 0 < \phi_{10}$. As we already have the previous constraint that $\phi_{00} + \phi_{01} < 1$, this can be combined to give $\phi_{10} \in (0, 1)$ as well. As $\phi_{00}, \phi_{01}, \phi_{10} \in (0, 1)$ and $\phi_{00} + \phi_{01} + \phi_{10} < 1$, this means that by calculating $\phi_{11} = 1 - \phi_{00} - \phi_{01} - \phi_{10}$,

$\phi_{11} \in (0,1)$ and $\phi_{00} + \phi_{01} + \phi_{10} + \phi_{11} = 1$ and so constraints 2 and 4 are both satisfied.

To then estimate the original parameters, we invert the above equations to get

- $\phi_{00} = a$,
- $\phi_{01} = 1 - \phi_{00} - b \times (1 - \phi_{00})$,
- $\phi_{10} = 1 - \phi_{00} - \phi_{01} \times c \times (1 - \phi_{00} - \phi_{01})$,
- $\phi_{11} = 1 - \phi_{00} - \phi_{01} - \phi_{10}$.

The other constraints are trivial to apply.

If the optimization algorithm requires unconstrained optimization, then the logit function can be used to convert a, b, c to new dummy parameters which instead have support of $(-\infty, \infty)$ whilst still maintaining the constraints on $\phi_{00}, \phi_{01}, \phi_{10}$ and ϕ_{11}. A logit transformation can also be used on ρ and log-transformation can be used on $\alpha_1, \alpha_2, \beta_1, \beta_2$ as well to convert to functions with support $(-\infty, \infty)$. The optimization algorithm can then be implemented with these dummy parameters and then the obtained results can be transformed back to the original parameters once the optimization is complete. The parameter estimates for $\alpha_1, \alpha_2, \beta_1, \beta_2, \rho, \phi_{00}, \phi_{01}, \phi_{10}, \phi_{11}$ are then updated.

Using these new parameter estimates, new conditional probabilities of $k_{1i}, k_{2i}, k_{3i}, k_{4i}$ can be calculated for each subject. Then, the Expectation and Maximization Steps are repeated until convergence is achieved.

End point

Once two consecutive values of the observed log-likelihood function are within $\epsilon = 0.001$, the algorithm terminates, and the final parameter estimates are returned. The expression of the observed log-likelihood is

$$L(\theta) = \sum_{i \in c_1} log(\phi_{11} f_{11}(t_{1i}, t_{2i})) + \sum_{i \in c_4} log(\phi_{11} S_{11}(t_{1i}, t_{2i}) + \phi_{10} S_{10}(t_{1i}) + \phi_{01} S_{01}(t_{2i}) + \phi_{00}) + \sum_{i \in c_2} log(-\phi_{11} S'_{11x}(t_{1i}, t_{2i}) + \phi_{10} f_{10}(t_{1i})) + \sum_{i \in c_3} log(-\phi_{11} S'_{11y}(t_{1i}, t_{2i}) + \phi_{01} f_{01}(t_{2i})). \tag{4.1.10}$$

The R code for this algorithm has been presented in the Appendix.

4.2 Starting Parameter Values

The EM algorithm can be sensitive to initial values as the algorithm can converge to a local and not global maximum [8]. Here, to find the model parameter estimates, a Weibull distribution can be fitted to the uncensored values for each event using the traditional MLE method. This will give starting parameters for $\alpha_1, \alpha_2, \beta_1$ and β_2. The observations can then be transformed into exponential variables and ρ can be calculated as the correlation between these uncensored values. If there are no pairs where both lifetimes are uncensored, then ρ is taken to be 0.5. There is not a clear method to find starting parameters for $\phi_{00}, \phi_{10}, \phi_{01}, \phi_{11}$ from a dataset and so 0.25 is used as the starting values of $\phi_{00}, \phi_{10}, \phi_{01}, \phi_{11}$; this ensures that the constraints are met. Upon setting $\phi_{00} = \phi_{01} = \phi_{10} = \phi_{11} = 0.25$, we get $a = 0.25$, $b = \frac{2}{3}$ and $c = \frac{1}{2}$ as the starting values for the dummy parameters.

Chapter 5

Empirical Study

In order to test the proposed algorithm, as well as to generate data replications for the parametric bootstrap method, sampling from the desired cure rate model is necessary. Once datasets have been simulated, the algorithm described in Chapter 4 can be implemented on each dataset and the parameter estimates can be obtained to dertermine the performance measures such as Bias and Mean Squared Error (MSE).

5.1 Data Simulation

To simulate a dataset, first of all, a bivariate cure status is generated for each subject. Then, for subjects susceptible to both events lifetimes are randomly generated from a Moran-Downton Bivariate Weibull distribution. For subjects susceptible to just one event the lifetime for this event will be randomly generated from a Weibull distribution. Next, the censoring times are randomly generated from an exponential distribution with rate parameter λ, with the censoring mechanism being independent

of lifetime. Finally, the lifetime will be compared to the censoring time and the smallest value is selected as the observed time for the susceptible subjects. For subjects who are cured, the observed time automatically will be set as the censored time.
To generate datasets when $\alpha_1, \alpha_2, \beta_1, \beta_2, \rho, \phi_{00}, \phi_{01}, \phi_{10}, \phi_{11}$ and λ are all specified the following algorithm is used, and the corresponding R code is presented in the Appendix:

1. Generate u from a uniform(0,1) distribution;

2. If $u < \phi_{00}$, then generate the censoring time from an exponential distribution with parameter λ for t_1 and t_2;

3. If $\phi_{00} < u < \phi_{00} + \phi_{01}$, then generate a lifetime, x, for event 1 from a marginal Weibull distribution with parameters α_1, β_1, and censoring time from an exponential distribution with parameter λ. Set t_1 as the smallest value out of censoring time and lifetime x. Set t_2 as the censoring time;

4. If $\phi_{00} + \phi_{01} < u < \phi_{00} + \phi_{01} + \phi_{10}$, then generate a lifetime, y, for event 2 from a marginal Weibull distribution with parameters α_2, β_2, and a censoring time from an exponential distribution with parameter λ. Set t_2 as the smallest value out of the censoring time and lifetime y and set t_1 as the censoring time;

5. If $\phi_{00} + \phi_{01} + \phi_{10} < u$, then generate two lifetimes, x_1 and y_1, from the Moran-Downton standard bivariate exponential distribution, using the MDBED package in R [25] with parameters $1, 1, \rho$. Then, transform to Weibull random variables by setting $x = \alpha_1 x_1^{\frac{1}{\beta_1}}$ and $y = \alpha_2 y_1^{\frac{1}{\beta_2}}$. Generate a censoring time from an exponential distribution with parameter λ. Now set t_1 as the smallest value

out of the censoring time and lifetime x, and set t_2 as the smallest out of the lifetime y and the censoring time;

6. Return t_1 and t_2 and the censoring indicators that indicate whether the returned values are lifetimes or censored times;

7. Repeat Steps 1-6 until a sample of suitable size has been generated.

5.2 Censoring Rate Compared to Cure Rate

For each dataset there will be some proportion of the observations that are censored. In the univariate/ marginal case the censoring proportion is the number of censored observations divided by the total number of observations. In the bivariate case, here censoring proportion will refer to the proportion of observations that have censored lifetimes for both events. There is a relationship between the parameters and the censoring proportion. For parametric bootstrap methods, the censoring proportion of the dataset to be replicated will be known, but the exponential rate parameter that would generate this censoring proportion will be unknown. For this reason, a suitable parameter for the censoring distribution will need to be found so that a dataset with the correct censoring proportion can be simulated. As the marginals of the cure rate model are univariate cure models, the probability of each event being censored can be seen as follows.

We let the random variable $C \sim f_c(c)$, $c > 0$, with $f_c(c) = \lambda exp(-\lambda c)$, represent the time of censoring for both events in a pair. For the random variable of lifetime for subjects susceptible to event 1, we use $X \sim f_{10}(x)$ with $f_{10}(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta_1 - 1} exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right)$, $x > 0$. As X and C are independent, their joint pdf is $f_{10}(x)f_c(c)$. If the subject is

cured of event 1, then the lifetime must be censored for event 1. This yields:

$$\begin{aligned}
P(\text{event 1 censored}) =& P(C < \text{observed}) \\
=& p_X + (1 - p_X)P(C < X) \\
=& \phi_{00} + \phi_{01} + (\phi_{10} + \phi_{11})P(C < X) \\
=& \phi_{00} + \phi_{01} + (\phi_{10} + \phi_{11}) \int_0^\infty \int_0^x f_{10}(x) f_c(c) dc dx \\
=& \phi_{00} + \phi_{01} + (\phi_{10} + \phi_{11}) \\
& \times \int_0^\infty \int_0^x \lambda exp(-\lambda c) \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta_1 - 1} exp\left(-\left(\frac{x}{\alpha}\right)^\beta\right) dc dx \\
=& \phi_{00} + \phi_{01} + (\phi_{10} + \phi_{11}) \left[1 - \int_0^\infty \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta_1 - 1} exp\left(-\left(\frac{x}{\alpha}\right)^\beta - \lambda x\right) dx\right].
\end{aligned} \tag{5.2.1}$$

We cannot analytically compute the above integral for all values of α and β.

For $\beta = 1$, i.e., in the exponential case, we have

$$\begin{aligned}
P(\text{event 1 censored}) &= \phi_{00} + \phi_{01} + (\phi_{10} + \phi_{11})P(c < X) \\
&= \phi_{00} + \phi_{01} + (\phi_{10} + \phi_{11})[1 - \int_0^\infty \frac{1}{\alpha} exp\left(-\left(\frac{x}{\alpha}\right) - \lambda x\right) dx] \\
&= \phi_{00} + \phi_{01} + (\phi_{10} + \phi_{11})[1 - \int_0^\infty \frac{1}{\alpha} exp\left(-\left(\frac{1}{\alpha} + \lambda\right) x\right) dx] \\
&= \phi_{00} + \phi_{01} + (\phi_{10} + \phi_{11}) \left(1 + \left[\frac{1}{\frac{1}{\alpha} + \lambda} exp\left(-\left(\frac{1}{\alpha} + \lambda\right) x\right)\right]_{x=0}^{x=\infty}\right) \\
&= \phi_{00} + \phi_{01} + (\phi_{10} + \phi_{11}) \left(1 - \frac{1}{\frac{1}{\alpha} + \lambda}\right).
\end{aligned} \tag{5.2.2}$$

Rearranging for λ, we get

$$\lambda = \frac{1 - P(\text{event 1 censored}) - \alpha + (\phi_{00} + \phi_{01}) \times \alpha}{\alpha \times (P(\text{event 1 censored}) - 1)} \tag{5.2.3}$$

This is the relationship between the censoring proportion and the parameters in the marginal distributions. For the censoring proportion of event 2, by symmetry, there is the following relationship:

$$\begin{aligned} P(\text{event 2 censored}) &= \phi_{00} + \phi_{10} + (\phi_{01} + \phi_{11}) \int_0^\infty \int_c^\infty f_c(c) f_{01}(y) dy dc \\ &= \phi_{00} + \phi_{10} + (\phi_{01} + \phi_{11}) \int_0^\infty f_c(c) S_{01}(c) dc. \end{aligned} \tag{5.2.4}$$

In the bivariate case, there is also the probability that both lifetimes can be censored. If the subject is cured of both events, then the observed times will automatically be the censoring times, and the probability of this is ϕ_{00}. If the subject is susceptible to event 1 but cured of event 2, then the subject will be censored if the lifetime of event 1 is greater than the censoring time, which is given by $P(X > C|I = 1)$. If the subject is cured of event 1 but susceptible to event 2, then the subject will be censored if the lifetime of event 2 is greater than the censoring time, which is given by $P(Y > C|J = 1)$. If the subject is susceptible to both events, then the probability of being censored is $P(X > C, Y > C|I = 1, J = 1)$. Combining all these cases

together, we obtain

$$\begin{aligned}
P(\text{both events censored}) =& \phi_{00} + \phi_{10}P(X > C|I = 1) + \phi_{01}P(Y > C|J = 1) \\
&+ \phi_{11}P(X > C, Y > C|I = 1, J = 1) \\
=& \phi_{00} + \phi_{10}\int_0^\infty \int_c^\infty f_c(c) f_{10}(x) dx dc + \phi_{01}\int_0^\infty \int_c^\infty f_c(c) f_{01}(y) dy dc \\
&+ \phi_{11}\int_0^\infty \int_c^\infty \int_c^\infty f_c(c) f_{11}(x, y) dx dy dc \\
=& \phi_{00} + \phi_{10}\int_0^\infty f_c(c) S_{10}(c) dc + \phi_{01}\int_0^\infty f_c(c) S_{01}(c) dc \\
&+ \phi_{11}\int_0^\infty f_c(c) S_{11}(c, c) dc.
\end{aligned} \tag{5.2.5}$$

The functions $S_{10}(c), S_{01}(c)$ and $S_{11}(c, c)$ are as given in (3.1.6), (3.1.7) and (3.1.4), respectively. If a specific censoring proportion, C_p, is required, then we need to find λ such that:

$$P(\text{both events censored}) - C_p = 0. \tag{5.2.6}$$

Using (5.2.5) a root finding algorithm can been used to find λ from (5.2.6).

5.3 Results

Four different sets of parameters with three different samples sizes, $n = 100$, $n = 50$ and $n = 25$, were used in the simulation study, with a total of 500 datasets generated for each parameter setting. Low ($\rho = 0.3$) and high ($\rho = 0.7$) correlation as well as low ($\lambda = 0.2$) and high ($\lambda = 0.7$) censoring levels were used to simulate data with censoring proportions and correlation for each of the parameter settings as presented in Table 5.1. The mean was calculated by taking the mean of the parameter estimates.

The bias was then calculated by taking the mean of the parameter estimates and subtracting the true value of the parameter. The MSE was calculated by subtracting the true value from each parameter estimate, squaring and then taking the mean of all these values. The obtained results are presented in Tables 5.3-5.5.

Parameter $(\alpha_1, \alpha_2, \beta_1, \beta_2, \rho)$	λ	Events censored			
		1 and 2	Marginally 1	Marginally 2	Neither
(1, 1.5, 1.5, 2, 0.3)	0.2	0.3246	0.5628	0.5976	0.1642
(1, 1.5, 1.5, 2, 0.7)	0.2	0.3758	0.5628	0.5976	0.2154
(1, 1.5, 1.5, 2, 0.3)	0.7	0.5054	0.7007	0.7718	0.0329
(1, 1.5, 1.5, 2, 0.7)	0.7	0.6549	0.7007	0.7718	0.1824

Table 5.1: The censoring proportions and correlation for the four different parameter settings

Furthermore, Table 5.2 presents the mean number of iterations to achieve convergence and the mean observed log-likelihood value for the final parameter estimates for each parameter setting in the simulation study.

In general, the algorithm performed well at predicting the true parameter values, with larger sample sizes having slightly better estimates. Notably the algorithm accurately estimates $\phi_{00}, \phi_{01}, \phi_{10}, \phi_{11}$ in all 12 considered settings. Unusually, high censoring leads to lower MSE and bias than low censoring. However, with high censoring, from Table 5.2, it can be seen that more iterations are needed for convergence with the algorithm finding parameter estimates that on an average gave a smaller observed log-likelihood value. This suggests that as the algorithm takes many more iterations, it is more likely to converge to a global maximum as opposed to a local maximum, which would explain why high censoring performs better than low censoring. A solution

Parameter $(\alpha_1, \alpha_2, \beta_1, \beta_2, \rho)$	Sample size	Censoring	Mean iterations	Mean ObsLL
$(1, 1.5, 1.5, 2, 0.3)$	100	low	4.47	-188.29
$(1, 1.5, 1.5, 2, 0.7)$	100	low	9.72	-147.40
$(1, 1.5, 1.5, 2, 0.7)$	100	high	14.36	-108.02
$(1, 1.5, 1.5, 2, 0.3)$	100	high	15.61	-110.98
$(1, 1.5, 1.5, 2, 0.3)$	50	low	4.01	-190.23
$(1, 1.5, 1.5, 2, 0.7)$	50	low	4.37	-182.68
$(1, 1.5, 1.5, 2, 0.7)$	50	high	15.26	-107.25
$(1, 1.5, 1.5, 2, 0.3)$	50	high	15.62	-110.42
$(1, 1.5, 1.5, 2, 0.3)$	25	low	3.74	-188.69
$(1, 1.5, 1.5, 2, 0.7)$	25	low	5.30	-181.74
$(1, 1.5, 1.5, 2, 0.7)$	25	high	14.71	-108.05
$(1, 1.5, 1.5, 2, 0.3)$	25	high	15.17	-110.33

Table 5.2: Mean observed log-likelihood (ObsLL) and mean number of iterations until convergence for each of the parameter settings

to the bad performance of low censoring could be to have multiple sets of starting parameters to ensure convergence to a global maximum, but this would come at the cost of higher computational complexity. Correlation does not appear to have a discernible impact on the parameter estimates either in terms of bias or MSE. Overall, the algorithm performs particularly well when there is a high censoring proportion present in the dataset, which makes it very applicable for real-life datasets.

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.30	0.21	0.27	0.23	0.29
Mean	1.15	1.88	1.47	1.78	0.36	0.21	0.27	0.23	0.29
Bias	0.15	0.38	-0.03	-0.22	0.06	0.00	0.00	0.00	0.00
MSE	0.55	1.83	0.05	0.16	0.05	0.00	0.00	0.00	0.00

(a) Simulation results for low correlation and low censoring

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.70	0.21	0.27	0.23	0.29
Mean	1.19	1.69	1.48	1.91	0.50	0.23	0.25	0.22	0.30
Bias	0.19	0.19	-0.02	-0.09	-0.20	0.02	-0.02	-0.01	0.01
MSE	3.70	0.91	0.07	0.15	0.13	0.00	0.00	0.00	0.01

(b) Simulation results for high correlation and low censoring

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.70	0.21	0.27	0.23	0.29
Mean	1.01	1.49	1.57	2.06	0.69	0.24	0.25	0.21	0.30
Bias	0.01	-0.01	0.07	0.06	-0.01	0.03	-0.02	-0.02	0.01
MSE	0.03	0.04	0.06	0.11	0.04	0.01	0.00	0.01	0.01

(c) Simulation results for high correlation and high censoring

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.30	0.21	0.27	0.23	0.29
Mean	1.01	1.50	1.54	2.08	0.34	0.24	0.24	0.21	0.31
Bias	0.01	0.00	0.04	0.08	0.04	0.03	-0.03	-0.02	0.02
MSE	0.03	0.04	0.05	0.11	0.08	0.01	0.01	0.01	0.01

(d) Simulation results for low correlation and high censoring

Table 5.3: Simulation results for sample size $n = 100$

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.30	0.21	0.27	0.23	0.29
Mean	1.15	1.87	1.46	1.78	0.36	0.21	0.26	0.23	0.29
Bias	0.15	0.37	-0.04	-0.22	0.06	0.00	-0.01	0.00	0.00
MSE	0.23	1.66	0.05	0.34	0.06	0.00	0.00	0.00	0.00

(a) Simulation results for low correlation and low censoring

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.70	0.21	0.27	0.23	0.29
Mean	1.18	1.75	1.45	1.80	0.66	0.21	0.27	0.23	0.30
Bias	0.18	0.25	-0.05	-0.20	-0.04	0.00	0.00	0.00	0.01
MSE	1.31	0.77	0.05	0.15	0.04	0.00	0.00	0.00	0.00

(b) Simulation results for high correlation and low censoring

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.70	0.21	0.27	0.23	0.29
Mean	1.02	1.50	1.56	2.09	0.68	0.24	0.25	0.20	0.31
Bias	0.02	0.00	0.06	0.09	-0.02	0.03	-0.02	-0.03	0.02
MSE	0.03	0.04	0.06	0.13	0.05	0.01	0.00	0.01	0.01

(c) Simulation results for high correlation and high censoring

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.30	0.21	0.27	0.23	0.29
Mean	1.03	1.49	1.57	2.08	0.35	0.24	0.25	0.21	0.31
Bias	0.03	-0.01	0.07	0.08	0.05	0.03	-0.02	-0.02	0.02
MSE	0.04	0.05	0.07	0.12	0.08	0.01	0.01	0.01	0.01

(d) Simulation results for low correlation and high censoring

Table 5.4: Simulation results for sample size $n = 50$

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.30	0.21	0.27	0.23	0.29
Mean	1.16	1.96	1.48	1.78	0.34	0.21	0.27	0.23	0.30
Bias	0.16	0.46	-0.02	-0.22	0.04	0.00	0.00	0.00	0.01
MSE	0.37	2.46	0.06	0.19	0.06	0.00	0.00	0.00	0.00

(a) Simulation results for low correlation and low censoring

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.70	0.21	0.27	0.23	0.29
Mean	1.29	1.82	1.44	1.81	0.67	0.20	0.27	0.23	0.30
Bias	0.29	0.32	-0.06	-0.19	-0.03	-0.01	0.00	0.00	0.01
MSE	6.36	1.83	0.06	0.17	0.04	0.00	0.00	0.00	0.00

(b) Simulation results for high correlation and low censoring

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.70	0.21	0.27	0.23	0.29
Mean	1.01	1.50	1.57	2.11	0.66	0.24	0.25	0.20	0.31
Bias	0.01	0.00	0.07	0.11	-0.04	0.03	-0.02	-0.03	0.02
MSE	0.04	0.04	0.06	0.13	0.05	0.01	0.00	0.01	0.01

(c) Simulation results for high correlation and high censoring

	Parameter								
	α_1	α_2	β_1	β_2	ρ	ϕ_{00}	ϕ_{01}	ϕ_{10}	ϕ_{11}
True value	1.00	1.50	1.50	2.00	0.30	0.21	0.27	0.23	0.29
Mean	1.02	1.48	1.58	2.09	0.35	0.24	0.25	0.21	0.30
Bias	0.02	-0.02	0.08	0.09	0.05	0.03	-0.02	-0.02	0.01
MSE	0.03	0.05	0.06	0.12	0.08	0.01	0.00	0.01	0.01

(d) Simulation results for low correlation and high censoring

Table 5.5: Simulation results for sample size $n = 25$

Chapter 6

Application to Retinopathy Data

6.1 The Dataset

Diabetic retinopathy is when, due to diabetes, the retina in the eye is damaged causing blindness. In 1971, a clinical trial was set up to investigate the effectiveness of using laser treatment, with each person in the trial being given the treatment to one eye and the other eye left untreated [48]. The time until blindness or until censoring was recorded, with blindness defined as when visual acurity drops below 5/200 two visits in a row. A dataset from this trial of a random sample of 197 patients is available in the survival package in R and is used for the analysis; see [55] . Each person was taken as a subject with going blind in each eye as the paired event. Many previous works have analyzed this dataset; some did not use a cure rate model [37] while other more recent papers did use cure rate models [20, 19, 21].

Within the treatment eye, 73% of lifetimes are censored and within the untreated eye, 49% of lifetimes are censored with 41% of observations having both lifetimes censored. Figure 6.1 is a scatterplot of the observed times from the dataset. The

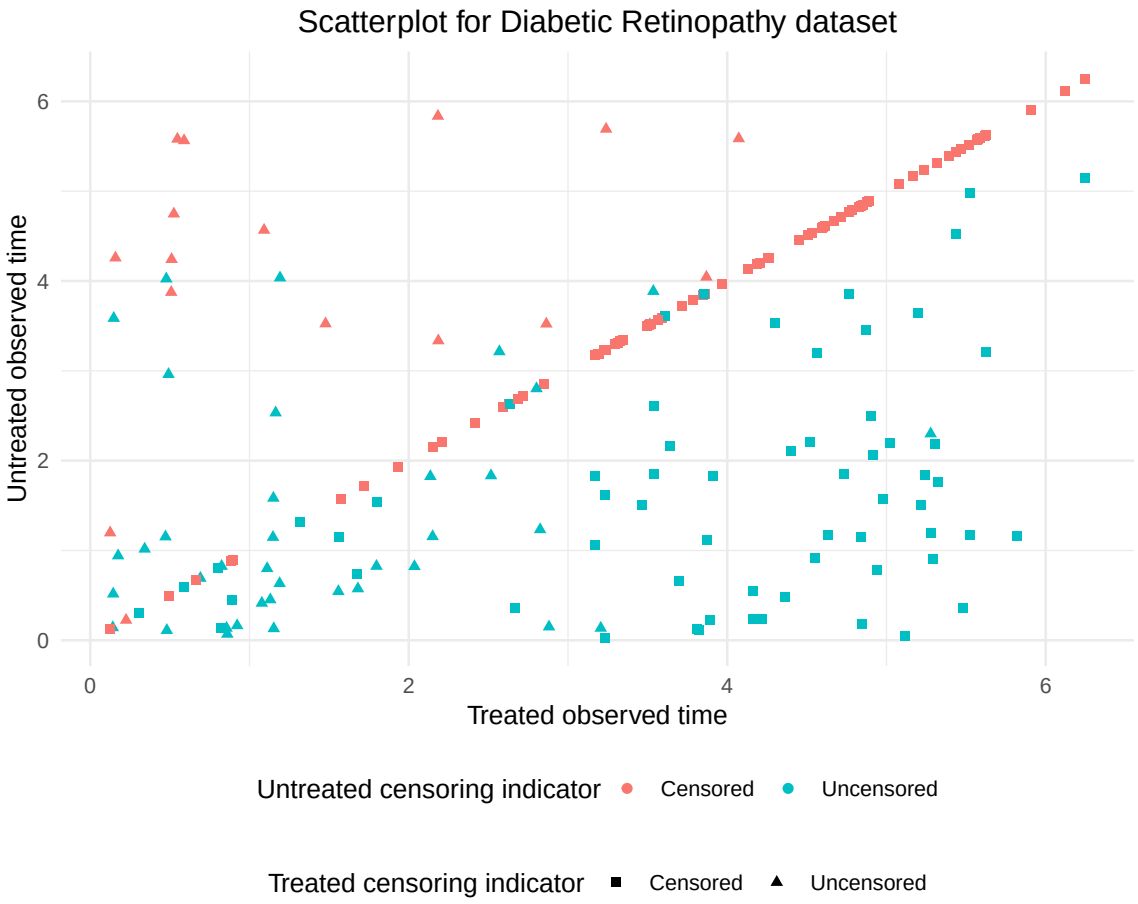

Figure 6.1: Scatterplot of Diabetic Retionopathy dataset showing a small positive correlation within uncensored observations

diagonal line represents when both observations are censored at the same time. This might be due to a person leaving the trial or the trial ending, so that both eyes are censored at the same time. Looking at the uncensored observations, a slight positive correlation may be observed.

6.2 Implementation of the Algorithm and Results

The EM algorithm described in Chapter 4 was implemented on the dataset and it converged in 14 iterations. To find the standard errors and confidence intervals, a parametric bootstrap algorithm was implemented, as described in Section 2.2.3. This meant that using the estimated parameters and censoring proportion of the data, the censoring parameter, λ, could be calculated using (5.2.5). Then, using the simulation method as outlined in Section 5.1, 1500 datasets were generated from the assumed model. The EM algorithm described in Chapter 4 was then implemented on each of the datasets to generate bootstrap replications. The standard deviation of these bootstrap replications was used as the standard error. To find 90% confidence intervals, the BC_a percentiles of the bootstrap replications were used.

The parameter estimates, standard error and confidence intervals obtained from the algorithm are presented in Table 6.1, with the confidence intervals being relatively wide suggesting uncertainty in the model.

Parameter	Estimate	SE	CI
α_1	2.5848	0.56	[1.92, 3.72]
α_2	2.4368	0.45	[1.86, 3.36]
β_1	1.0755	0.18	[0.81, 1.34]
β_2	0.9893	0.13	[0.79, 1.17]
ρ	0.4729	0.21	[0.00, 0.72]
ϕ_{00}	0.2797	0.06	[0.19, 0.37]
ϕ_{01}	0.3549	0.06	[0.26,0.46]
ϕ_{10}	0.0685	0.04	[0.02, 0.14]
ϕ_{11}	0.2968	0.06	[0.21, 0.40]

Table 6.1: Results from the EM algorithm applied to the Diabetic Retionpathy dataset

The observed log-likelihood value, based on this set of parameter values, was computed to be -440.7872 with the corresponding AIC as 899.5745. From the results presented in Table 6.1 and from (2.3.9) the cure rate for the treated eyes is $p_X = \phi_{00} + \phi_{01} = 0.28 + 0.35 = 0.63$ and for untreated eyes it is $p_Y = \phi_{00} + \phi_{10} = 0.28 + 0.07 = 0.35$. This suggests that the laser treatment drastically increased the probability that an eye would be cured from blindness. Using (3.1.14) the estimate for the correlation of the Weibull variables, ρ_w, is 0.4725. This suggests a medium positive correlation exists between the time to blindness in each eye in a patient. The covariance between the probability of a subject being susceptible to both events, ω, can then be calculated as $\phi_{00} - p_X p_Y = 0.2797 - (0.63)(0.35) = 0.0592$ which suggests a small positive correlation. Both the treated eyes and untreated eyes have similar parameter estimates for α and β, but different probabilities of being susceptible to going blind. This suggests that the treatment reduces the probability of being susceptible, but does not slow down blindness for the eyes that are susceptible. Furthermore, both β values are close to 1, with the confidence intervals containing 1, which suggests that exponential marginals may also be appropriate. The Kaplan-Meier plot and Weibull survival curves for the treated and untreated eyes can be seen in Figure 6.2 which shows a very close agreement between the Kaplan-Meier survival function and the marginal weibull distribution survival functions fitted in the model. This suggests the Weibull distribution provides a good fit.

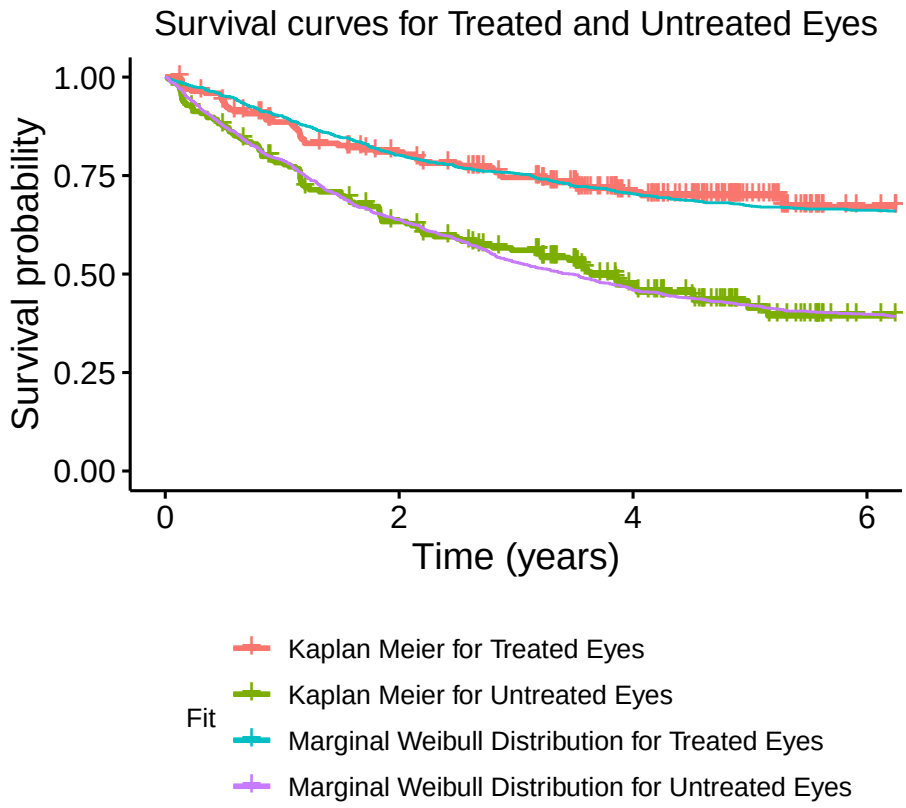

Figure 6.2: Marginal Survival Plots for Treated and Untreated Eyes

6.3 Comparison with Other Models

The estimates from this model can be compared to those from some other models that have been fitted to the dataset. Table 6.2 compares the results from this model, the Moran-Downton bivariate Weibull (MDBW), with the Basu–Dhar bivariate geometric (BDBG) cure rate model [19], a Bivariate Weibull with Generalized Farlie-Gumbel-Morgenstern (GFGM) copula cure rate model [21], and a Bivariate Weibull with Clayton-Oakes distribution, but not including cure fraction [37]. The log-likelihood values, AIC values and correlation estimates were also calculated for all the models [1]. It can be seen that the GFGM model fitted the data the best, closely followed by the Moran-Downton bivariate Weibull model. The non-cure rate model performed the worst suggesting that a cure rate model is essential for this dataset. The Weibull parameter estimates and the cure proportion estimates were very similar between the GFGM and Moran-Downton bivariate Weibull models, but their estimates for the correlation between susceptible subject lifetimes differed. The Weibull estimates for the cure rate models differed from those of the non-cured model, but that is to be expected as they are based on different assumptions. The BDBG model estimated a much smaller correlation between the event times of susceptible subjects, but this model had the largest AIC, and so did not provide a good fit. The MDWB model found very similar estimates for $\phi_{00}, \phi_{01}, \phi_{10}$ and ϕ_{11} compared to the BDBG model. Overall, the Moran-Downton bivariate Weibull model fits the Diabetic Retinopathy dataset reasonably well, and also provides clear evidence that a cure rate model is necessary for this dataset.

	BDBG [19]	GFGM copula [21]	Weibull not cured [37]	MDBW
α_1		2.5913	15.53	2.5848
α_2		3.0413	5.81	2.4368
β_1		0.9454	0.79	1.0755
β_2		1.0340	0.82	0.9893
ρ_w	0.0986	0.2080	0.63	0.4725
ϕ_{00}	0.2862	0.2482	0	0.2797
ϕ_{01}	0.3383	0.3226	0	0.3549
ϕ_{10}	0.0601	0.0624	0	0.0685
ϕ_{11}	0.3154	0.3668	1	0.2968
LogLikelihood	-590.99	-401.82	-829.58	-440.79
AIC	1187.98	827.63	1677.16	899.58

Table 6.2: Comparison of parameter estimates for the Diabetic Retinopathy dataset based on different models

Chapter 7

Conclusions

This thesis has successfully developed and fitted a new bivariate cure rate model based on the Moran-Downton bivariate Weibull distribution. Weibull marginals within cure rate models had been studied before due to their wide range of applications, but this is the first time when the Moran-Downton bivariate Weibull distribution is used in cure rate modelling.

Furthermore, this thesis has developed and applied an EM algorithm to the bivariate cure rate model for the first time. An empirical study, with various settings, has been performed showing that the proposed algorithm performs well. The algorithm is slow, most likely due to the high computational complexity of the Moran-Downton bivariate Weibull distribution and did not perform as well on datasets with low censoring rates. The algorithm appeared to perform particularly well at predicting the cure fractions across all the settings considered.

The algorithm has been used on a dataset on Diabetic Retinopathy; the parameter estimates were in line with previously published research analyzing the same data. The parameter estimates suggest that exponential marginals may also be appropriate

for modelling this dataset. The bootstrap method has been used to find the standard errors and confidence intervals for the parameters, with the confidence intervals being relatively wide suggesting uncertainty within the model fitted to the Diabetic Retinopathy dataset.

Further problems to consider could include looking at other bivariate lifetime models fitted via EM algorithms, adaptations to the algorithm that improve the speed and accuracy, other censoring mechanisms and also incorporation of covariates into the model.

Appendix A

R code

A.1 EM Algorithm

For calculating the conditional expectations as seen in (4.1.3) - (4.1.6):

```
opifun<-function(x,y,a1,a2,b1,b2,rho, c1,c2, phi00, phi01, phi10, lg, lg2){
  phi11<-1-phi00-phi10-phi01
  probs=matrix(data=NA, nrow=length(x), ncol=4)
  for(i in 1:length(x)){
    if(c1[i]==0 & c2[i]==0){
      probs[i,1]<-0
      probs[i,2]<-0
      probs[i,3]<-0
      probs[i,4]<-1}
    if(c1[i]==0 & c2[i]==1){
      probs[i,1]<-0
```

```
      probs[i,2]<-0
      probs[i,3]<-phi10*(dweibull(x[i],b1,a1))/(phi10*(dweibull(x[i],b1,a1))
        -phi11*omdbwsdashx(x[i],y[i],a1,a2,b1,b2, rho, lg, lg2))
      probs[i,4]<-1-probs[i,3]}
    if(c1[i]==1 & c2[i]==0){
      probs[i,1]<-0
      probs[i,2]<-phi01*(dweibull(y[i],b2,a2))/(phi01*(dweibull(y[i],b2,a2))
        -phi11*omdbwsdashx(y[i],x[i],a2,a1,b2,b1, rho, lg ,lg2))
      probs[i,3]<-0
      probs[i,4]<-1-probs[i,2]}
    if(c1[i]==1 & c2[i]==1){
      bottom<-phi11*omdbwsurv(x[i],y[i],a1,a2,b1,b2, rho, lg, lg2)
        + phi10*(1-pweibull(x[i],b1,a1)) +phi01*(1-pweibull(y[i],b2,a2))+phi00
      probs[i,1]<-phi00/bottom
      probs[i,2]<-phi01*(1-pweibull(y[i],b2,a2))/bottom
      probs[i,3]<-phi10*(1-pweibull(x[i],b1,a1))/bottom
      probs[i,4]<-1- probs[i,1]- probs[i,2]- probs[i,3]}}
  return(probs)
}
```

For a single observation the Q-function, as seen in (4.1.7) is:

```
olikelihoodfun<-function(x,y,a1,a2,b1,b2,rho,phi00,phi01,phi10, pi1, pi2,
    pi3, pi4, c1, c2, lg, lg2){
  phi11<-1-phi00-phi10-phi01
```

```
  if(c1==0 & c2==0){ ##both observed
    ans<-log(phi11*omdbwpdf(x,y,a1,a2,b1,b2, rho))}
  if(c1==0 & c2==1){
    ans<-pi4*log(-phi11*omdbwsdashx(x,y,a1,a2,b1,b2, rho, lg, lg2))
        +pi3*log(phi10*dweibull(x,b1,a1))}
  if(c1==1 & c2==0){
    ans<-pi4*log(-phi11*omdbwsdashx(y,x,a2,a1,b2,b1, rho, lg, lg2))
        +pi2*log(phi01*dweibull(y,b2,a2))}
  if(c1==1 & c2==1){
    ans<-pi4*log(phi11*omdbwsurv(x,y,a1,a2,b1,b2, rho, lg, lg2))
        +pi3*log(phi10*(1-pweibull(x,b1,a1)))
        +pi2*log(phi01*(1-pweibull(y,b2,a2)))+pi1*log(phi00)}
  return(ans)
}
```

The value of the Q-function for an entire dataset

```
onegloglikefun2<-function(paraest, piest, dataset, lg, lg2){
  a1<-paraest[1]
  a2<-paraest[2]
  b1<-paraest[3]
  b2<-paraest[4]
  rho<-paraest[5]
  phi00<-exp(paraest[6])/(1+exp(paraest[6]))
  phi01<-1-phi00-exp(paraest[7])*(1-phi00)/(1+exp(paraest[7]))
  phi10<-1-phi00-phi01-exp(paraest[8])*(1-phi00-phi01)/(1+exp(paraest[8]))
```

```
  x<-dataset[,1]
  y<-dataset[,2]
  c1<-dataset[,3]
  c2<-dataset[,4]
  pi1<-piest[,1]
  pi2<-piest[,2]
  pi3<-piest[,3]
  pi4<-piest[,4]
  ans<-0
  for (i in 1:length(x)){
    b<-olikelihoodfun(x[i],y[i],a1,a2,b1,b2,rho,phi00,phi01,phi10,
        pi1[i], pi2[i], pi3[i], pi4[i], c1[i], c2[i], lg, lg2)
    ans<-ans+b}
  return(-ans)
}
```

For the joint pdf of the Moran Downton Bivariate Weibull distribution as seen in (3.1.2).

```
omdbwpdf<-function(x,y,a1,a2,b1,b2, rho){
  x1<-(x/a1)^b1
  y1<-(y/a2)^b2
  ans<-b1*b2*x1*y1/((1-rho)*x*y)*exp(-(x1+y1)/(1-rho))*
    besselI(2*sqrt(rho*x1*y1)/(1-rho),0)
  return(ans)
```

```
}
```

For the joint survival function of the Moran Downton Bivariate Weibull distribution as seen in (3.1.4)

```
omdbwsurv<-function(x,y,a1,a2,b1,b2, rho, lg, lg2){
  Term<-rep(0,n)
  x1<-(x/a1)^b1
  y1<-(y/a2)^b2
  for(j in 1:n){
    lgj<-as.function(lg[[j]])
    k<-j-1
    Term[j]<-((rho^(k+1))/(k+1)^2)*lgj(x1)*lgj(y1)}
  ans<-exp(-x1-y1)+x1*y1*exp(-x1-y1)*sum(Term)
  return(ans)
}
```

For the joint survival function differentiated with respect to x of the Moran Downton Bivariate Weibull distribution as seen in (4.1.8). Due to symmetry this function was also used for Equation (4.1.9) with the input order changing to $(y, x, a_2, a_1, b_2, b_1, \rho)$.

```
 omdbwsdashx<-function(x,y,a1,a2,b1,b2, rho, lg, lg2){
  Term<-rep(0,n)
  x1<-(x/a1)^b1
  y1<-(y/a2)^b2
  for(j in 1:n){
    lgj<-as.function(lg[[j]])
```

```
    k<-j-1
    Term[j]<-((rho^(k+1))/(k+1)^2)*lgj(x1)*lgj(y1)
  }
  ans<- -b1/x*x1*exp(-x1-y1)+(b1/x-b1*x1/x
    -2*b1*x1/(x^2))*x1*y1*exp(-x1-y1)*sum(Term)
  return(ans)
}
```

For the observed loglikelihood of the dataset, as needed to test for convergence, from (4.1.10):

```
obsloglike<-function(x,y,a1,a2,b1,b2,rho,phi00,phi01,phi10, c1, c2, lg, lg2){
  phi11<-1-phi00-phi10-phi01
  if(c1==0 & c2==0){
    ans<-log(phi11*omdbwpdf(x,y,a1,a2,b1,b2,rho))
  }
  if(c1==0 & c2==1){
    ans<-log(-phi11*omdbwsdashx(x,y,a1,a2,b1,b2,rho, lg, lg2)
        +phi10*dweibull(x,b1,a1))
  }
  if(c1==1 & c2==0){
    ans<-log(-phi11*omdbwsdashx(y,x,a2,a1,b2,b1,rho, lg, lg2)
        +phi01*dweibull(y,b2,a2))
  }
  if(c1==1 & c2==1){
    ans<-log(phi11*omdbwsurv(x,y,a1,a2,b1,b2,rho, lg, lg2)
```

```
        +phi10*(1-pweibull(x,b1,a1))+phi01*(1-pweibull(y,b2,a2))+phi00)
  }
  return(ans)
}

compobslikefun<-function(x,y,a1,a2,b1,b2,rho,phi00,phi01,phi10, c1,c2,
    lg, lg2){
  ans<-0
  for (i in 1:length(x)){
    ans<-ans+obsloglike(x[i],y[i],a1,a2,b1,b2,rho,phi00,phi01,phi10,
        c1[i], c2[i], lg, lg2)}
  return(ans)
}
```

To find the starting parameters:

```
startingparafun<-function(dataset){
  xnocen<-dataset[which(dataset$cenx==0), 1]
  ynocen<-dataset[which(dataset$ceny==0), 2]
  bothnocen<-dataset[which(dataset$cenx==0 & dataset$ceny==0), 1:2]
  wf<-fitdist(xnocen, "weibull")
  a1<-wf$estimate[2]
  b1<-wf$estimate[1]
  wf<-fitdist(ynocen, "weibull")
  a2<-wf$estimate[2]
```

```
  b2<-wf$estimate[1]
  ifelse(length(bothnocen[,1])==0,
    rho<-0.5, {bothnocen<-cbind(bothnocen$xobs^(1/b1),
    bothnocen$yobs^(1/b2)); rho<-cor(bothnocen[,1], bothnocen[,2])})
  if(rho< 0){
    rho<-0.001}
  if(rho>0.99){
    rho<-0.99}
  return(c(a1,a2,b1,b2,rho))
}
```

The overall function that implements the full EM algorithm:

```
emfun<-function(dataset, it, paraest,lg, lg2){
  k=0
  x<-dataset[,1]
  y<-dataset[,2]
  c1<-dataset[,3]
  c2<-dataset[,4]
  comp<-0
  a1<-paraest[1]
  a2<-paraest[2]
  b1<-paraest[3]
  b2<-paraest[4]
  rho<-paraest[5]
```

```
phi00<-exp(paraest[6])/(1+exp(paraest[6]))
phi01<-1-phi00-exp(paraest[7])*(1-phi00)/(1+exp(paraest[7]))
phi10<-1-phi00-phi01-exp(paraest[8])*(1-phi00-phi01)/(1+exp(paraest[8]))
comp1<-compobslikefun(x,y,a1,a2,b1,b2,rho,phi00,phi01,phi10, c1, c2, lg, lg2)
while (k<it & abs(comp-comp1)> 0.0001 ){
  k=k+1
  piest<- opifun(x,y,a1,a2,b1,b2,rho, c1,c2, phi00, phi01, phi10, lg, lg2)
  a<-optim(paraest, fn=onegloglikefun2, piest=piest, dataset=dataset,
      method="L-BFGS-B", lower=c(0.001,0.001,0.001,0.001,0,-Inf,-Inf,-Inf),
      upper=c(100,100,100,100, 0.99, Inf, Inf, Inf), lg=lg, lg2=lg2)
  paraest<-a$par
  a1<-paraest[1]
  a2<-paraest[2]
  b1<-paraest[3]
  b2<-paraest[4]
  rho<-paraest[5]
  phi00<-exp(paraest[6])/(1+exp(paraest[6]))
  phi01<-1-phi00-exp(paraest[7])*(1-phi00)/(1+exp(paraest[7]))
  phi10<-1-phi00-phi01-exp(paraest[8])*(1-phi00-phi01)/(1+exp(paraest[8]))
  phi11<-1-phi00-phi01-phi10
  comp<-comp1
  comp1<-compobslikefun(x,y,a1,a2,b1,b2,rho,phi00,phi01,phi10, c1, c2,
      lg, lg2)}
return(c(a1,a2,b1,b2,rho,phi00,phi01, phi10, phi11, comp1,k))
```

```
}
```

A.2 Empirical Study

This code generates a Dataset using the algorithm outlined in Section 5.1.

```
newdatagen1<-function(a1,a2,b1,b2,n, rho, phi00, phi10, phi01, xrate, yrate){
  xobs<-c()
  yobs<-c()
  cenx<-c()
  ceny<-c()
  expdata<-rBED(rho=rho, 1,1,n)
  for(i in 1:n){
    u<-runif(1)
    if(u< phi00){
      cenx[i]<-1
      ceny[i]<-1
      xobs[i]<-rexp(1, xrate)
      yobs[i]<-rexp(1, yrate)
    }
    if(phi00<u & u< (phi00+phi01)){
      xobs[i]<-a1*expdata[i,1]^(1/b1)
      centimex<-rexp(1, xrate)
      ifelse(centimex<xobs[i], {xobs[i]<-centimex; cenx[i]<-1}, cenx[i]<-0)
      yobs[i]<-rexp(1, yrate)
      ceny[i]<-1
```

```
    }
    if(phi00+phi10<u & u<phi00+phi01+phi10){
      yobs[i]<-a2*expdata[i,2]^(1/b2)
      centimey<-rexp(1, yrate)
      ifelse(centimey<yobs[i], {yobs[i]<-centimey; ceny[i]<-1}, ceny[i]<-0)
      xobs[i]<-rexp(1, xrate)
      cenx[i]<-1
    }
    if(phi00+phi10+phi01<u){
      xobs[i]<-a1*expdata[i,1]^(1/b1)
      yobs[i]<-a2*expdata[i,2]^(1/b2)
      centimex<-rexp(1, xrate)
      centimey<-rexp(1, yrate)
      ifelse(centimex<xobs[i], {xobs[i]<-centimex; cenx[i]<-1}, cenx[i]<-0)
      ifelse(centimey<yobs[i], {yobs[i]<-centimex; ceny[i]<-1}, ceny[i]<-0)}}
  return(data.frame(xobs,yobs, cenx, ceny))
}
```

A.2.1 Calculation of ρ_w

This calculates ρ_w according to (3.1.14).

```
b1<-1
b2<-1
rho<-0.7
lst<-seq(from=0, to=150, by=1)
```

```
infsum<-sum(rho^lst/factorial(lst)^2*gamma(lst+1/b1+1)*gamma(lst+1/b2+1))
top<-(1-rho)^(1/b1+1/b2+1)*infsum-gamma(1+1/b1)*gamma(1+1/b2)
bottom<-(gamma(1+2/b1)-gamma(1+1/b1)^2)^(1/2)*(gamma(1+2/b2)
    -gamma(1+1/b2)^2)^(1/2)
ans<-top/bottom
```

A.2.2 Expected censoring proportion

This calculates the censoring proportion for completely censored observations using (5.2.5):

```
fun1 <- function(c,cenrate, a, b){
  (1-pweibull(c,shape=b,scale=a))*cenrate*exp(-cenrate*c)
}
funstar<-function(c, a1,a2,b1,b2,cenrate, rho){
  n=50
  lg<-orthopolynom::glaguerre.polynomials(n=n, alpha=1, normalized=FALSE)
  Term<-rep(0,n)
  x1<-(c/a1)^b1
  y1<-(c/a2)^b2
  for(j in 1:n){
    lgj<-as.function(lg[[j]])
    k<-j-1
    Term[j]<-((rho^(k+1))/(k+1)^2)*lgj(x1)*lgj(y1)}
  ans<-exp(-x1-y1)+x1*y1*exp(-x1-y1)*sum(Term)
  cenrate*exp(-cenrate*c)*(exp(-x1-y1)+x1*y1*exp(-x1-y1))*sum(Term)
```

```
}

compcenprob<-function(a1,a2,b1,b2,cenrate, phi00, phi01, phi10, rho){
  phi11=1-phi00-phi01-phi10
  ans<-phi00+phi10*integrate(f=fun1, lower=0, upper= Inf, cenrate=cenrate,
    a=a1, b=b1)$value+phi01*integrate(f=fun1, lower=0, upper= Inf,
    cenrate=cenrate, a=a2, b=b2)$value+phi11*integrate(Vectorize(funstar),
    lower=0, upper= Inf, cenrate=cenrate,
    a1=a1, a2=a2, b1=b1,b2=b2, rho=rho)$value
  return(ans)
}
```

For marginal censoring proportion within each outcome, using Equation 5.2.1:

```
partialcenprob<-function(a,b,cenrate, phi00, phi01, phi10, rho){
  phi11=1-phi00-phi10-phi01
  ans<-phi00+phi01+(phi10+phi11)*integrate(f=fun1,
    lower=0, upper= Inf, cenrate=cenrate, a=a1, b=b1)$value
    +phi01*integrate(f=fun1, lower=0, upper= Inf, cenrate=cenrate, a=a2,
    b=b2)$value+phi11*integrate(Vectorize(funstar), lower=0,
    upper= Inf, cenrate=cenrate, a1=a1, a2=a2, b1=b1,b2=b2, rho=rho)$value
  return(ans)
  }
```

Bibliography

[1] J. A. Achcar, N. Davarzani, and R. M. Souza. Basu–Dhar bivariate geometric distribution in the presence of covariates and censored data: a Bayesian approach. *Journal of Applied Statistics*, 43(9):1636–1648, 2016.

[2] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*. John Wiley & Sons, New York, 1992.

[3] S. G. Baker. A simple method for computing the observed information matrix When using the EM algorithm with categorical data. *Journal of Computational and Graphical Statistics*, 1(1):63–76, 1992.

[4] N. Balakrishnan and C. D. Lai. *Continuous bivariate distributions*. Springer, New York, 2009.

[5] O. S. Balogun, X. Z. Gao, E. T. Jolayemi, and S. A. Olaleye. Generalized cure rate model for infectious diseases with possible co-infections. *PLoS ONE*, 15, 9 2020.

[6] A. V. Belikov. The number of key carcinogenic events can be predicted from cancer incidence. *Scientific Reports*, 7(1):12170–8, 2017.

[7] J. Berkson and R. P. Gage. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47:501–515, 1952.

[8] C. Biernacki, G. Celeux, and G. A. E. Govaert. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis*, 41:561–575, 2003.

[9] J. W. Boag. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11:15–53, 1949.

[10] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5): 1190–19, 09 1995.

[11] G. Celeux, S. Chrétien, F. Forbes, and A. Mkhadri. A component-wise EM algorithm for mixtures. *Journal of Computational and Graphical Statistics*, 10: 697–712, 2001.

[12] N. Chatterjee and J. Shih. A bivariate cure-mixture approach for modeling familial association in diseases. *Biometrics*, 57:779–786, 2001.

[13] D. G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151, 1978.

[14] D. Cox and D. Oakes. *Analysis of Survival Data*. Chapman and Hall/CRC Press, London, 1984.

[15] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society*, 34:187–220, 1972.

[16] K. Davies, S. Pal, and J. A. Siddiqua. Stochastic EM algorithm for generalized exponential cure rate model and an empirical study. *Journal of Applied Statistics*, 48(12):2112–2135, 2020.

[17] A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, England, 1997.

[18] R. de la Cruz, C. Fuentes, and O. Padilla. A Bayesian mixture cure rate model for estimating short-term and long-term recidivism. *Entropy*, 25, 1 2023.

[19] R. P. de Oliveira, J. A. Achcar, D. Peralta, and J. Mazucheli. Discrete and continuous bivariate lifetime models in presence of cure rate: a comparative study under Bayesian approach. *Journal of Applied Statistics*, 46:449–467, 2019.

[20] R. P. de Oliveira, M. V. de Oliveira Peres, E. Z. Martinez, and J. A. Achcar. A new cure rate regression framework for bivariate data based on the Chen distribution. *Statistical Methods in Medical Research*, 31:2442–2455, 2022.

[21] M. V. de Oliveira Peres, J. A. Achcar, and E. Z. Martinez. Bivariate lifetime models in presence of cure fraction: a comparative study with many different copula functions. *Heliyon*, 6, 2020.

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

[23] Y. Dodge and F. H. C. Marriott. *The Oxford Dictionary of Statistical Terms*. Oxford University Press, Oxford, England, sixth edition, 2003.

[24] F. Downton. Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32:408–417, 1970.

[25] L. F. Duque. *MDBED: Moran-Downton Bivariate Exponential Distribution*, 2020. URL `https://CRAN.R-project.org/package=MDBED`. R package version 1.0.0.

[26] B. Efron. Bootstrap methods: Another look at the jacknife. *The Annals of Statistics*, 7:1–26, 1979.

[27] B. Efron. Censored data and the bootstrap. *Journal of the American Statistical Association*, 76:312–319, 1981.

[28] B. Efron. Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397):171–185, 1987. URL `http://www.jstor.org/stable/2289144`.

[29] B. Efron and B. Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Sciemce*, 1:54–77, 1986.

[30] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, New York, 1993.

[31] C. Forbes. *Statistical Distributions*. Hohn Wiley & Sons, 4th edition, 2011.

[32] M. Fréchet. Sur les tableaux dont les marges et des bornes sont données. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 28(1/2):10–32, 1960.

[33] A. J. Gross. *Survival Distributions : Reliability Applications in the Biomedical Sciences.* John Wiley & Sons, New York, 1975.

[34] E. J. Gumbel. Bivariate exponential distributions. *Journal of the American Statistical Association*, 55:698–707, 1960.

[35] M. R. Gupta and Y. Chen. Theory and use of the EM algorithm. *Foundations and Trends in Signal Processing*, 4:223–296, 2010.

[36] P. Hougaard. Frailty models for survival data. *Lifetime data analysis*, 1(3): 255–273, 1995.

[37] W. J. Huster, R. Brookmeyer, and S. G. Self. Modelling paired survival data with covariates. *Biometrics*, 45:145–156, 1989.

[38] A. Høyland. *System reliability theory models and statistical methods.* Wiley series in probability and mathematical statistics. John Wiley & Sons, Hoboken, New Jersey, 2nd edition, 2004.

[39] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data.* Springer, New York, 2nd edition, 2002.

[40] Y. J. Kim. Cure rate model with bivariate interval censored data. *Communications in Statistics: Simulation and Computation*, 46:7116–7124, 2017.

[41] D. G. Kleinbaum and M. Klein. *Statistics for Biology and Health Survival Analysis A Self-Learning Text*. Springer, New York, 3rd edition, 1996.

[42] S. W. Lagakos. General right censoring and its impact on the analysis of survival data. *Biometrics*, 35(1):139–156, 1979.

[43] C. D. Lai, D. Murthy, and M. Xie. *Weibull Distributions and Their Applications*. Springer, London, 2006. doi: 10.1007/978-1-84628-288-1_3.

[44] L. M. Leemis. Lifetime distribution identities. *IEEE Transactions on Reliability*, 35(2):170–174, 1986.

[45] G. J. McLachlan and T. Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, Hoboken, New Jersey, 2nd edition, 2008.

[46] P. A. P. Moran. Testing for correlation between non-negative variates. *Biometrika*, 54:385–394, 1967.

[47] M. Nagao and M. Kadoya. Two-variate exponential distribution and its numerical table for engineering application. *Bulletin of the Disaster Prevention Research Institute*, 1971.

[48] A. Patz, S. Fine, and et al. Preliminary report on effects of photocoagulation therapy. *American Journal of Ophthalmology*, 1 1976.

[49] Y. Peng and K. B. G. Dear. A nonparametric mixture model for cure rate estimation. *Biometrics*, 56:237–243, 2000.

[50] R Core Team. *R: A Language and Environment for Statistical Computing*. R

Foundation for Statistical Computing, Vienna, Austria, 2022. URL `https://www.R-project.org/`.

[51] J. Rodrigues, M. de Castro, V. G. Cancho, and N. Balakrishnan. COM-Poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference*, 139:3605–3611, 2009.

[52] P. Schmidt and A. D. Witte. Predicting criminal recidivism using 'split population' survival time models. *Journal of Econometrics*, 40:141–159, 1989.

[53] S. Selvin. *Survival Analysis for Epidemiologic and Medical Research*. Practical Guides to Biostatistics and Epidemiology. Cambridge University Press, Cambridge, England, 2008.

[54] M. Sklar. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231, 1959.

[55] T. M. Therneau. A Package for Survival Analysis in R, 2023.

[56] W. Van Assche. Ordinary special functions. In *Encyclopedia of Mathematical Physics*, pages 637–645. Academic Press, Oxford, 2006.

[57] G. J. Van Den Berg. Chapter 55 - duration models: Specification, identification and multiple durations. In *Handbook of econometrics*, volume 5, pages 3381–3460. Elsevier, Amsterdam, 2001.

[58] R. Varadhan and C. Roland. Simple and globally convergent methods for accelerating the convergence of any EM algorithm. *Scandinavian Journal of Statistics*, 35:335–353, 2008.

[59] J. W. Vaupel, K. G. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454, 1979.

[60] W. Weibull. A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 18:293–297, 1951.

[61] A. Wienke. Frailty Models. Working Paper, Max Planck Institute for Demographic Research, 2003.

[62] A. Wienke, N. V. Holm, K. Christensen, A. Skytthe, J. W. Vaupel, and A. I. Yashin. The heritability of cause-specific mortality: a correlated gamma-frailty model applied to mortality due to respiratory diseases in Danish twins born 1870–1930. *Statistics in Medicine*, 22(24):3873–3887, 2003.

[63] A. Wienke, P. Lichtenstein, and A. I. Yashin. A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. *Biometrics*, 59: 1178–1183, 2003.

[64] A. Wienke, I. Locatelli, and A. I. Yashin. The modelling of a cure fraction in bivariate time-to-event data. *Austrian Journal of Statistics*, 35:67–76, 2006.